

# JĘZYKI OGÓLNE, SPECJALISTYCZNE I KONTROLOWANE A EFEKTYWNOŚĆ TŁUMACZENIA MASZYNOWEGO

Karolina Stefaniak

*Autorka jest doktorantką w Zakładzie Komunikacji Językowej Instytutu Lingwistyki Stosowanej Uniwersytetu Warszawskiego. Zajmuje się również działalnością translatorską. Jej zainteresowania badawcze obejmują wykorzystanie tłumaczenia maszynowego w praktyce tłumaczenia oraz tłumaczenie tekstów specjalistycznych, a także zagadnienia komunikacji specjalista-laik, lingwistyki korpusowej i analizy dyskursu. Przygotowuje pracę doktorską na temat rozumienia dyskursu medycznego w komunikacji lekarz-pacjent.*

Celem artykułu jest określenie cech, jakimi powinien charakteryzować się tekst, aby do jego tłumaczenia można było wykorzystać system tłumaczenia maszynowego, oczekując jednocześnie wysokiej efektywności tego procesu. Efektywność tłumaczenia maszynowego zależy bowiem po pierwsze od cech systemu tłumaczeniowego, a po drugie od cech tekstu (Reinke 1997:102). Efektywność tłumaczenia maszynowego rozumiana będzie tutaj jako wzrost szybkości procesu tłumaczenia i/lub wzrost jakości produktu tłumaczenia.

Przez termin „tłumaczenie maszynowe” oraz termin pokrewny „tłumaczenie wspomagane maszynowo” rozumie się najczęściej każdy system, który przetwarza tekst w języku źródłowym na tekst w języku docelowym. Jest to definicja bardzo niejednoznaczna, nie precyzuje bowiem ani stopnia interakcji między człowiekiem a maszyną, ani kręgu użytkowników tego systemu. Nie określa również, czy „tłumaczenie” oznacza proces czy też produkt otrzymywany w wyniku tego procesu. Prowadzić to może do wielu nieporozumień, zwłaszcza kiedy w grę wchodzi ocena możliwości i ograniczeń tłumaczenia maszynowego. Na potrzeby niniejszego artykułu termin „tłumaczenie maszynowe” zostanie więc zawężony do tłumaczenia (produktu) wykonanego przez tłumacza przy użyciu systemu działającego w oparciu o pamięć tłumaczeniową, czyli tzw. zintegrowanego systemu tłumaczenia maszynowego.

Zintegrowane systemy tłumaczenia maszynowego (*translator's workstations*) pojawiły się pod koniec lat 80-tych. Podstawowym elementem tych systemów jest pamięć tłumaczeniowa (*translation memory*). Oprócz tego zawierają one programy do zarządzania terminologią i edytory tekstu oraz opcjonalnie elektroniczne słowniki, terminologiczne bazy danych i systemy całkowicie automatycznego tłumaczenia maszynowego. Właśnie połączenie wszystkich tych narzędzi w jednym programie i pojawienie się komputerów osobistych, w których program ten można było stosować, stało się przełomem w wykorzystywaniu tłumaczenia maszynowego przez tłumaczy.

Pamięć tłumaczeniowa jest rodzajem bazy danych, w której przechowywane są teksty źródłowe i ich tłumaczenia w postaci zsynchronizowanych ze sobą jednostek tłumaczeniowych (najczęściej zdań), tak że jednej jednostce tekstu źródłowego odpowiada jedna jednostka tekstu tłumaczenia. Proces synchronizowania tekstów i ich tłumaczeń określany jest jako *alignment*. Większość zintegrowanych systemów tłumaczenia maszynowego umożliwia jego automatyczne przeprowadzenie, niemniej jednak niezbędna jest też zazwyczaj ręczna korekta wykonana przez tłumacza. Tłumaczenie nowego tekstu odbywa się w następujący sposób: najpierw program dzieli tekst w języku źródłowym na jednostki tłumaczeniowe, które następnie porównuje z jednostkami tłumaczeniowymi w języku źródłowym zgromadzonymi w pamięci tłumaczeniowej i określa, które są identyczne

(*exact match*) bądź podobne (*fuzzy match*). Następnie przywołane zostają wcześniejsze tłumaczenia znalezionych jednostek tłumaczeniowych, ale decyzja, czy tłumaczenie danej jednostki włączyć do nowego tłumaczenia podejmowana jest każdorazowo przez tłumacza. Tym sposobem tłumacz może skupić się na czynnościach kreatywnych, a uwolnić od rutynowych i powtarzających się aspektów procesu tłumaczenia, zachowując nad nim jednocześnie pełną kontrolę.

Efektywność tłumaczenia maszynowego, nawet przy zastosowaniu zintegrowanych systemów tłumaczenia maszynowego może różnić się zależnie od cech przetwarzanego materiału językowego. Z praktycznych więc powodów konieczne jest ustalenie kryteriów wyboru tekstów nadających się do tłumaczenia maszynowego, tj. gwarantujących wzrost efektywności pracy tłumacza. Sager (1994:291) wymienia następujące kryteria:

- tekst źródłowy dostarczony w formacie elektronicznym
- terminologia tekstu źródłowego zawarta w słownikach systemu
- stylistyczna prostota tekstu
- poprawność gramatyczna i ortograficzna tekstu
- „pełność” tekstu źródłowego, tj. brak elips i ominięć.

Dla Haverkorta (1991:11) jakość tłumaczenia maszynowego jest tym wyższa, im prostsze i bardziej jednoznaczne z leksykalnego i syntaktycznego punktu widzenia są zdania, z jakich zbudowany jest tekst źródłowy; im bardziej ograniczona jest składnia i słownictwo tekstów źródłowych, jak np. w dokumentacjach technicznych. Bernhard (1994:51) podkreśla, że do tłumaczenia maszynowego nie nadają się teksty dowolne, ale tylko terminologicznie jednorodne, jednoznacznie sformułowane i względnie ubogie pod względem stylistycznym teksty specjalistyczne i techniczne. Reinke (1997:103) wprowadza kryterium rekurencji zewnętrznej i wewnętrznej. Im bardziej w obrębie tekstu przeznaczonego do tłumaczenia powtarzają się zdania, frazy bądź dłuższe fragmenty tekstu (wewnętrzna rekurencja), oraz im więcej materiałów referencyjnych posiada tłumacz (zewnętrzna rekurencja), tym większej efektywności tłumaczenia maszynowego można się spodziewać. Także tłumaczenie tekstów napisanych według określonych wskazówek stylistycznych może podnieść efektywność tłumaczenia.

Tak więc tekst optymalny z punktu widzenia tłumaczenia maszynowego, czyli gwarantujący jego największą efektywność, powinien być w formie elektronicznej, cechować się jednorodnym bądź ograniczonym słownictwem i strukturami gramatycznymi, być jednoznaczny, odpowiadać normom poprawności i zawierać powtarzające się elementy leksykalne i/lub gramatyczne. Dostępność materiałów referencyjnych, czyli stworzenie bądź importowanie pamięci tłumaczeniowej, oraz odpowiednie skonfigurowanie słowników i terminologicznych baz danych, nie należą do cech tekstu i stanowią raczej działanie poprzedzające właściwy proces tłumaczenia i wykonywane niezależnie od niego.

W warunkach rzeczywistej sytuacji tłumaczeniowej teksty można podzielić na mówione i pisane, a wśród tekstów pisanych wyróżnić teksty literackie i informatywne. Teksty informatywne dzielą się dalej na ogólne i specjalistyczne (Grucza 1981:14). W tłumaczeniu maszynowym w rozumieniu niniejszego artykułu liczą się jedynie teksty pisane informatywne, gdyż literackie nigdy nie są ani jednorodne, ani jednoznaczne. Istotne wydaje się natomiast porównanie cech tekstu optymalnego z punktu widzenia tłumaczenia maszynowego z cechami tekstów ogólnych i tekstów specjalistycznych. Cechy te z kolei należy umieścić na szerszym tle cech języka naturalnego i sztucznego.

Język naturalny to twór powstały jako wynik historyczno-kulturowego rozwoju określonej wspólnoty ludzkiej (Lukszyn 1998). Cechuje się przede wszystkim maksymalną formalną wolnością na wszystkich poziomach wyrażania, otwartością na zmiany, elastycznością. Wolność języka naturalnego polega na braku zewnętrznych kodyfikacji i ograniczona jest jedynie wzajemnym zrozumieniem. Otwartość na zmiany pozwala odzwierciedlać nowe doświadczenia użytkowników danego języka. Elastyczność języka naturalnego czyni go zaś zdolnym do parafrazy i dlatego może być on używany jako metajęzyk. Język naturalny ma nieograniczone funkcje, dlatego jest niezastąpiony do wyrażania znaczeń emotywnych, estetycznych, społecznych i fatycznych. W języku naturalnym wyrazy są wieloznaczne, bazuje on bowiem na polisemii i homonimii, których potrzebuje jako bodźca do rozwoju. Jeżeli chodzi o reguły, język naturalny rozróżnia między językowymi regułami a normami społecznymi (Sager 1994).

Po drugiej stronie ograniczeń formalnych znajduje się język sztuczny. Sager (1994:34) definiuje go jako język, którego reguły zostały określone przed jego powstaniem. Język sztuczny stanowi stały repertuar środków językowych i reguł ich tworzenia i łączenia, które są zawsze takie same i mogą być zmienione lub dodane tylko w wyniku celowego działania. Tworzenie języka sztucznego może być więc rozumiane jako nakładanie ograniczeń na język naturalny, przy czym język naturalny traci swoje charakterystyczne właściwości. Język sztuczny nie jest elastyczny, tylko jednoznaczny, ogranicza polisemię i homonimie, a element leksykalny ma zazwyczaj tylko jedno znaczenie. Dlatego w przeciwieństwie do języka naturalnego nie może być używany jako metajęzyk. Poza tym język sztuczny ma zazwyczaj jedną funkcję, tworzony jest do celów opisowych, klasyfikujących lub informacyjnych.

Język sztuczny jest więc węższy pod każdym względem od języka naturalnego, z którego powstał, i właśnie te ograniczenia decydują o jego nadrzędności w kontekście tłumaczenia maszynowego. Koncepcja języka naturalnego i sztucznego zakłada ewolucję języka, tj. przejście od jednego typu do drugiego wraz ze wzrostem określonych ograniczeń. Motywacja i ewolucja tych ograniczeń wyjaśnia rozwój wariantów języka, z których dla tłumaczenia i tłumaczenia maszynowego najważniejsze są język ogólny (*general language*), języki specjalistyczne (*languages for special purposes, LSP*) i języki kontrolowane (*controlled languages*).

Język ogólny używany jest w codziennej niespecjalistycznej komunikacji w obrębie danej wspólnoty językowej. Język ogólny cechuje się minimalnym wzrostem ograniczeń w stosunku do języka naturalnego. Podstawowymi jego cechami jest otwartość i dynamizm, które pozwalają mu zachować elastyczność i wrażliwość na zmiany społeczne, funkcjonować jako łącznik między wszystkimi członkami danej wspólnoty językowej i stanowić podstawę dalszych ograniczeń na potrzeby komunikacyjne określonych grup w obrębie wspólnoty. (Sager 1994).

Język ogólny obejmuje słownictwo odzwierciedlające podstawowe elementy świata danej wspólnoty ludzkiej, ważne dla wszystkich jej członków, w przeciwieństwie do języków specjalistycznych, które odzwierciedlają wiedzę specjalistyczną członków grupy w ramach jakiejś wspólnoty zajmującej się tym samym rodzajem pracy (Grucza 2002). Dlatego też języki specjalistyczne są „komplementarnymi składnikami, albo lepiej: uzupełnieniami (...) części podstawowych czy rdzennych” etnolektów (Grucza 1994:14). Cechami uniwersalnymi języków specjalistycznych są: występowanie złożonych konstrukcji syntaktycznych, uproszczenie systemu gramatycznego względem bazowego języka ogólnego oraz neutralność stylistyczna. Ponadto do typowych cech języków specjalistycznych należą też otwartość na

zapożyczenia z języków światowych oraz ogólna tendencja do umiędzynarodowienia systemów terminologicznych i orientacja na źródłosłów języków klasycznych przy tworzeniu nowych terminów. Spójność semantyczna tekstów specjalistycznych ujawnia się natomiast między innymi w zjawisku wielokrotnego powtarzania tych samych terminów (Lukszyn 2002). Jednoznaczność semantyczna nie jest natomiast wyróżnikiem języków specjalistycznych, które mogą, ale nie muszą charakteryzować się znaczeniową ścisłością składających się na nie elementów leksykalnych (Grucza 2002).

Większość języków specjalistycznych to języki naturalne, z nałożonymi pewnymi ograniczeniami w kierunku języka sztucznego. Do języka sztucznego można natomiast zaliczyć już języki kontrolowane, derywowane z języka naturalnego, ogólnego lub specjalistycznego, posiadają one bowiem sztywne reguły dopuszczalnych słów i wyrażeń oraz konstrukcji gramatycznych. Tworzenie tekstów w języku kontrolowanym polega na trzymaniu się ustalonych wytycznych stylistycznych i korzystaniu z list słownictwa. Możliwe są przy tym dwa sposoby postępowania: określenie listy dopuszczalnego słownictwa i konstrukcji gramatycznych (lista pozytywna) lub listy wykluczonych słów i wyrażeń (lista negatywna). Używanie standaryzowanej, jednorodnej terminologii ma przynajmniej dwie zalety: zwiększa jednorodność leksykalną tekstów, a przez to ich zrozumiałość, i ułatwia ich późniejsze tłumaczenie. Dlatego języki kontrolowane konstruowane są przede wszystkim z myślą o automatycznej korekcie tekstów oraz tworzeniu tekstów i tłumaczeń w sposób standaryzowany i automatyczny (Mügge 2002).

Porównując cechy języków naturalnych i sztucznych, a w ich obrębie języka ogólnego, języków specjalistycznych i kontrolowanych, z cechami tekstów optymalnych z punktu widzenia tłumaczenia maszynowego (jednorodne i/lub ograniczone słownictwo i struktury gramatyczne, jednoznaczność, poprawność, powtarzalność) łatwo zauważyć, iż efektywność tłumaczenia maszynowego jest tym większa, im więcej cech języka sztucznego ma tekst źródłowy. W ten sposób teksty ogólne nie spełniają żadnego z tych kryteriów, teksty specjalistyczne nie zawsze spełniają warunek jednorodności i jednoznaczności leksykalnej, wszystkie cechy tekstów idealnych posiadają natomiast konstruowane między innymi dla celów tłumaczenia maszynowego specjalistyczne języki kontrolowane. Stanowią jednak niewielki procent wszystkich tekstów występujących w rzeczywistych sytuacjach tłumaczeniowych, gdyż tworzone są zwykle na wewnętrzne potrzeby konkretnego przedsiębiorstwa (np. Caterpillar) bądź organizacji (np. AECMA Simplified English).

Z praktycznego punktu widzenia największą efektywność tłumaczenia maszynowego zapewniają teksty specjalistyczne. Otwartą i wymagającą szczegółowych badań kwestią jest natomiast efektywność tłumaczenia maszynowego konkretnych rodzajów tekstów specjalistycznych z określonych dziedzin i w określonych językach. Nie istnieje bowiem jeden uniwersalny język specjalistyczny, lecz jest on różny w różnych językach; język specjalistyczny nie jest też jednorodny, lecz różny w zależności od dziedziny; a i w ramach jednej dziedziny występują zróżnicowania w zależności od typu tekstu (Spillner 1983).

## **Bibliografia**

Bernhard, U., 1994, Maschinelle Übersetzung in der Praxis, s. 49-52, *Lebende Sprachen*, nr 2/1994

Grucza, F., 1981, Zagadnienia translatoryki, s. 9-29, w: Franciszek Grucza (red.): *Glottodydaktyka a translatoryka. Materiały z IV Sympozjum zorganizowanego przez Instytut Lingwistyki Stosowanej UW. Jachranka, 3 – 5.11.1976*. Warszawa: Wydawnictwo UW

Grucza, F., 1994, O językach specjalistycznych (technolektach) jako pewnych składnikach rzeczywistych

- języków ludzkich, s. 7-29, w: Grucza F., Z. Kozłowska (red.): *Języki specjalistyczne*. Warszawa: Akapit
- Grucza, F., 2002, Języki specjalistyczne – indykatory i/lub determinanty rozwoju cywilizacyjnego, s. 9-26, w: Jan Lewandowski (red.): *Języki specjalistyczne. Problemy technolingwistyki*. Warszawa: Poligrafia
- Haverkort, K., 1991, Was Übersetzer schon immer über maschinelle Übersetzung wissen wollten, sich aber nicht zu fragen trauten, s. 8-12, *Lebende Sprachen*, nr 1/1991
- Lukszyn, J. (red.), 2002, *Tezaurus terminologii translatorycznej*. Warszawa: Wydawnictwo Naukowe PWN
- Lukszyn, J., 2002, Uniwersalia tekstów specjalistycznych, s. 41-48, w: Jan Lewandowski (red.): *Języki Specjalistyczne. Problemy technolingwistyki*. Warszawa: Poligrafia
- Mügge, U., 2002, Möglichkeiten für das Realisieren einer einfachen Kontrollierten Sprache, s. 110-114, *Lebende Sprachen*, nr 3/2002
- Reinke, U., 1996, Der rechnergestützte Übersetzungsarbeitsplatz im Wandel. Von isolierten Einzelsystemen zur integrierten „Translator's Workstation“, s. 169-185, w: Angelika Lauer, Heidrun Gerzymisch-Arbogast, Johann Haller, Erich Steiner (red.): *Übersetzungswissenschaft im Umbruch. Festschrift für Wolfram Wills zum 70. Geburtstag*. Tübingen: Günter Narr Verlag
- Reinke, U., 1997, Integrierte Übersetzungssysteme. Betrachtungen zu Übersetzungsprozess, Übersetzungsproduktivität, Übersetzungsqualität und Arbeitssituation, s. 97-106, *Lebende Sprachen*, nr 3/1997
- Reinke, U., 1994, Zur Leistungsfähigkeit integrierter Übersetzungssysteme, s. 97-104, *Lebende Sprachen*, nr 3/1994
- Sager, J. C., 1994, *Language Engineering and Translation. Consequences of Automation*. Amsterdam/Philadelphia: John Benjamins Publishing Company
- Schmitt, P. A., 1996, Computereinsatz in der Translation, s. 187-195, w: Angelika Lauer, Heidrun Gerzymisch-Arbogast, Johann Haller, Erich Steine (red.): *Übersetzungswissenschaft im Umbruch. Festschrift für Wolfram Wills zum 70. Geburtstag*. Tübingen: Günter Narr Verlag
- Spillner, B., 1983, Methodische Aufgaben der Fachsprachenforschung und ihre Konsequenzen für den Fachsprachenunterricht, w: Heinrich P. Kelz (red.): *Fachsprache. Band 1. Sprachanalyse und Vermittlungsmethoden*. Bonn: Dümmler Verlag

## SUMMARY

The article deals with the question of the efficiency of machine translation, seen as depending on the features of the machine translation system on the one hand and on the qualities of the text to be translated on the other hand. The author focuses on the qualities the text to be machine translated must possess to ensure a high efficiency of the process: homogenous vocabulary, simple syntactical structures, no ambiguity, no errors and a high level of recurrence. The qualities of an optimal text to be machine translated are then compared against the qualities of general language texts, language for special purposes texts and controlled language texts. LSP texts are found to be best suited for machine translation and further research possibilities in this area are indicated.