LUBELSKIE MATERIAŁY NEOFILOLOGICZNE • 1976

Grażyna Kryszczuk

Granice i możliwości automatycznego tłumaczenia tekstów

Pierwsza koncepcja przekładu maszynowego wyszła od radzieckiego wynalazcy Trojańskiego już w r.1933, kiedy opatentował on maszynę do wykonywania przekładów. W tym czasie jednak kiedy nie istniały jeszcze elektronowe maszyny cyfrowe, nie można było mówić o opłacalnym sposobie wykonywania tłumaczeń przy pomocy maszyny. Dopiero w czasie II wojny światowej, dwaj amerykańscy uczeni A.Weawer i A.Boot, w oparciu o maszyny elektronowe, służące do rozszyfrowania tajnych meldunków nieprzyjaciela, skonstruowali urządzenie, przy pomocy którego można było zrealizować przekład z jednego języka naturalnego na inny. Pierwszy doświadczalny przekład maszynowy zrealizowany został w Stanach Zjednoczonych w r.1954.

Od tego czasu okres badań nad przekładem automatycznym znacznie się rozszerzył. Obecnie prowadzi się prace nad zbudowaniem urządzeń nie tylko tłumaczących, ale jednocześnie czytających i streszczających.

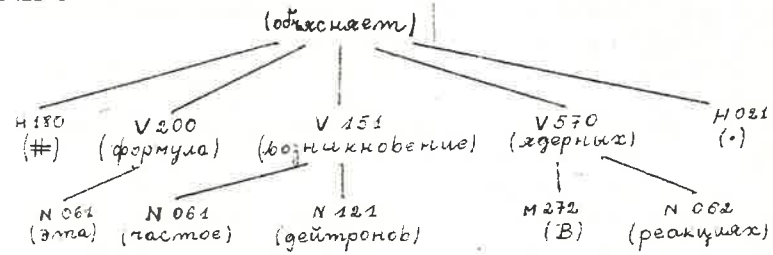
Nauką bardzo pomocną przy rozwiązywaniu problemów tłumaczenia maszynowego okazała się lingwistyka strukturalna, która bada język metodami matematycznymi, tzn. pomija materialną stronę znaku językowego oraz sens wyrażeń językowych, zakładając, że tekst to tylko szereg znaków uporządkowanych według określonych praw. Lingwistyka strukturalna traktuje więc język jako strukturę zorganizowaną, dającą się ściśle określić.

Matematyczne metody badania języka polegają na sztucznym konstruowaniu tzw. modeli językowych. Dzięki takim modelom moż-

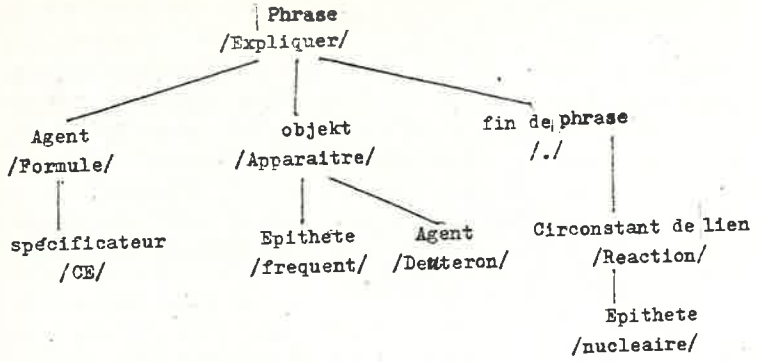
na łatwiej udzielić odpowiedzi na pytanie: jak można mając dane jednostki języka stworzyć za pomocą skończonej ilości formalnych reguł tekst w danym języku.

Istnieje kilka modeli języka, których twórcą jest N. Chomsky. Pierwsze modele zbudowane przez Chomsky'ego są jednak niewystarczające i niedoskonałe, ponieważ machanizm tworzenia zdań nie jest związany z logiczną i gramatyczną strukturą zdań. Modele te są niedoskonałe również z tego względu, że wymagają budowy ogromnej ilości schematów. Modele językowe próbowano jednak z powodzeniem wykorzystać w badaniach poświeconych tłumaczeniom automatycznym. W roku 1967 w Centrum Tłumaczenia Automatycznego w Grenoble, odbyła się konferencja poświęcona tłumaczeniom maszynowym. Na konferencji tej, B.Vauquis, posługując się serią modeli językowych, zrealizował przekład z języka rosyjskiego na język francuski. Do maszyny IBM 7044 wprowadzono rosyjskie zdanie z podrecznika fizyki atomowej, zakodowane przez francuską operatorkę nie znającą języka rosyjskiego,rozróżniającą tylko znaki alfabetu rosyjskiego. Jednym z tłumaczonych zdań było nastepujące zdanie rosyjskie: „Эта формула объясняет частое возникновение в ядерных реакциях.
Dla zdania ustalono 250 reguł składniowych, z których każda przypisana była poszczególnym słowom, a zdanie przedstawione zostało w formie następującego drzewa genealogicznego

Model M 2



Po zastosowaniu reguł składniowych otrzymano następujące zdanie francuskie: "Cette formule explique que dans les réactions nucléaires les denterons apparaissent fréquement", które także można przedstawić w formie drzewa genealogicznego:



W pierwszych eksperymetalnych tłumaczeniach maszynowych próbowano metody najprostszej - przekładu dosłownego, który potem miał być częściowo korygowany według reguł gramatycznych. Realizacja takiego tłumaczenia wyglądała następująco:do pamięci maszyny wprowadza się dwa słowniki:, każde słowo jednego języka otrzymuje adres równoważnika danego słowa w drugim języku i na wyjściu drukuje odpowiednik dosłowny.Trudności występujące przy tego rodzaju tłumaczeniach są oczywiste. Przede wszystkim słowa danego języka nie są jednoznaczne, a więc jednemu słowu może odpowiadać kilka przekładów. Wielo - znaczność powstaje wtedy, gdy jedno słowo w języku A, ma kilka odpowiedników w języku B, a każdy z tych odpowiedników jest równoznaczny z innym słowem. Zależnie od kontekstu polskie słowo - "wydawać" może mieć 20 odpowiedników w języku niemieckim:

1. Werk - herausgeben
2. Geld - ausgeben
3. Munition - verteilen
4. Waren - ausliefern
5. Manifest - ergehen lassen
6. Stadt - der Plünderung preisgeben
7. Befehl - erlassen

itd.

Język niemiecki zna także polisemię. Słowo "Hieb" ma jedenaś-

cie odpowiedników w języku polskim:
1. uderzenie, cios, cięcie
2. raz, cios, pchnięcie białą bronią
3. raz, uderzenie, baty itd.

Zjawisko polisemii jest bardzo częste. Kalifornijski badacz Kenneth E.Harper oznaczył liczbę słów wieloznacznych w rosyjskim tekście naukowym. Każde słowo rosyjskie miało co najmniej dwa różne angielskie odpowiedniki: W takich wypadkach maszyna podaje na wyjściu wszystkie możliwe znaczenia, co z kolei wymaga dalszego opracowywania tekstu, tzn.: postredagowania. Jako przykład posłużyć może grupa wyrazów, która odpowiada polskiemu słowu "rodzeństwo"

język znaczenie	j.węgierski	j.niemiecki	j.polski	j.malajski
starszy brat	bārya	Bruder	brat	sudarā
młodszy brat	öccs			
starsza siostra	néne	Schwester	siostra	
młodsza siostra	húg			

W języku polskim słowo "brat" oznacza zarówno starszego jak i młodszego brata i używane jest w znaczeniu "syn tych samych rodziców. W języku malajskim słowo "sudarā" odnosi się do brata i siostry i oznacza "dziecko tych samych rodziców"

Przekład dosłowny nie uwzględnia także wyrażeń idiomatycznych. Tak więc angielskie "heat dog" /bułka z gorącą parówką/, w przekładzie dosłownym będzie brzmiało "gorący pies".

Przy realizacji przekładu dosłownego występują także trudności z szykiem zdań. Np. język angielski i niemiecki mają ściśle określoną strukturę i szyk zdania, podczas gdy języki słowniańskie mają szyk dość dowolny. Dlatego też nie można tłumaczyć słowa po słowie, lecz należy tłumaczyć całe konstrukcje zdaniowe.

Powyższe trudności można usunąć, wykonując tzw. preredagowanie tekstu.

Ponieważ jednak preredagowanie tekstu nie rozwiązało problemu tłumaczenia automatycznego, rozpoczęto próby stworzenia specjalnej gramatyki, ujętej w precyzyjne reguły. Przedstawienie gramatyki w postaci ścisłych reguł matematycznych pozwoliłoby na tłumaczenie obcego tekstu w sposób całkowicie automatyczny, tylko przy pomocy słownika. Matematyczne ujęcie całości gramatyki, nie jest jednak jeszcze w chwili obecnej możliwe. W tej dziedzinie badań poczyniono dopiero pierwsze kroki.

Jedną z podstawowych trudności przy realizacji tłumaczenia maszynowego jest istnienie leksykalno-gramatycznej homonimii. Homonimy utrudniają jednoznaczne określenie słowa i dlatego też, aby przeprowadzić analizę zdania, należy najpierw zlikwidować homonimię. Przy realizacji tłumaczenia automatycznego zdanie języka wyjściowego /np. angielskiego/ powinno być rozłożone na formy elementarne, a następnie z odpowiadających im form języka wyjściowego, należy tworzyć tłumaczone zdanie. Takiej analizy i syntezy zdania można dokonać tylko wtedy, gdy wiadomo, do jakiej klasy słów należy każde ze słów zdania tłumaczonego. W języku angielskim istnieje wiele leksykalno- gramatycznych homonimów, tzn. form słownych zewnętrznie pokrywających się, lecz przynależnych do różnych klas gramatycznych.

Pierwszy etap przy rozróżnianiu homonimii polega na wykorzystaniu danych morfologii. Przynależność danych słów do określonych klas można ustalić jednoznacznie na podstawie ich końcówek. Aby odróżnić homonimy według cech morfologicznych,należy wydzielić końcówki, które są charakterystyczne dla każdej klasy, a nie są możliwe dla innej klasy. Jeśli w badanej formie słownej można ustalić końcówkę charakterystyczną dla określonej klasy słów, to w ten sposób zostaje określona jej przynależność do tej klasy. W języku angielskim, w wypadku homonimii czasownik - przymiotnik, czasownik określa się wedle końcówek -s, -ed, - ing, natomiast przymiotnik - według końcówek -er, -est. Podobnie rozróżnia się homonimy typu: rzeczownik - przymiotnik, rzeczownik - czasownik. Istnieją jednak przypadki, kiedy według cech morfologicznych nie udaje się ustalić przynależności słowa do określonej klasy; np. w języku angielskim końcówka - s nie jest charakterystyczna ani dla rzeczownika, ani dla czasownika/. W takim wypadku rozpatrzyć należy kombinacje, które słowo tworzy z sąsiednimi słowami w zdaniu. Jeśli słowo tworzy kombinację charakterystyczną dla jednej

klasy słów /np. dla rzeczownika/ a nie dla innej/np. dla czasownika/ to uważamy to słowo za rzeczownik.

Po rozpatrzeniu zasad morfologicznych bada się przypadki, gdy w zdaniu istnieje obok siebie kilka słów posiadających homonimy. W zdaniu We can see that the solution is inique - see może być czasownikiem zwykłego typu i czasownikiem używanym w konstrukcji typu I see the boy run - that może być spójnikiem lub zaimkiem wskazującym.

Może się zdarzyć, że zaistnieją przypadki jednostkowej homonimii. Należy je wtedy rozpatrywać oddzielnie, aby nie przedłużać dalszych operacji. W informacji do rdzenia słowa znajduje się zawsze wskazówka, czy jest ono rzeczownikiem,czy przymiotnikiem. Dla każdego przypadku homonimii układa się specjalną tabelę, która zawiera wykaz charakterystycznych kombinacji dla analizowanych homonimów.

Przy realizacji tłumaczenia maszynowego konieczne jest rozwiązanie problemu idiomów występujących w tekście.Problem idiomów jest łatwiejszy do rozwiązania niż problem homoninów. Rozwiązuje się go przez budowanie słowników idiomów.Najpierw zestawia się słowniki wszystkich tematów. Słowa głównego.Do każdego tematu słowa dodana jest informacja o bieżącym numerze tematu danego słowa w słowniku.

Jako słowo "główne" wybiera się takie słowa z tekstu, że przy wystąpieniu tego słowa istnieje największe prawdopodobieństwo, że mamy do czynienia z idiomem.

Słownik idiomów opracowuje się po ukończeniu analizy morfologicznej.tekstu. Każdy idiom zostaje opatrzony odpowiednią informacją. Przy tłumaczeniu, każde słowo oznaczone tą informacją jest sprawdzone w słowniku idiomów. Słownik taki składa się z dwóch części:

a/ część identyfikacyjna,

b/ część informacyjna.

Zestawienie takiego słownika idiomów jest konieczne, ponieważ opracowanie idiomów przeprowadza się przed analizą syntaktyczną. Gdyby pominięto opracowanie idiomów, to analiza syntaktyczna byłaby bardzo utrudniona, a w niektórych przypadkach nawet niemożliwa.

Jednym z najbardziej dotychczas udanych sposobów tłumaczenia automatycznego jest przekład zbiorowy polegający na, tym, że maszyna umie przełożyć dowolny z "n" języków na dowolny inny z tych języków. Tłumaczenie takie odbywa się za pomocą tzw. języka pośredniego. Każdy tekst z dowolnego języka "n" maszyna tłumaczy na język pośredni i z języka pośredniego, na każdy inny dowolny język "n". Przekład polega na tym że w procesie tłumaczenia na język pośredni, zachowany zostaje tylko sens tłumaczonego zdania, a potem sens ten przekładany jest na inny język. Język pośredni nazywany językiem informacyjnym. Język taki ma ściśle określony słownik i gramatykę. /Słownik jest to spis oznaczeń pojęć danej nauki/

Przekład oznaczeń pojęć według J.Lewina:

"∧"	-	spójnik "i"
"∨"	-	oznacza wyrażenie "lub"
"→"	-	odpowiada wyrażeniu "jeżeli ... to ..."tzn. /A→ B/ /oznacza to "jeśli A to B",lub z A wynika "B"/
"∼"		oznacza "nie"
"⋀"		"dla każdego x"
"⋁"		"istnieje takie "x", że ..." .
d_1, d_2		oznaczają "prosta"
P		oznacza własność "być równoległym do prostej "b"

Posługując się powyższymi symbolami, twierdzenie:"jeżeli dwie proste są równoległe do trzeciej prostej, to są one równoległe do siebie" możemy przedstawić następująco:

$$\bigwedge d_1 \bigwedge d_2 \{\bigvee d_3 \; [P(d_1, d_3) \wedge P(d_2, d_3)] \rightarrow P(d_1, d_2)\}$$

W ten sposób, przy pomocy języka informacyjnego, można zapisać każde twierdzenie geometryczne ponieważ słownik geometrii elementarnej może być tylko wykazem pojęć. Bardziej skomplikowaną sprawą jest utworzenie takiego słownika dla innych dziedzin nauki. W takim przypadku wprowadza się tzw. "mnożniki semantyczne". Są to pojęcia elementarne, składające się na pojęcia bardziej złożone. Pojęcia złożone są iloczynami pojęć elementarnych. J.Lewin uznaje następujące pojęcia za elementarne "przyrząd" /B/, pomiar "/M/", "temperatura" /T/ "ciśnienie" /P/, "natężenie prądu" /I/, "napięcie" /V/, "siła" /Q/ i zbiór pojęć przedstawia w formie iloczynów:

Termometr	-	B M T
manometr	-	B M P
amperomierz	-	B M J

woltomierz - B M V
dynamometr - B M Q

Metoda ta nie jest jednak szczególnie przydatna, ponieważ nie wskazuje na związki między pojęciami; BN można odczytać jako przyrząd do nagrzewania lub jako nagrzewanie przyrządu.

Jedną z najbardziej udanych prób stworzenia języka informacyjnego jest propozycja W.Perry'ego i A.Kenta. Około 15 lat temu stworzyli oni język informacyjny który pozostaje w użyciu do dziś. W słowniku tego języka jest 214 mnożników semantycznych, z których każdy oznaczony jest trzema literami, z przerwą po pierwszej literze:

M - CH - urządzenie
L - CT - elektryczność
M - SR - pomiar

W miejsce przerwy wstawiana jest litera oznaczająca stosunek pojęcia elementarnego do pojęcia złożonego.

Język ten wygodny jest z tego względu, że opracowanie algorytmu takiego tłumaczenia nie jest zbyt skomplikowane.

Opracowanie tekstu dla maszyny, a więc uproszczenie go i wyeliminowanie wieloznaczności - to programowanie. Wszystkie operacje przeprowadzane nad tekstem mają na celu doprowadzenie do ustalonej kolejności jednoznacznych rozkazów dla maszyny, ponieważ tylko tak może ona tekst przetłumaczyć. Układ rozkazów, które określają przebieg procesu tłumaczenia tekstu to algorytm tłumaczenia.

Po opracowaniu algorytmu tekst przekazywany jest programiście, który przedstawia go w formie kolejnych rozkazów dla maszyny. Rozkazy dla maszyny zawsze przedstawia się w systemie ósemkowym, który potem przekodowuje się na system dwójkowy.

Zakodowany program zostaje zapisany na specjalnych taśmach lub kartach perforowanych i wprowadzony do maszyny za pomocą urządzenia wejściowego.

Tłumaczenie maszynowe i problemy z nim związane to zagadnienie nowe i na ogół jeszcze mało znane. Przekładem maszynowym zaczęto zajmować się dopiero po II wojnie światowej, ale mimo tak krótkiej historii, w porównaniu z innymi dyscyplinami wiedzy, nauka o tłumaczeniu maszynowym zrobiła ogromne postępy. Zrealizowanie tłumaczenia całkowicie automatycznie jest bardzo skomplikowane i przy obecnym stanie wiedzy jeszcze niemożliwe. W trakcie eksperymentów wyłaniają się coraz to nowe problemy natury lingwistycznej. Język ludzki jest za mało sformalizowany. Problemy homonimii, synominii, szyku zdania i wyrażeń idiomatycznych, do dnia dzisiejszego nie zostały rozwiązane w takim stopniu, aby przestały być przeszkodą w trakcie realizacji przekładu automatycznego. Już po przeprowadzeniu wstępnych prób tłumaczeń maszynowych okazało się, że na razie można dyskutować tylko na temat tłumaczeń naukowo - technicznych, ze względu na stosunkowo ubogi słownik, brak indywidualnych cech stylu, brak wyrażeń idiomatycznych i mało skomplikowaną składnię zdania.

Pokonanie wszystkich trudności związanych z realizacją tłumaczenia automatycznego jest jeszcze sprawą dalekiej przyszłości, ponieważ warunkiem dokonania tłumaczenia jest dokładne sprecyzowanie gramatyki języków naturalnych, co w obecnym stadium rozwoju teorii przekładu automatycznego nie jest niestety jeszcze możliwe.

Zusammenfassung

Etwa vor 20 Jahren entwickelte sich eine neue Wissenschaft - die Kybernetik. Man erzeugte die ersten elektronischen Maschinen. Gegenwärtig führt man Forschungen über den Bau von Übersetzungsanla - gen. Man führt zur Zeit in fünfzehn Ländern Untersuchungen.

Alle Arbeiten über die maschinelle Übersetzung betreffen nur wissenschaftlich-technische Texte; literarische Texte werden zur Zeit nicht übersetzt. Es liegen hier einige Gründe vor. Vor allem sind Übersetzungen von Fachliteratur besonders nötig. Außerdem ist für wissenschaftlich-technische Texte ein verhältnismäßig armer Wortschatz charakteristisch; sie tragen auch fast keine individuellen Merkmale des Stils. Es ist auch wichtig, daß die Eindeutigkeit der Wörter größer ist; es fehlt an idiomatischen Wendungen. Auch die Syntax ist einfach. Das Problem der maschinellen Übersetzung literarischer Texte, besonders die Übersetzung der Poesie, befindet sich außerhalb des Bereichs der heutigen Erwägungen.

Man diskutiert die Möglichkeit der maschinellen Übersetzung aus

natürlichen Sprachen in die Maschinensprachen. Es handelt sich darum, daß man die natürliche Sprache in Befehle umgestalten kann, die eine Maschine ausführen könnte. Die Sprache, die die Maschine umgestalten kann, muß nach exakten, mathematischen Regeln bearbeitet werden. Die Ausarbeitung einer formalisierten Sprachtheorie erwies sich als notwendig. Bei dieser Theorie muß man die materielle Seite des Sprachzeichens und der sprachlichen Ausdrücke übergehen. Man muß voraussetzen, daß ein Text oder eine Rede nur eine Kette von Zeichen ist, die nach bestimmten Regeln miteinander verbunden sind. Die Sprache als etwas Ganzes muß man als eine bestimmte, organisierte Struktur behandeln.

Die moderne strukturelle Linguistik untersucht die Sprache mit mathematischen Methoden und zwingt zu ihrem Gebrauch. Eine der Richtungen der Sprachuntersuchung mit mathematischen Methoden kann man logisch-mathematische Richtung nennen

Bibliografia

1. O.C.Kułagina, Ob odnom sposobie opriedielienija gra - mmaticzeskich poniatij na bazie tieorii mnozestw: Problemy kibiernietiki" Moskwa 1958 Gosudarstwiennoje Izdatielstwo Fizyko-Matiematiczeskoj Litieratury.
2. H.Lewicka Na drodze do przekładu maszynowego. "Kwartalnik Neofilologiczny" t.15, zeszyt 4, 1968.
3. J.Lewin, J.Gastiew J.Rozanow, Język, matematyka, cybernetyka Warszawa 1967 PWN.
4. D.J.Lewin, G.N.P. Niekrasow Sistiema programirowanija dla zadacz maszinnogo pierewoda. "Problemy Kibiernietili"/ Moskwa 1971, Izdatiel - stwo "Nauka"
5. T.N. Mołosznaja, Woprosy rozliczenija omonimow pri maszinnom poriewodzie s anglijskogo jazyka na russkoj,"Problemy Kibiernietiki, Moskwa 1958.

6. S.Nudel, G.Klimonow, I.Starke, I.Brand Automatische Sprachübersetzung russisch-deutsch.Akademie - Verlag Berlin 1969.
7. A.N.Romanow, G.A.Frotow, Opierator programmist.Osnowy wyczyslitielnoj miechaniki DOSAAF Izdatielstwo Moskwa 1972.
8. J.Ryder, Engineering electronics whith Industrial Applications and Control Mc Craw-Hill Book Company, ING, New York, Toronto, London 1961.
9. A.Wierzbicka O języku dla wszystkich Warszawa 1968 Wiedza Powszechna.
10. A.Ludskanow, Tłumaczy człowiek i maszyna cyfrowa Warszawa 1973 Wyd.Naukowo-Techniczne.