

MIROSLAW GAJER

Wyspecjalizowany system automatycznego przekładu zrealizowany metodą wzorców translacyjnych

Streszczenie

Tematyka niniejszej monografii stanowi bezpośrednią kontynuację uprzednio wydanej książki autora noszącej tytuł *Wielojęzyczne systemy automatycznego przekładu oparte na metodzie wzorców translacyjnych*. W książce tej zostały przedstawione podstawowe koncepcje teoretyczne leżące u podstaw opracowanej przez autora metody komputerowego przekładu opartej na tzw. wzorcach translacyjnych (ang. *Pattern-Based Machine Translation*). Ponadto w wymienionej książce zostały zamieszczone wyniki obszernych analiz dotyczących możliwości i ograniczeń zastosowania metody wzorców translacyjnych w przypadku wyboru różnych języków świata jako występujących w roli języka źródłowego bądź języka docelowego przekładu. Na podstawie przeprowadzonych analiz wykazano, że metoda wzorców translacyjnych jest metodą w wysokim stopniu uniwersalną i po stosownych modyfikacjach może zostać zastosowana w celu dokonania przekładu komputerowego między właściwie dowolnie wybraną parą języków. Istota metody automatycznego przekładu opartego na wzorcach translacyjnych polega na przeniesieniu procesu translacji z poziomu pojedynczych wyrazów, co stanowi cechę charakterystyczną dla wszelkiego typu systemów komputerowego przekładu, których zasady działania zostały oparte na regułach (ang. *Rule-Based Machine Translation*), do poziomu wielowyrazowych fraz, które same w sobie zawierają znaczną porcję informacji kontekstowej definiującej jednoznacznie ich znaczenie. W ten sposób w znacznej mierze wyeliminowane zostaje niebezpieczeństwo pojawienia się w trakcie realizacji procesu przekładu jakiegokolwiek wieloznaczności, która stanowi podstawową przeszkodę w sprawnym funkcjonowaniu systemów automatycznego przekładu opartych na regułach. Istota metody wzorców translacyjnych polega na tym, że zasady jej działania wywodzą się bezpośrednio z technik tłumaczenia komputerowego opartego na przykładach (ang. *Example-Based Machine Translation*), w przypadku których system automatycznej translacji wykorzystuje wiedzę i uprzedni dorobek translatorski tłumaczy specjalizujących się w przekładzie określonego typu tekstów. Opracowana przez autora metoda wzorców translacyjnych pozwala na zastosowanie technik tłumaczenia opartego na przykładach również w przypadku języków fleksyjnych, takich jak na przykład język polski, ponieważ wzorce translacyjne zawierają także informację dotyczącą reguł składniowych, które pozwalają na automatyczne dopasowywanie form fleksyjnych tłumaczonych fraz.

Prezentowana monografia stanowi kompleksowy raport z wdrożenia systemu komputerowego przekładu opartego na metodzie wzorców translacyjnych. Omawiany w monografii tłumacz komputerowy jest systemem wyspecjalizowanym, czyli takim, który przewidziany jest do sporządzania automatycznego przekładu tekstów o ściśle określonej tematyce. Wy-

bór takiego rozwiązania podyktowany został faktem związanym z bardzo dużymi nakładami pracy, które są wymagane w celu budowy komputerowego tłumacza przeznaczonego do automatycznego przekładu tekstów o dowolnie wybranej przez użytkownika tematyce. Budowa takiego uniwersalnego systemu komputerowego tłumacza wymagała będzie wielu lat pracy licznego zespołu badawczego skupiającego specjalistów z różnych dziedzin i jest przedsięwzięciem, którego może podjąć się jedynie potężna firma informatyczna. W kontekście powyższych uwag opracowanie przez autora wyspecjalizowanego systemu komputerowego przekładu należy postrzegać przede wszystkim jako wyraz realizmu i trzeźwej oceny możliwości praktycznej budowy tego typu systemów. W wyniku realizacji omawianego projektu powstał system komputerowego tłumacza przewidziany do dokonywania automatycznej translacji tekstów komunikatów dotyczących oceny bieżącej sytuacji panującej na globalnym rynku wymiany walutowej – *forex*. Rozważane komunikaty charakteryzują się stereotypową budową, a występujące w nich słownictwo i swoista frazeologia stanowią zbiór już w znacznej mierze zamknięty. Taki stan rzeczy umożliwił opracowanie potrzebnych do dokonania automatycznego przekładu wzorców translacyjnych. W efekcie powstał system komputerowego tłumacza, który z dużą precyzją i wiernością jest w stanie dokonywać w pełni automatycznych przekładów komunikatów dotyczących zachowania się wybranych par walutowych. Rozważane komunikaty są publikowane regularnie i licznie na odpowiednich stronach internetowych, co uzasadnia budowę tego rodzaju systemu wyspecjalizowanego. Ponadto, jak wykazały przeprowadzone badania, opracowany przez autora komputerowy system automatycznego przekładu jest w stanie w wielu wypadkach tłumaczyć rozważane komunikaty lepiej, niż zrobiły to przypadkowy tłumacz, który nie specjalizuje się w przekładzie rozważanego typu tekstów i występująca w nich terminologia jest w znacznym zakresie mu obca.

Prezentowany w niniejszej monografii system automatycznej translacji stanowi bezpośredni dowód na to, że przekład komputerowy o wysokiej jakości, przypominający swą postacią rezultat pracy profesjonalnego tłumacza, jest możliwy w praktyce. Niestety, na obecnym etapie badań ceną, którą musimy za odpowiednio wysoką jakość przekładu zapłacić, jest konieczność drastycznego ograniczenia tematyki tłumaczonych tekstów jedynie do wybranej wąskospecjalistycznej dziedziny. Ponadto tłumaczone dokumenty muszą odznaczać się odpowiednio strukturalizowaną postacią i wysokim stopniem stereotypowości ich treści.

MIROSLAW GAJER

Dedicated Machine Translation System Implementing the Method of Translation Patterns

Summary

The topic of the monographic study at hand constitutes a direct continuation of the author's previous book publication *Multilingual Pattern-Based Machine Translation Systems*. The book presented the basic theoretical concepts underlying the method of Pattern-Based Machine Translation (PBMT) designed by the author. Further, the publication discussed results of broad analyses of potential and limitations involved in the application of the PBMT method for various natural languages used as source or target languages. On the basis of the conducted analyses the PBMT method has been demonstrated to be highly universal; after appropriate modification the method can be applied for doing machine translation between almost any chosen pair of languages. The essence of the PBMT method consists in shifting the translation process from the level of individual words, which is characteristic for all types of Rule-Based Machine Translation (RBMT) systems, to the level of multi-word phrases, which in themselves contain a considerable load of contextual information defining the meaning of words in a phrase. In this way the problem of encountering ambiguity in the translation process has been eliminated to a large extent; ambiguity constitutes one of the basis obstacles in efficient functioning of RBMT systems. The core of the PBMT method consists in drawing directly on techniques of Example-Based Machine Translation (EBMT) where the machine translation system uses experience and translation databases of translators who specialize in translation of certain text types. The PBMT method devised by the author makes it possible to apply EBMT techniques also for inflected languages, such a Polish, since translation patterns contain information about syntactic rules which allows the system to adjust automatically the inflected forms of the translated phrases.

The present monographic study is a complex report on the implementation of the Pattern-Based Machine Translation system. The computer translator discussed in the monographic study is a specialized system, i.e. it is devised for automatic translation of texts of a strictly specified subject matter and of a repetitive and standardized form. Such a choice has been dictated by a vast work load necessary to construct a computer translator for automatic translation of texts dealing with any subject matter of the user's preference. The construction of such a universal system of computer translation will require many years of work of a numerous research team, made up of specialists from various areas; this is an enterprise which can be taken up only by a powerful informatics company. In this context, the construction of a specialized computer translation system is to be seen as, above all, the expression of the author's realistic approach. As a result of the project implementation a computer translation system has been constructed, aimed at automatic translation of reports assessing the current

situation on the global foreign exchange market, forex. The considered reports are characterized by a stereotypical structure, while their vocabulary and phraseology constitute, to a large extent, a finite set. These characteristic features made it possible to work out translation patterns for the realization of automatic translation. As a result, a computer translation system has been constructed which is able to conduct translation – fully automatically and with high precision and fidelity – of reports on the volatility of chosen currency pairs. The reports considered are posted regularly and in large number on relevant websites which justifies the construction of this type of a specialized translation system. Further, as the author's experiments have proved, the devised computer translation system translates the considered reports in many cases more efficiently as translators trainees or average translators who do not specialize in such text types and who are, to a large extent, not familiar with the terminology contained therein.

In conclusion, the automatic translation system presented in the monographic study at hand proves the feasibility of high-quality computer translation whose results resemble the results of professional translators' work. Unfortunately, at the current stage of research, the price for the high quality of translation is the need to drastically limit the subject matter of translated texts to a selected, highly-specialized field only and to focus on text types whose macro- and micro-structure is highly standardized and repetitive.