

# Wybrane aspekty tłumaczenia maszynowego japońsko-polskiego ze szczególnym uwzględnieniem zjawiska anafory

## Streszczenie

Przetwarzanie języka naturalnego stanowi obecnie jedną z najważniejszych poddziedzin językoznawstwa stosowanego. W dziedzinie tej tworzenie programów zdolnych przetłumaczyć tekst źródłowy na tekst docelowy w innym języku naturalnym stanowi szczególnie trudne zadanie. Obecnie wykorzystywane algorytmy tłumaczeniowe nie są w stanie przełożyć tekstu w sposób w pełni satysfakcjonujący, głównie ze względu na trudności w rozpoznaniu przez maszynę niuansów semantycznych występujących w języku naturalnym. Podobnie dzieje się w przypadku wyrażen anaforycznych, jakich interpretacja zwykle nie sprawia problemów zwykłemu użytkownikowi danego języka, a które w wielu przypadkach pozostają z punktu widzenia komputera jako odbiorcy przekazu właściwie całkowicie niezrozumiałe. Poprawne zinterpretowanie wyrażenia anaforycznego lub uzupełnienie brakujących semantycznych komponentów wyrażenia eliptycznego wiąże się bowiem często z przeprowadzeniem procedur dedukcji i z wykorzystaniem wiedzy o otaczającym użytkownika świecie pozajęzykowym, co zwyczajnie nie leży w zakresie możliwości obecnie używanych komputerów.

Jako jeden z języków nastroczających licznych problemów związanych z funkcjonowaniem anafory w przekazie można rozpoznać język japoński, w którym elementy zdania często są opuszczane w strukturze powierzchniowej wypowiedzi. Jedno z najczęściej występujących tego typu zjawisk to brak wyraźnie zaznaczonego podmiotu lub dopełnienia bliższego. Ze względu na brak morfologicznych wykładników osoby, liczby oraz rodzaju gramatycznego, wskazanie referenta musi w takich przypadkach opierać się wyłącznie na dedukcji z wykorzystaniem innych danych językowych – takich jak wzorce składniowe wskazywane dla danego predykatu – lub ogólnodostępnych dla każdego rozmówcy w postaci wiedzy o otaczającym go świecie. Proces ten stanowi umiejętność, jaka prawdopodobnie jeszcze długo nie będzie możliwa do zaimplementowania w obecnie używanych komputerach.

Niniejsza rozprawa służy rozważeniu wyżej przedstawianych kwestii w kontekście możliwości uzyskania efektywnego systemu służącego do automatycznego wykonywania tłumaczenia japońsko-polskiego. Głównym jej celem jest przedstawienie stanu dziedziny badań nad tłumaczeniem maszynowym, w szczególności w kontekście interpretacji wyrażen anaforycznych, oraz odniesienie wniosków z analizy do ogólnie pojętego tłumaczenia

maszynowego japońsko-polskiego. Odniesienie to służy przede wszystkim wyszczególnieniu potencjalnych punktów problemowych w przyszłych systemach automatycznego przekładu japońsko-polskiego oraz ocenie realistycznych oczekiwań stawianych wobec takich systemów. W języku japońskim natomiast odniesienia anaforyczne, w tym przypadki elipsy, występują stosunkowo często, stąd zasadne wydaje się przyjąć, że kwestia ich poprawnej interpretacji jest kluczowa w tłumaczeniu z użyciem języka japońskiego jako źródłowego.

W rozdziale pierwszym podjęto próbę opisu głównych zagadnień tłumaczenia maszynowego. W tej części rozprawy przedstawiono historię badań w dziedzinie prowadzonych na przestrzeni ostatnich kilkudziesięciu lat, omówienie zastosowań automatycznego przekładu oraz opis metod ewaluacji jakości tekstu docelowego. Przede wszystkim jednak skupiono się na omówieniu najważniejszych technik tłumaczenia maszynowego. Informacje zawarte w rozdziale pierwszym mają w zamierzeniu zobrazować ogólnie rozwój dziedziny, jak również zdefiniować przynajmniej w przybliżeniu obecny zakres możliwości systemów tłumaczeniowych, ze szczególnym wskazaniem na ich ograniczenia.

Rozdział drugi w całości poświęcono omówieniu zagadnienia anafory. Przedstawiono w nim w pierwszej kolejności definicje rozmaitych wyrażeń anaforycznych oraz omówiono ich najważniejsze klasyfikacje. Ze względu na fakt, że problematyka niniejszej rozprawy dotyczy tłumaczenia z języka japońskiego, przykłady występowania wyrażeń anaforycznych w tekście zdecydowano się podawać po japońsku. W kolejnej części rozdziału zebrano informacje dotyczące uwzględniania istotnych czynników językowych mających wpływ na interpretację wyrażeń anaforycznych. Czynniki te podzielono schematycznie na trzy grupy: morfologiczno-składniowe, semantyczne oraz pragmatyczne. W ten sposób dążono do wstępnego i formułowanego na niezmiernie ogólnym poziomie orzekania wskazania, jakiego typu informacje należy zakodować w systemach maszynowych, by były one w stanie dokonać poprawnej interpretacji tekstu, a w konsekwencji jego tłumaczenia.

W rozdziale trzecim skupiono się na omówieniu najważniejszych narzędzi przetwarzania języka naturalnego stworzonych dla języka japońskiego. Przedstawione wyliczenie technik przetwarzania języka uwzględnia podany w rozdziale drugim trójpodział czynników językowych na morfologiczno-składniowe, semantyczne oraz pragmatyczne. Tam, gdzie było to możliwe, umieszczono rysunki pokazujące format, w którym zapisane są informacje w poszczególnych narzędziach. Trzy pierwsze części rozdziału trzeciego ukazują metody kodowania informacji niezbędnych dla dokonania maszynowej interpretacji wyrażenia anaforycznego oraz ich ograniczenia. Sekcja czwarta poświęcona została zagadnieniu sieci

neuronowych, których wykorzystanie postrzegane jest obecnie jako najbardziej obiecująca metoda tłumaczenia maszynowego.

Rozdział czwarty poświęcono opisowi powstałych dotychczas systemów interpretacji wyrażeń anaforycznych w języku japońskim. W systemach tych wykorzystuje się narzędzia omówione już w rozdziale trzecim. Pierwsza część rozdziału czwartego zawiera przegląd najważniejszych programów do interpretacji anafory stworzonych dla języka japońskiego. Sekcja druga natomiast poświęcona jest zagadnieniu Winograd Schema Challenge; jest to konkurs mający na celu wyłonienie systemu sztucznej inteligencji zdolnego przeprowadzić wnioskowanie porównywalne z procesami myślowymi zachodzącymi u człowieka. Zadania w tym konkursie polegają na wskazaniu poprawnej interpretacji dla zawartego w zdaniu zaimka. W rozdziale tym wykazano również niedobory omówionych systemów, zarówno tych przeznaczonych do interpretacji anafory w języku japońskim, jak i tych konkurujących w Winograd Schema Challenge.

Rozdział piąty zawiera podsumowanie omawianych informacji oraz próbę udzielenia odpowiedzi na najważniejsze pytania badawcze formułowane w ramach niniejszej rozprawy, a mianowicie: do jakiego stopnia zasadne jest oczekiwanie uzyskania rozpoznania przez systemy komputerowe zjawiska anafory i w jakim stopniu ustalenia te można zastosować w tłumaczeniu w parze językowej japoński-polski. W pierwszej części rozdziału wyszczególniono punkty problemowe dotyczące rzeczywistego działania programów służących do interpretacji wyrażeń anaforycznych. Następnie przeprowadzono próbę odniesienia sformułowanych ustaleń do całości problematyki japońsko-polskiego tłumaczenia maszynowego. W dalszej części rozdziału wskazano również na trudności mogące wystąpić w maszynowym tłumaczeniu japońsko-polskim, o których wnioskowano na podstawie ogólnej analizy obecnej sytuacji w dziedzinie. Dzięki temu zdołano wyszczególnić najważniejsze kwestie, na które należy zwrócić uwagę przy konstruowaniu systemu automatycznego przekładu dedykowanego parze językowej japoński-polski.

Zakończenie rozprawy zawiera podsumowanie rozważań oraz krótką dyskusję na temat kierunków rozwoju systemów przetwarzania języka naturalnego, w tym systemów tłumaczeniowych z potencjalnym odniesieniem do tłumaczenia w perspektywie japońsko-polskiej. Na podstawie badań przeprowadzonych na potrzeby niniejszej rozprawy można wysnuć następujące wnioski:

- Istnieje konieczność opracowania wyważonych pod względem tematycznym i stylistycznym korpusów równoległych japońsko-polskich.

- Obecnie prowadzone nad zjawiskiem anafory badania należy rozszerzyć na różnorodne typy wyrażeń anaforycznych.
- System tłumaczenia japońsko-polskiego należy opracować z pominięciem języka angielskiego jako pośredniego, zamiast tego dedykując taki system jedynie wzmiankowanej parze językowej.
- Należy kontynuować badania nad interpretacją wyrażeń anaforycznych w języku japońskim w celu polepszenia wyników osiąganych przez obecnie istniejące systemy.
- Należy doskonalić obecnie istniejące narzędzia przetwarzania języka naturalnego pod kątem reprezentacji różnorodnych informacji językowych oraz ich przeszukiwania.
- Warto rozważyć ponowne zdefiniowanie oczekiwań stawianych obecnie istniejącym systemom tłumaczenia maszynowego w świetle ich ograniczeń oraz różnic w porównaniu do przekładu wykonywanego przez człowieka.
- Prowadzone współcześnie badania w dziedzinie efektywności procedur wnioskowania realizowanych przez komputer, w szczególności w zakresie opisanego w rozdziale czwartym konkursu Winograd Schema Challenge, mogą znacząco przyczynić się do poprawy jakości automatycznego przekładu, w tym przekładu japońsko-polskiego.

Obecnie nie istnieją systemy tłumaczeniowe dedykowane parze językowej japoński-polski, uznano jednak, że omówione w rozprawie kwestie mogą stanowić podstawę teoretyczną dla stworzenia takich programów w przyszłości.