

MASZYNOWE TŁUMACZENIE USTNE  
CZYLI AUTOMATYCZNE TŁUMACZENIE DIALOGÓW  
W CZASIE RZECZYWISTYM

MAŁGORZATA TRYUK

Mówiąc o tłumaczeniu należy rozróżnić dwa główne typy przekładu. Tłumaczenia pierwszego typu dokonuje człowiek, a jego odmianami są: tłumaczenie pisemne (przekształcanie tekstu pisanego w języku A na tekst pisany w języku B), tłumaczenie a vista (przekształcanie tekstu pisanego w języku A na tekst mówiony w języku B), tłumaczenie audiowizualne (w tym tłumaczenie w formie napisów, w którym tekst mówiony w języku A przekształcany jest w tekst pisany w języku B), tłumaczenie ustne (przekształcanie tekstu mówionego w języku A na tekst mówiony w języku B w sposób konsekwentny lub symultaniczny) i tłumaczenie na język migowy (przekształcanie tekstu mówionego w języku A na tekst przedstawiony za pomocą systemu gestów odpowiednich do języka A lub innego języka naturalnego lub odwrotnie). Drugi typ przekładu to tłumaczenie maszynowe rozwijane od ponad pół wieku za pomocą specjalnych programów tłumaczących. Najbardziej znanymi przykładami pisemnego tłumaczenia maszynowego są kanadyjskie TAUM-Meteo (program tłumaczący prognozy meteorologiczne w parze językowej angielski i francuski, opracowany przez ekipę badaczy z Uniwersytetu w Montrealu) oraz wielojęzyczny program SYSTRAN używany m.in. w instytucjach europejskich, w NATO i licznych koncernach międzynarodowych, tłumaczący w parach językowych większości języków urzędowych UE. Programy te mają liczne zalety, z których najważniejsze to szybkość i skuteczność przekładu; mają też liczne wady, m.in. konieczność poddania tekstów tłumaczonych preedycji, a rezultatów tłumaczenia postedycji.

Równocześnie z maszynowym tłumaczeniem pisemnym prowadzi się prace nad ustnym tłumaczeniem maszynowym, czyli nad takim tłumaczeniem dialogów w sposób konsekwentny lub symultaniczny, w którym tłumaczem między uczestniczącymi w nim osobami jest specjalny program komputerowy. Głównym wymaganiami przy takim systemie jest tłumaczenie w czasie rzeczywistym, stąd raczej należy mówić o automatycznym tłumaczeniu dialogów w czasie rzeczywistym.

Maszynowe tłumaczenie ustne, czyli automatyczne tłumaczenie dialogów w czasie rzeczywistym podobne jest do tłumaczenia ustnego dokonywanego przez człowieka. Celem operacji przeprowadzonej przez maszynę jest:

- przekazanie interakcji werbalnych między dwoma rozmówcami,
- rozpoznanie każdej sekwencji w języku wyjściowym oraz ustalenie jej sensu, a następnie przetłumaczenie jej w języku docelowym,
- podejmowanie bezpośredniego dialogu przez program, bez tłumaczenia, z jednym z rozmówców w celu uzyskania niezbędnego wyjaśnienia.

Tego typu tłumaczenie opiera się na przesłance, że komunikacja między człowiekiem i komputerem jest możliwa. Dialog człowieka z maszyną w czasie rzeczywistym może przybierać dwie postaci:

1. mowa ludzka jest przetwarzana na tekst pisemny (ang. speech-to-text communication): ma to miejsce np. w zapytaniach kierowanych do banków danych terminologicznych,
2. tekst pisemny przetwarzany jest w mowę (ang. text-to-speech communication): tak dzieje się np. w poczcie elektronicznej, chatach, itp.

Od dwóch dziesięcioleci do najważniejszych i oczekiwanych zastosowań głosowej komunikacji człowiek – komputer należy automatyczne tłumaczenie foniczne w parze dwu języków w czasie rzeczywistym, które zakłada komunikację głosową człowiek – maszyna (ang. speech-to-speech communication). Oznacza to, że mówcy prowadzący dialog wprowadzają wypowiedź głosem w swoim języku i odbierają wypowiedź rozmówcy przykazywaną w jego mowie i tłumaczoną na ich język.

Już w latach osiemdziesiątych dwudziestego wieku w japońskich ośrodkach naukowych powszechnie było przekonanie, że za dwadzieścia lat każdy będzie posiadał małe kieszonkowe urządzenie, które rozpozna głos, zrozumie zdanie i wypowie je w innym języku. Ta swoista fikcja naukowa okazała się trudniejsza do zrealizowania, niż przewidywali to japońscy naukowcy.

Zanim przedstawię trudności związane z budową pierwszych systemów tłumaczących dialogi, należy wyjaśnić, dlaczego przekazywanie informacji

dialogowych głosem jest interesujące i atrakcyjne tak dla badaczy jak i przyszłych użytkowników. Systemy tłumaczące dialogi w czasie rzeczywistym charakteryzują się następującymi cechami:

- nie wymagają uczenia się specjalnego sformalizowanego kodu,
- przekazywanie informacji głosem może być realizowane znacznie szybciej niż za pomocą klawiatury alfanumerycznej, a przekazywanie informacji do komputera/maszyny za pomocą sygnału mowy umożliwia zwolnienie rąk operatora, co jak się okazuje jest bardzo istotne w mowie, biorąc pod uwagę możliwości wykorzystania elementów niewerbalnych,
- przekazanie informacji głosem może mieć miejsce w różnych nietypowych sytuacjach i pozycjach rozmówców,
- prowadzenie dialogu głosem nie wymaga specjalnego przygotowania ani – treningu, nie wymaga dużego wysiłku, nie stawia fizycznych ograniczeń,
- jedynym wymaganym urządzeniem jest mikrofon,
- jest rozumiane przez ludzi, tj. dialog głosem jest naturalny,
- w końcu ten sposób przekazywania informacji jest kompatybilny z siecią telefoniczną, największą na świecie siecią przekazującą informacje.

W pierwszym etapie badań nad automatycznym tłumaczeniem dialogów zakładano, że aby taki system był uniwersalny, porozumiewanie się w parze języków A i B musi następować za pośrednictwem języka C (por. Basztura 1996). Takie rozwiązanie zostało zastosowane w systemie automatycznego tłumaczenia dialogów VERBMOBIL w jego pierwszej wersji dla pary językowej niemiecki – japoński, gdzie funkcję takiej platformy pośredniczącej pełnił język angielski. W nowszej wersji VERBMOBILU (etap II), którego podstawą jest wielojęzyczność wprowadzono jednocześnie tłumaczenie z języka niemieckiego na angielski i na język japoński. Ponadto w systemie tym występuje dwukierunkowość w przekładzie, tj. system tłumaczy w następujących parach językowych niemiecki – angielski, angielski – niemiecki, japoński – angielski, angielski – japoński.

Zaletą systemów tłumaczących jako wyspecjalizowanych systemów dialogowych jest możliwość rozwiązywania problemów równoległe, bez oczekiwań na satysfakcjonujące w pełni rozwiązania poszczególnych komplementarnych zagadnień.

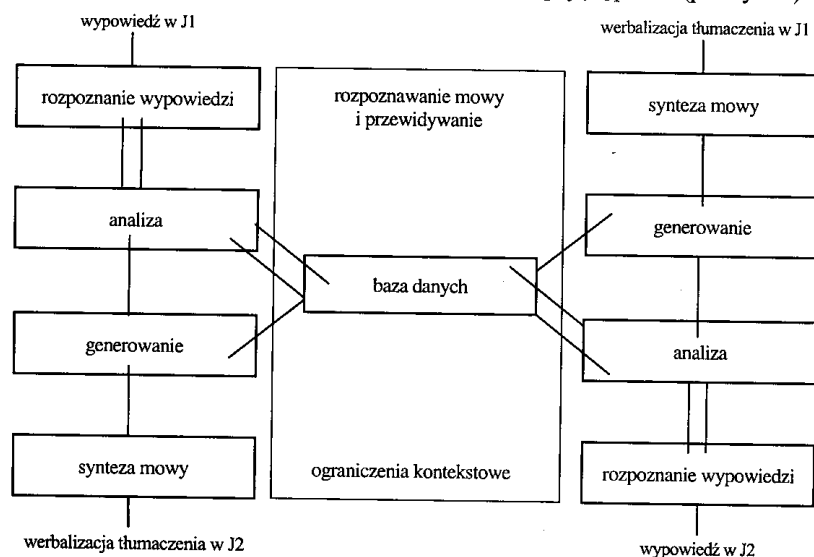
Do głównych i najtrudniejszych zadań przy tworzeniu takiego systemu należy wprowadzenie informacji głosem, a więc rozpoznanie mowy ciągłej, w sposób całkowicie lub w pewnym stopniu niezależny od cech osobniczych głosu rozmówcy.

Główną różnicą między tłumaczeniem ludzkim a automatycznym jest to, że:

- w tłumaczeniu ustnym ludzkim, przekłada się różne wypowiedzi, nawet nieprzewidzianą wymianę zdań lub fragmentów zdań, także poszczególne wyrazy; w tłumaczeniu automatycznym mamy natomiast do czynienia z pełnymi zdaniami, sekwencjami tworzącymi dialog, przy którym rozmówcy muszą być uprzedzeni co do możliwości programu. Tak jest np. w systemie VERBMOBIL – etap I.;
- drugą różnicą jest założenie, że tłumacz człowiek posiada wiedzę ogólną i szczegółową oraz że zna intencje mówcy. Tego nie można powiedzieć o automatycznych systemach tłumaczących.

Jednym z ważniejszych problemów w trakcie tworzenia architektury systemów tłumaczenia dialogów ustnych jest określenie, czy system ten jest odmianą tłumaczenia maszynowego, do którego dodano moduł rozpoznawania mowy, czy też jest odmianą badań nad automatycznym rozpoznaniem i syntezą mowy, do której dołączony jest komponent tłumaczeniowy.

Model systemu maszynowego tłumaczenia dialogów pierwszego typu (tj. pisemne tłumaczenie maszynowe plus moduł analizy i syntezy mowy) zaproponował Kitano (1993). Architektura ta ma następującą postać (por. rys. 1.):



Rys. 1. Architektura systemu tłumaczenia języka mówionego (Kitano 1993, 410)

Zaproponowany model zakłada interakcję między bazą danych i poszczególnymi etapami przetwarzania mowy (rozpoznanie – analiza – generowanie – synteza mowy).

Jak już wspomniano, głównym wymaganiem dotyczącym systemu jest tłumaczenie w czasie rzeczywistym. W celu spełnienia tego i innych wymagań, system musi być wyposażony w szybkie moduły przetwarzania akustycznego, analizy syntaktycznej, semantycznej oraz w moduł sterowania dialogowego zwany komponentem dialogowym (por. Basztura 1996). Ważną rolę spełnia też moduł transferu, który wykorzystuje informacje semantyczne i pragmatyczne w celu sporządzenia mapy odpowiedników między językiem źródłowym i docelowym. Do bardzo ważnych komponentów systemu do tłumaczenia dialogów w czasie rzeczywistym należą komponenty semantyczny i dialogowy. Komponent semantyczny spełnia dwa zadania:

- tworzy reprezentację semantyczną wypowiedzi na podstawie syntaktycznej reprezentacji wypowiedzi wprowadzonej przez rozmówcę, niezależnie od kontekstu interpretacji semantycznej. Taka reprezentacja semantyczna musi być oparta na reprezentacji określonych struktur dialogu i uzupełniona o informacje semantyczne oraz tematycznie zorientowane pojęcia,
- drugą funkcją jest dalsza specyfikacja reprezentacji semantycznej w celu transferu; osiąga się to za pomocą oceny reprezentacji semantycznej w zakresie konceptualnego modelu dialogu wprowadzonego do systemu. Model konceptualny zawiera elementy anaforyczne, poszukuje odpowiedników oraz aktualnia reprezentację.

Zadaniem tak rozumianego komponentu semantycznego jest przetwarzanie pewnych typów fragmentarycznych danych wejściowych.

Najważniejszym elementem systemu jest komponent dialogowy, który podwyższa jego efektywność, dostarcza informacji w celu wspierania procesu tłumaczenia. Informuje o bieżącym stanie dialogu, o dotychczasowej interakcji. W przypadku, gdy informacja potrzebna do poprawnego przetworzenia (tj. tłumaczenia) dialogu jest niekompletna, lub jej brakuje, moduł dialogowy (poprzez bieżącą interakcję z danym rozmówcą) przeprowadza wyjaśnienie dialogów. Ten moduł składa się z trzech podmodułów: modułu statystycznego, modułu skończenie stanowego i modułu planistycznego dialogu. Wszystkie trzy komponenty opierają się na założeniu, że interakcje między modułami dadzą się zmodelować – chodzi tu o poszczególne możliwe sekwencje dialogu. Moduł dialogowy składa się zatem z opisu wszystkich potencjalnych sekwencji dialogu. Moduł statystyczny zawiera zakodowaną informację pochodzącą ze zbioru dialogów z ustalonymi sekwencjami dialogowymi. Dla każdego stanu dialogu moduł dostarcza następujące po nim możliwe teksty wraz z ich prawdopodobień-

stwem występowania Moduł skończenie stanowy dokonuje analizy gramatycznej sekwencji dialogu (dostarczonych przez komponent semantyczny) i sprawdza je pod kątem zgodności z podstawowym modelem dialogu. Moduł ten dostarcza też informacje o tym, jakie następne fazy dialogu według modelu są możliwe. Moduł planistyczny dialogu zawiera reprezentację modelu dialogowego i jest modelem hierarchicznym. Planowanie poprzedza specyficzne fazy dialogu (np. inicjowanie, wyjaśnianie, poprawianie). Jest odpowiedzialny za powiększającą się pamięć dialogu, która zawiera informacje intencjonalne, tematyczne i adnotacyjne dotyczące dialogu, który do tego czasu się odbył. O rzeczywistej efektywności systemu decyduje interakcja między komponentem semantycznym i dialogowym. Komponenty te wchodzi z sobą w interakcję. Różne poziomy informacji dotyczące wypowiedzi początkowej są ujęte w pojedynczej, ustrukturyzowanej reprezentacji. W celu przetworzenia dialogu, system posiada sekwencje działania zwane planami dyskursu, które umożliwiają przewidywanie elementów dialogu. Na podstawie tych przewidywań oraz innych dostępnych danych system tworzy hipotezy co do przebiegu dialogu i które są przetwarzane przez komputer.

System automatycznego tłumaczenia dialogów w czasie rzeczywistym opiera się więc na trzech zasadach:

- jednoczesności lub prawie jednoczesności wypowiedzianego zdania lub sekwencji zdań i tłumaczenia przez system,
- powtarzalności zdań lub sekwencji wypowiedzianych przez rozmówców,
- przewidywalności zdań lub sekwencji występujących w dialogu o określonej tematyce.

Automatyczne tłumaczenie tekstów mówionych w czasie rzeczywistym nie ma na celu zastąpić tłumacza człowieka, lecz znajduje zastosowanie w sytuacjach wysoce skonwencjonalizowanych. Takimi sytuacjami są np. rejestracja uczestników konferencji, rezerwacja pokoju hotelowego, biletów lotniczych, do kina lub do teatru, zamawianie posiłków, wynajem samochodu, umawianie się na spotkanie czy też wizyta u lekarza.

Już na początku badań nad automatycznym systemem tłumaczenia dialogów naukowcy zdali sobie sprawę z ograniczeń, jakie wiążą się z tym typem przekładu, a mianowicie: rozpoznanie mowy nie bierze pod uwagę wszystkich elementów mowy (np. zwroty eliptyczne, intonacja), wiedza językowa, jaką system może obejmować, jest ograniczona, system nie przetwarza wiedzy pozawerbalnej kontekstowej, rozróżnienie elementów nierelevantnych (wahania, wielokrotne zaczynania mowy, wtrącania, powtórzenia) i elementów relevantnych dla mowy nie jest możliwe. Wobec

takich ograniczeń systemowych w celu zbudowania sprawnego i efektywnego urządzenia tłumaczącego należy:

- zdefiniować rodzaj ekwiwalencji, jaka będzie stosowana przez system tłumaczący,
- określić techniki służące w osiągnięciu tej ekwiwalencji,
- zdefiniować dziedziny tematyczne interakcji tłumaczonej przez system,
- opisać przewidywalne struktury dialogu,
- określić przewidywalne elementy językowe, które wystąpią w dialogu,
- wyjaśnić ewentualne dwuznaczności,
- przewidzieć elementy, które trzeba będzie odtworzyć, np. w parze językowej angielski – japoński trzeba będzie odtworzyć aż 70% zwrotów eliptycznych, np. honoryfiki.

Przy każdym typie przekładu tłumacz stoi przed wyborem strategii i techniki tłumaczenia w celu osiągnięcia zakładanej ekwiwalencji, np. formalnej i/lub funkcjonalnej. Automatyczne tłumaczenie dialogów dotyczy jedynie pewnego typu skonwencjonalizowanych dialogów, a więc zdefiniowanie ekwiwalencji musi odnosić się do typu lub tematu dialogu, który ma być przetłumaczony. Oznacza to, że ekwiwalencja (jej typ) musi być z góry określona w zależności od typu dialogu, dziedziny itp. Zazwyczaj tłumacz człowiek wykorzystuje dwie techniki w celu zachowania ekwiwalencji między tekstem źródłowym i tekstem docelowym. Tymi technikami są redukcja i kompensacja. Pierwsza technika zwana również generalizacją lub skróceniem polega na eliminowaniu w tekście docelowym powtórzeń, pleonazmów lub niedociągnięć stylistycznych. Kompensacja zaś jest techniką przekładu polegającą na wyrównywaniu straty informacji powstałej przy tłumaczeniu fragmentu tekstu, inaczej mówiąc jej celem jest odnalezienie brakującego ogniwa. Obie techniki są często stosowane w tłumaczeniu, np.

#### **Redukcja**

zdanie w języku oryginalnym: *Wednesdays or Wednesdays, Fridays or Saturdays are the best for me. I'm something a long lines of. Perhaps Saturday, December eighteenth.*

tłumaczenie: *Wie sieht's Samstag achtzehnten Dezember aus.*

#### **Kompensacja**

zdanie w języku oryginalnym: *Also, Sie brauchen eine Tastatur, die sowohl Deutsch als auch English versteht?*

tłumaczenie: *That means, you need a keyboard that understands both German and English, if I understand you correctly?*

Obie techniki znajdują zastosowanie w tłumaczeniu maszynowym. Pozwalają na realizację bezpośredniej komunikacji interkulturowej między rozmówcami: redundancja jest usuwana a brakujące elementy są dodawane, np. gdy są konieczne dla osiągnięcia celu tłumaczenia (np. grzeczność językowa).

Przykładem najbardziej rozwiniętych i ze sobą konkurujących systemów automatycznych tłumaczenia dialogów w czasie rzeczywistym jest japoński system Interpreting Telephony oraz niemiecki VERBMOBIL.

### 1. INTERPRETING TELEPHONY

System ten jest rozwijany w laboratorium Advanced Telecommunications Research Laboratories ATR na Uniwersytecie w Kyoto w Japonii. Interpreting Telephony jest systemem tłumaczenia dwukierunkowego w czasie rzeczywistym dialogu, który odbywa się przez telefon. Celem komunikacji jest uzyskanie informacji o konferencji międzynarodowej. Tłumaczony dialog prowadzi japońska sekretarka/recepcjonistka i angielskojęzyczny uczestnik konferencji. Przykładowy scenariusz dialogu zawiera następujące sekwencje: angielskojęzyczny przyszły uczestnik konferencji („A”) telefonuje, wygłasza formułę powitalną i przedstawia się po angielsku. Japońska recepcjonistka („J”) odpowiada po japońsku na pozdrowienie. „A” prosi o informacje odnośnie dat, terminów, programu konferencji, programu dla osób towarzyszących, wycieczek itp. „J” przekazuje te informacje, prosi „A” o podanie jego adresu w celu przesłania dokumentacji, itd.

Tak jak w tłumaczeniu ludzkim, system dokonuje trzech operacji:

1. zapewnia interakcję między rozmówcami,
2. umożliwia wysłuchanie i wypowiedanie każdej sekwencji, tak by było możliwe i przetłumaczenie każdej sekwencji,
3. podejmuje rozmowę z jednym z rozmówców w celu wyjaśnienia pewnych elementów wypowiedzi, prowadzi osobny dialog maszyna – człowiek:
  - w celu powtórzenia:  
*what has he said?*
  - dla uzyskania informacji:  
*has Dr. Ebihara answered the phone?,*  
*is she still listening?,*  
*how long have we been talking?*
  - dla uzyskania relewantnej informacji:  
*what time is in Tokio?*
  - aby kontrolować dialog:  
*hang up,*  
*dial 555-1234,*  
*answer*

- dla wyjaśnienia:  
*what was seven nonety-two?*

Komponent dialogowy systemu Interpreting Telephony zawiera 5 odmiennych struktur dialogowych:

1. dialog rozmówca – rozmowca  
W tym module zawarte są różne sekwencje mowy, figury retoryczne.
2. dwa dialogi jednojęzyczne (japońsko-japoński i angielsko-angielski), według schematu:

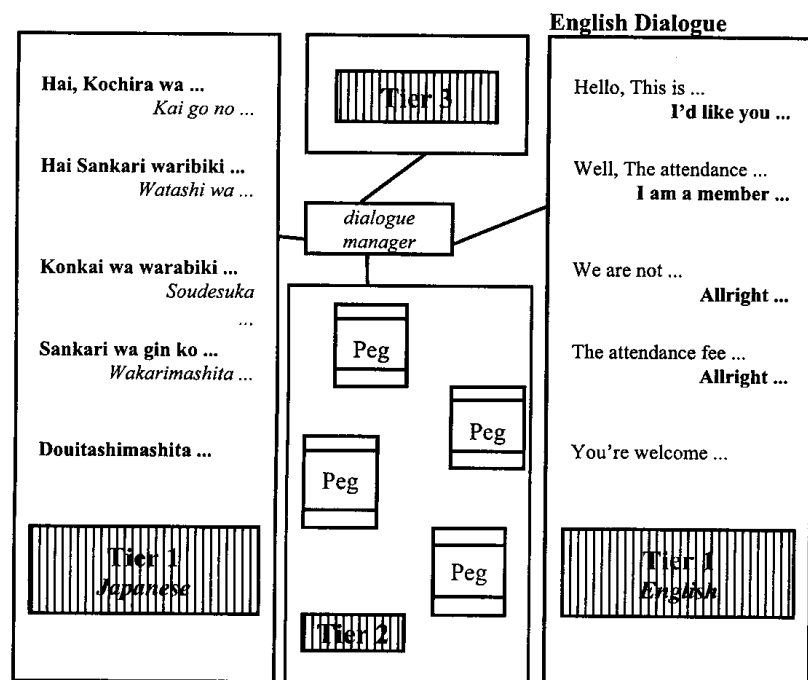
angielski – japoński  
jap. *Hai. Okurimasu (ang. Yes. Send)*  
japoński – angielski  
ang. *OK. I'll send you one*  
angielski

Ten komponent pozwala na otworzenie elementów specyficznych dla danego języka, takich jak np. honoryfiki w języku japońskim czy anaforyki w języku angielskim.

3. dialog człowiek – maszyna

System Interpreting Telephony zawiera w sumie trzy komponenty: moduł 1 obejmuje dwie jednojęzyczne reprezentacje dialogów w języku japońskim i angielskim, moduł 2 jest modułem dialogowym, a moduł 3 stanowi bazę danych odnoszących się do sytuacji związanej z wybranym tematem dialogu.

Architekturę Interpreting Telephony przedstawia rys. 2.



Rys. 2. Architektura systemu Interpreting Telephony (LuperFoy 1996, 240)

## 2. VERBMOBIL

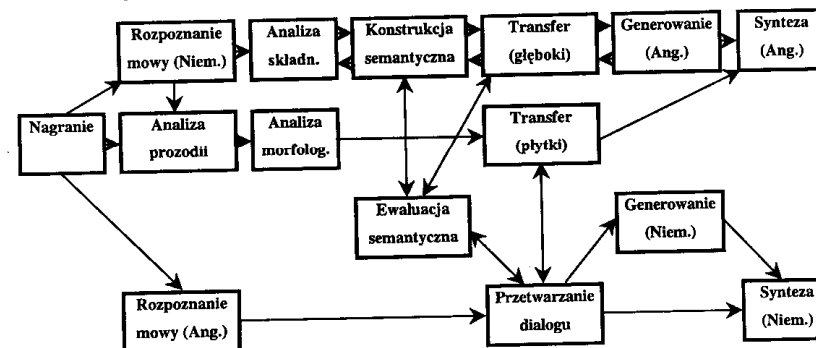
System ten był rozwijany w dwóch etapach: etap I w latach 1993-1996 i etap II w latach 1997-2000. Jest to system tłumaczenia mowy spontanicznej rozwijany i finansowany przez Federalne Ministerstwo Edukacji, Nauki, Badań i Technologii, koncerny Daimler-Benz, Siemens, wyższe uczelnie, uniwersytety i politechniki w Niemczech.

VERBMOBIL jest przenośnym urządzeniem, które umożliwia tłumaczenie rozmów nt. spotkań biznesowych między niemieckimi i japońskimi użytkownikami programu, którzy posiadają zaledwie bierną znajomość języka angielskiego. Wykorzystywane jest podczas bezpośrednich spotkań, co pozwala na kontrolowanie jakości tłumaczenia przez rozmówców, którzy mogą poprawiać, dodawać lub uzupełniać je elementami niewerbalnymi (gesty, wymiana wizytówek, pokazywanie kalendarza, robienie notatek itp.).

System VERBMOBIL jest przenośnym urządzeniem, które:

- rozpoznaje dialogi w sytuacji bezpośredniej,
- służy komunikacji w pewnych specjalistycznych dziedzinach,
- posiada trzy scenariusze językowe (angielski, niemiecki i japoński), przy czym język angielski jest językiem dialogu.

Struktura systemu ma następującą postać (por. rys. 3.):



Rys. 3. Architektura systemu VERBMOBIL (Jekat &amp; Klein 1996, 15)

Czas przetwarzania zdania wyjściowego (input) jest sześć razy dłuższy niż generowanie zdania wyjściowego (output). W systemie VERBMOBIL czas wygenerowania zdania wynosi 0,7 sek. Czas ten potrzebny jest na:

- rozpoznanie mowy (38% czasu)
- analizę prozodyczną (17%)
- analizę składniową i semantyczną (25%)
- semantyczną ocenę dialogu (14%)
- transfer (3%)
- generowanie (3%).

System VERBMOBIL tłumaczy jedynie centralną informację, wyrazy lub zdanie kluczowe użyte w dialogu. System wyróżnia następujące sekwencje dialogów:

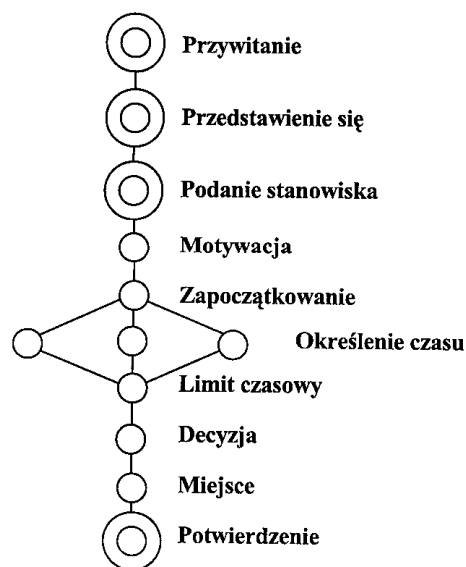
- akceptację (data, miejsce spotkania i czasu trwania spotkania), np.  
*Yes a quarter to three would suit me fine.*
- odmowę, np.  
*That's not so good.*
- sugestię lub prośbę, np.  
*Couldn't we say half past two then?*  
*When would it suit you?*

Tłumaczy on bezpośrednio wyrażenia czasowe oraz sekwencje dialogów. Podobnie czyni człowiek tłumaczący dialog z wykorzystaniem techniki redukcji i kompensacji w celu zachowanie ekwiwalencji między tekstem wyjściowym i tekstem docelowym. W systemie VERBMOBIL ekwiwalenty tłumaczeniowe są zdefiniowane wstępnie, nie ma indywidualnego podejmowania decyzji odnośnie technik tłumaczenia, sekwencja tłumaczonego dialogu może być niekompletna, ekwiwalenty definiuje się w zależności od dziedziny i typu dialogu.

Struktura dialogu oparta jest na następujących przesłankach:

- rozmówcy, którzy chcą się porozumieć, współpracują,
- mają oni wspólny cel komunikacji: umawianie spotkania,
- między nimi istnieje symetria, mają porównywalny status społeczny i zawodowy,
- rozmówcy są zobowiązani zachować grzeczność,
- rozmowa ma charakter rzeczowy, w dialogu nie ma figur retorycznych, gier słownych ani kłamstw.

Struktura dialogu dotyczącego spotkania przybiera następującą postać (por. rys. 4.):



Rys. 4. Struktura dialogu na temat spotkania (Jekat 1997, 151)

Do głównych zalet systemu Verbmobil należą:

- zmniejszenie ilości błędów na poziomie wyrazów z 50% do 14%; w chwili obecnej jest to najlepszy wynik na świecie dla systemów tłumaczących mowę spontaniczną,
- możliwość rozpoznania informacji prozodycznych: komponent ten przyspiesza analizę składniową o 92%, i redukuje liczbę interpretacji,
- pozytywna ocena poprawności tłumaczeń jest stosunkowo wysoka: na 25 tysięcy zbadanych przekładów, ocenianych przez tłumaczy profesjonalnych, 74,2% tłumaczeń poprawnie oddaje treść,
- system kompensuje informację wejściową, gdyż jego głównym zadaniem jest oddanie sekwencji dialogu i treści zdania,
- system oddaje różnice kulturowe i językowe: rozpoznaje centralną funkcję każdej wypowiedzi (np. formułki powitalne), dlatego też dodawanie nowych języków nie powinno być skomplikowane,
- system pozwala na zachowanie równowagi między ekwiwalencją formalną i funkcjonalną.

Pomimo wymienionych powyżej zalet w systemie VERBMOBIL następujące kwestie pozostają wciąż nierozwiązane:

- segmentacja sekwencji na jednostki odpowiadające sekwencjom dialogów,
- identyfikacja centralnej sekwencji dialogu.

Dlatego też w etapie II VERBMOBILU wprowadzono następujące zmiany:

- system jest wielojęzyczny: tłumaczy dialogi w parach językowych angielski – niemiecki, niemiecki – japoński i japoński – angielski. Słownik angielsko-niemiecki zawiera 10 tys. wyrazów, a słownik niemiecko-japoński 2,5 tys. wyrazów,
- system jest polifunkcyjny, dotyczy rozmaitych tematów (nie tylko umawianie spotkań), dialogi mogą być prowadzone w mowie swobodnej, spontanicznej, niezależnie od mikrofonu.
- system jest samosterujący i rozpoznaje początek i koniec sekwencji do przetworzenia,
- możliwe jest prowadzenie multimedialnej konferencji z udziałem więcej niż 2 partnerów, tłumaczenie jest dwukierunkowe jednoczesne z niemieckiego na angielski i japoński
- w systemie poprawiono jakość mowy i nie chodzi tutaj tylko o artykulację, lecz o prozodię zdania, tj. zjawiska ponadzdaniowe, jak akcent logiczny, który może być realizowany poprzez konstrukcje składniowe, pauzy, czas trwania samogłosek, prozodię lub kombinację tych elementów.

### Podsumowanie

Oba omówione powyżej systemy automatycznego tłumaczenia dialogów w czasie rzeczywistym wciąż stoją przed szeregiem nierozwiązanych problemów, do których jako najważniejsze należą:

- intonacja i jej rola pragmatyczna oraz ekspresywna; systemy nie potrafią odpowiednio zanalizować np. rozkazu, ironii, czy też zdziwienia rozmówcy wyrażonych za pomocą intonacji. Stąd też podejmuje się osobne badania nad konwencjami intonacyjnymi w różnych językach, oraz nad tworzeniem specjalnych korpusów intonacyjnych (por. Jassem 1996), nad ich rolą w rozpoznawaniu elementów prozodycznych, np.
  - franc. *Vous le prenez?*  
*Vous le prenez.*
- analiza sekwencji mowy i ich identyfikacja w celu stworzenia odpowiednich ekwiwalentnych par dialogu, np.
  - ang. *can you ...?* – action request  
– yes/no question
  - jap. *Hai* – ang. *yes*  
*Hai* – ang. *OK, all right*

Automatyczne systemy tłumaczenia dialogów w czasie rzeczywistym są w dobie obecnej w dwojaki sposób ograniczone. Pierwsze ograniczenie jest natury informatycznej i dotyczy sztucznej inteligencji. Pomimo ogromnych nakładów finansowych i badań nad głosową komunikacją między człowiekiem i komputerem zbudowanie słowników, gramatyk, systemów, które będą w stanie rozpoznać i dokonać syntezy mowy we wszystkich jej aspektach (tj. dialekty, prozodia, akcenty) i jednocześnie tłumaczyć dialog wciąż nie jest możliwe. Drugie ograniczenie jest natury translatorycznej. Jeżeli maszyna poradzi sobie z problemem segmentacji i rozpoznawania sekwencji mowy, jeżeli programy tłumaczenia będą w stanie objąć rozmaite tematy konwersacji (dziś są one bardzo ograniczone), wciąż pozostaje aktualne pytanie, skąd maszyna będzie wiedziała, której techniki tłumaczenia użyć w celu zapewnienia odpowiedniej ekwiwalencji. Pojawia się zatem pytanie, czy maszyna może wykonać twórczą pracę tak jak człowiek? Również kwestia jakości tłumaczenia, którą w automatycznym tłumaczeniu pisemnym zapewnia pre- i postedykcja pozostawia wiele do życzenia.

Choć minęło już ponad 20 lat od obietnic badaczy japońskich, nie nosimy jeszcze małych kieszonkowych urządzeń, które rozpoznają czyjś głos, „rozumieją” zdania i wypowiadają je w innym języku. W rezultacie maszyna nie jest jeszcze w stanie zastąpić człowieka tłumacza, na przekór przewidywaniom naukowców.

### Bibliografia

- Apfelbaum B., Wadensjö C., 1997, „How does Verbmobil Affect Conversation? Discourse Analysis and Machine-supported Translatory Interaction”, in: Hauenschild Ch & Heizmann S. (red.), 93-122.
- Barczyńska A., 2000, *La désambiguïsation lexicale en TA*, niepublikowana praca magisterska ILS UW.
- Basztura Cz., 1996, „Założenia modelu głosowego wejścia do systemu komputerowego tłumaczenia w czasie rzeczywistym”, in: Basztura Cz., & Dobrogowska K. (red.), 3-25.
- Basztura Cz., & Dobrogowska K. (red.), 1996, *Podstawowe założenia fonetyczne tłumaczenie różnojęzycznego dialogu w czasie rzeczywistym*, PTF, Poznań.
- D’allesandro Ch. & Tzoukermann E., (red.), 2001, „Synthèse de la parole à partir du texte”, numéro spécial de „Traitement automatique des langues”, vol. 42, no 1.
- Hauenschild Ch. & Heizmann S. (red.), 1997, *Machine Translation and Translation Theory*, Mouton de Gruyter, Berlin, New York.
- Jassem W., 1996, „Intonacje rdzenne w dialogu angielskim. Analiza akustyczna i statystyczna”, in: Basztura Cz., & Dobrogowska K. (red.), 85-104.
- Jekat S., 1997, „Authomatic Interpreting of Dialogue Acts”, in: Hauenschild Ch. & Heizmann S. (red.), 145-55.
- Jekat S., Klein A., 1996, „Machine Interpretation. Open Problems and Sole Solutions”, „Interpreting” 1 (1), 7-20.
- Kay M., Gawron J.-M., Norwig P., 2000, *Verbmobil: A Translation System for Face-to-Face Dialog*, Stanford CA, Center for the Study of Language and Information.
- Kitano H., 1993, „La traduction de la langue parlée”, in: Clas A. & Bouillon P. (éds.) *La traductique: Etudes et recherches de traduction par ordinateur*, Montréal, Presses de l’Université de Montréal, 408-22.
- Krauer S. (red.), 2000, *Spoken Language Translation*, special issue of „Machine Translation”, vol. 15, no -2.
- LuperFoy S., 1996, „Machine Interpretation of Bilingual Dialogue”, „Interpreting” 1 (2), ss 213-233.
- LuperFoy S., 1997, „Discourse Processing for Voice-to-Voice Machine Translation”, in: Hauenschild Ch. & Heizmann S. (red.), 223-250.
- Moser-Mercer B., 1997, „Process Models in Simultaneous Interpretation”, in: Hauenschild Ch. & Heizmann S. (red.), 3-17.



- Prahl B., Petzolt S., 1997, „Translation Problems and Translation Strategies Involved in Human and Machine Translation: Empirical Studies”, in: Hauenschild Ch. & Heizmann S. (red.), 123-144.
- Rayner M., Bouillon P., Carter D., et al. (red.), 2000, *Spoken Language Translator*, Cambridge.
- Seligman M., 2000, „Nine Issues in Speech Translation”, „Machine Translation”, no 1-2, vol. 15, 149-185.
- Schmitz B., 1998, *Pragmatikbasiertes Maschinelles Dolmetschen*, Heidelberg, J. Gross.
- Tomokiyo M., 1994, „Transfert de la langue parlée japonais-anglais dans le système de traduction automatique ASURA”, in: Clas A. & Bouillon P. (éds.), *TAO-TA: Recherches de pointe et application immédiates*. Montréal, Presses de l'Université de Montréal, 149-164.
- Wahlster W. (red.), 2000, *Verbmobil: Foundations of Speech-to-Speech Translation*, Berlin, Springer.