

# **WNIOSEK O PORTFOLIO:**

## **Weryfikacja koncepcji możliwości budowy wielojęzycznego systemu komputerowego przekładu typu Human-Aided Machine Translation opartego na wykorzystaniu idei języka pośredniczącego**

**Autorzy:**

*dr inż. Mirosław Gajer, AGH w Krakowie, wydział EAIIB, Katedra Informatyki Stosowanej*  
*dr hab. Joanna Dybiec-Gajer, Uniwersytet Pedagogiczny w Krakowie Wydział Neofilologii*  
*dr inż. Zbigniew Handzel, Wydział Zarządzania i Komunikacji Społecznej*

**Centrum Inteligentnych Systemów Informatycznych** Akademia Górniczo-Hutnicza im. Stanisława Staszica al. Mickiewicza 30, 30-059 Kraków  
budynek C-2 pokój 426 tel.: 12 617 44 53 [www.isi.agh.edu.pl](http://www.isi.agh.edu.pl) [isi@agh.edu.pl](mailto:isi@agh.edu.pl)

# Opis merytoryczny

## 1. Wprowadzenie

Celem wnioskowanego projektu jest dokonanie weryfikacji koncepcji dotyczącej możliwości budowy i praktycznego wykorzystania wspomaganego przez użytkownika systemu komputerowego przekładu typu Human-Aided Machine Translation (HAMT), opartego na zastosowaniu koncepcji języka pośredniczącego (tzw. interlingua).

Podlegający analizie system z założenia ma być systemem wielojęzycznym oraz ma umożliwić dokonywanie przekładu w dowolnym kierunku pomiędzy dowolnie wybraną parą dostępnych w nim języków naturalnych (a także i sztucznych, jak na przykład esperanto lub ido). W początkowym etapie weryfikacji koncepcji Human-Aided Machine Translation wnioskowany system analizowany będzie głównie pod kątem wykorzystania w nim języków polskiego i angielskiego.

W etapie kolejnym zostaną rozpatrzone również i inne języki, przede wszystkim z germańskiej grupy językowej (niemiecki, niderlandzki, afrikaans, szwedzki, norweski, duński, islandzki) oraz z grupy języków romańskich (francuski, portugalski – w wersji europejskiej i brazylijskiej, hiszpański, włoski, kataloński, rumuński). Dalsza weryfikacja koncepcji będzie miała na celu zbadanie możliwości rozszerzenia funkcjonalności systemu o wybrane języki słowiańskie (rosyjski, ukraiński, białoruski, czeski, słowacki, słoweński, chorwacki, serbski, bośniacki, macedoński, bułgarski) oraz inne języki należące do rodziny indoeuropejskiej (grecki, albański, ormiański, walijski, bretoński, irlandzki, perski, hindi, urdu, pendżabski, bengali). Docelowo wnioskowany system uwzględni także wybrane języki nie należące do indoeuropejskiej rodziny językowej, które są określane mianem języków nostratycznych (arabski, hebrajski, turecki, węgierski, fiński, estoński, koreański, japoński). W przyszłości planowane jest także podjęcie próby rozszerzenia funkcjonalności systemu na inne wybrane języki orientalne, takie jak chiński (mandaryński i kantoński), tajski, wietnamski, indonezyjski, malajski i tagalski.

Już wstępie należy wyraźnie zaznaczyć, że przeznaczenie planowanego systemu jest nieco odmienne, niż ma to miejsce w przypadku w pełni automatycznych systemów komputerowego przekładu, gdzie na wejście tego rodzaju systemu zadawany jest zawsze gotowy już tekst zapisany w języku wyjściowym przekładu, który tłumaczony jest następnie przez komputer, a na wyjściu systemu otrzymujemy w związku z tym gotowy produkt w postaci przekładu rozważanego tekstu na wybrany język docelowy.

Tymczasem w przypadku systemu typu Human-Aided Machine Translation tłumaczenie tekstu ma miejsce bezpośrednio na etapie jego powstawania, przy czym tekst podlegający przekładowi tworzony jest przez użytkownika w tzw. języku kontrolowanym, który dopuszcza do użytku tylko wybrane konstrukcje składniowe danego języka naturalnego oraz ma precyzyjnie zdefiniowaną warstwę leksykalną. Nałożenie tego rodzaju więzów na tekst tworzony w języku wyjściowym stanowi gwarancję odpowiednio wysokiej jakości uzyskiwanych przekładów, ponieważ w dużym stopniu wyeliminowane zostaje wówczas zjawisko wieloznaczności języka naturalnego, zarówno na poziomie jego leksyki, jak i morfologii oraz składni danego języka.

Analizowany system ma być w zamierzeniu systemem otwartym zarówno pod względem leksykalnym, jak i składniowym. Otwarcie leksykalne polega na tym, że dopuszczalna jest rozbudowa zasobów słownictwa języka poprzez dodawanie do lingwistycznych baz danych nowych rekordów – jednostek leksykalnych. Z kolei otwarcie składniowe polega na dopuszczeniu możliwości rozbudowy zestawu dozwolonych struktur składniowych o nowe typy związków wyrazowych.

Jeżeli chodzi o zasoby leksykalne tworzonego systemu, to oprócz podstawowego słownictwa należącego do sfery języka ogólnego, planowane jest rozwijanie zasobów leksykalnych, głównie pod kątem przyszłych zastosowań systemu w obszarze nauk ścisłych, (matematycznych, technicznych, ekonomicznych), a także lingwistyki oraz nauk przyrodniczych, biologicznych i medycznych.

Zaproponowane podejście oparte na zastosowaniu języków kontrolowanych jako języków wyjściowych przekładu powinno istotnie podnieść jakość tłumaczeń otrzymywanych przy użyciu systemu typu Human-Aided Machine Translation.

Wnioskowany system może być prawdopodobnie z powodzeniem stosowany w roli interaktywnych rozmówek turystycznych, z możliwością dalszej rozbudowy w kierunku dodania do nich modułów rozpoznawania i syntezy mowy.

Ponieważ działanie systemu oparte jest na idei wykorzystania języka pośredniczącego przekładu (interlingua), ewentualna rozbudowa systemu o dalsze języki powinna być stosunkowo prosta, gdyż polegać będzie na dodaniu modułu tłumaczącego z danego języka do języka pośredniczącego i z języka pośredniczącego na dany język.

Reasumując, wnioskowany system powinien być z założenia systemem otwartym zarówno semantycznie, jak i składniowo z możliwością systematycznej rozbudowy o kolejne języki. Tłumaczenie z języków kontrolowanych jest gwarantem wysokiej jakości uzyskiwanych przekładów.

W realizacji projektu zaangażowany będzie docelowo przynajmniej kilkunastoosobowy zespół składający się zarówno z informatyków (programowanie systemów sieciowych i urządzeń mobilnych), jak i osób z wykształceniem językoznawczym lub filologicznym (tworzenie lingwistycznych baz danych potrzebnych do funkcjonowania systemu).

Wnioskowany system ma w założeniu umożliwić sporządzanie wielu różnych wersji językowych dokumentów tekstowych. System umożliwi ich równoczesne wytworzenie dzięki koncepcji zastosowania języka pośredniczącego przekładu. W ten sposób można będzie na przykład jednocześnie tworzyć wiele wersji językowych wiadomości SMS lub tekstów poczty elektronicznej.

Ponadto proponowany system może z powodzeniem zastąpić tradycyjne papierowe rozmówki turystyczne, zapewniając tym samym efektywną wzajemną komunikację osób nie znających żadnego wspólnego im języka.

Co istotne, ponieważ wnioskowany system ma być z założenia systemem wielojęzycznym, dlatego może znaleźć nabywców nie tylko w Polsce, ale również i w innych krajach, których języki narodowe będą w systemie uwzględnione.

W oparciu o wnioskowany system mogą powstać także mobilne wielojęzyczne rozmówki turystyczne z dodatkową opcją rozpoznawania i syntezy mowy.  
Do silnych stron projektu można zaliczyć:

- Nowatorskie podejście do zagadnienia przekładu komputerowego z wykorzystaniem języków kontrolowanych (system typu Human-Aided Machine Translation);
- Oparcie działania systemu na wykorzystaniu języka pośredniczącego przekładu – łatwość rozszerzania systemu o nowe języki;
- Otwartość leksykalna systemu – możliwość poszerzania zasobów słownikowych o nowe jednostki;
- Otwartość składniowa systemu – możliwość wprowadzania nowych konstrukcji gramatycznych;
- Wielojęzyczność systemu – możliwość tłumaczenia z dowolnego języka uwzględnionego w systemie na dowolny inny język w nim występujący.  
Natomiast za słabą stroną projektu można uznać stosunkowo duży nakład pracy związany z akwizycją tzw. wzorców translacyjnych z wielu różnych języków świata potrzebnych do sprawnego funkcjonowania systemu.  
Ponadto pewnym problemem może być fakt, że do korzystania z systemu typu Human-Aided Machine Translation konieczne jest posiadanie przez jego użytkownika elementarnej wiedzy o języku (wiadomości typu, co to jest podmiot, orzeczenie i dopełnienie zdania, co to są zaimki osobowe, wskazujące i dzierżawcze, czym różni się strona czynna od biernej itp.), co może stanowić w praktyce swego rodzaju barierę utrudniającą jego powszechne wykorzystanie. Być może krótkie szkolenie dotyczące zasad korzystania z systemu będzie w stanie skutecznie rozwiązać ten problem.  
Ponieważ podejście do komputerowego tłumaczenia typu Human-Aided Machine Translation oparte na językach kontrolowanych i języku pośredniczącym przekładu jest podejściem nowatorskim i zgodnie z wiedzą wnioskodawców nie istnieją jeszcze tego typu wielojęzyczne systemy, jest zapewne sprawą dyskusyjną, czy takie nowe rozwiązanie może kiedyś w przyszłości zdobyć większą popularność i stać się dużym sukcesem rynkowym oraz czy potencjalni użytkownicy do pracy z tego typu systemem będą się w stanie przekonać i w pewnym stopniu przyzwycząić. Jest przy tym rzeczą niezmiernie ważną, aby obsługa systemu i współpraca z nim były możliwie jak najbardziej intuicyjne oraz aby obsługa systemu wymagała od jego użytkownika jedynie absolutnie niezbędnego minimum wiedzy o języku.

## 2. Przekład komputerowy wspomagany przez człowieka

W chwili obecnej historia badań nad automatyzacją przekładu liczy już ponad 60 lat. Pierwszy eksperyment, podczas którego tłumaczenie powierzono komputerowi, miał miejsce już w roku 1954. Wówczas za wielki sukces uznano przetłumaczenie kilkudziesięciu prostych zdań języka rosyjskiego na język angielski, przy czym komputer posiłkował się w tym celu jedynie 16 regułami gramatycznymi i słownikiem dwujęzycznym liczącym zaledwie 250 haseł [1].

W owym czasie panowało powszechne przekonanie, że całkowite zastąpienie tłumaczy przysięgłych odpowiednimi programami komputerowymi jest jedynie kwestią czasu. Jednak kolejne lata pokazały, jak bardzo się wówczas mylono, a przekład komputerowy o jakości porównywalnej z efektami pracy wykształconego w tym celu człowieka po dziś dzień pozostaje jedynie niedoścignionym ideałem [2].

Za taki stan rzeczy odpowiedzialne są przede wszystkim takie zjawiska, jak wieloznaczność i kontekstowość wypowiedzi formułowanych w dowolnym języku naturalnym. Występowanie wieloznacznych wyrazów bądź wieloznacznych konstrukcji składniowych, których znaczenie zależy dodatkowo od kontekstu ich występowania, nie stanowi bynajmniej większego problemu, który utrudniałby w jakimś istotnym stopniu komunikację międzyludzką, aczkolwiek i tutaj zdarzają się niekiedy pomyłki (czasami nawet komiczne) wynikające z błędnej interpretacji, czy też niezrozumienia kierowanej do rozmówcy wypowiedzi [3].

Tymczasem zastąpienie ludzkiego tłumacza programem komputerowym powoduje, że uzyskanie wysokiej jakości przekładu staje się niemożliwym do osiągnięcia celem, właśnie na skutek występowania różnego typu wieloznaczności i kontekstowości wypowiedzi formułowanych w języku naturalnym [4].

W tym miejscu można postawić bardzo ważne pytanie o to, czy dalsze badania nad automatyzacją przekładu mają jeszcze jakikolwiek sens i czy w ich przypadku można spodziewać się dojścia do rozwiązania satysfakcjonującego dla użytkowników tego typu oprogramowania?

Zdaniem autorów projektu odpowiedź na tak postawione pytanie jest twierdząca, jednakże pod warunkiem spełnienia kilku istotnych postulatów [5 – 18].

Po pierwsze, system tłumaczenia komputerowego musi być zorientowany na tłumaczenie tekstów z pewnej określonej dziedziny (na przykład teksty dotyczące nauk technicznych, przyrodniczych, medycznych itp.) Zakładanie, że możliwe jest stworzenie systemu, który będzie tłumaczył dosłownie wszystko, co tylko zostanie zadane na jego wejście, jest w praktyce niemożliwą do urzeczywistnienia utopią.

Po drugie, kontroli podlegać musi semantyka tłumaczonych tekstów. W tym celu należy opracować dla danej dziedziny odpowiednie słowniki i leksykony, których hasła posiadają precyzyjnie określone znaczenia.

Po trzecie, muszą zostać nałożone odpowiednie więzy na język, w którym tworzone są teksty mające podlegać przekładowi przez komputer. Więzy takie dotyczą przede wszystkim konstrukcji gramatycznych, które są dopuszczalne w użyciu.

Spełnienie powyższych postulatów prowadzi do powstania tzw. języka kontrolowanego, który stanowi pewien podzbiór języka naturalnego, z którego się pierwotnie wywodzi. Każde zdanie języka kontrolowanego powinno być zarazem poprawnym zdaniem danego języka naturalnego, natomiast nie każde zdanie języka naturalnego jest zdaniem należącym do języka kontrolowanego [19].

Języki kontrolowane są najczęściej zamknięte składniowo, co oznacza, że w ich przypadku dopuszczalne są tylko najprostsze i absolutnie niezbędne z punktu widzenia funkcjonowania języka konstrukcje gramatyczne. Natomiast ich leksyka może podlegać ciągłemu rozwojowi poprzez dodawanie do słownika określonych haseł związanych z daną dziedziną [20].

Główny nurt prowadzonych obecnie badań nad komputerowym przekładem skupia się głównie na pracach dotyczących systemów w pełni automatycznych, w przypadku których na wejściu zadawany jest gotowy już tekst zapisany w języku źródłowym, a na wyjściu systemu oczekujemy pojawienia się jego przekładu w języku docelowym [1].

Niestety, pomimo wielu lat wyłożonych badań, które są prowadzone w rozważanej dziedzinie nauki, w pełni automatyczny przekład wciąż pozostaje swego rodzaju niedoścignionym ideałem. Co gorsze, nic nie wskazuje na to, aby w najbliższej przyszłości

coś miało się w tym zakresie istotnie zmienić, nawet pomimo łożenia na ten cel potężnych środków finansowych przez takich potentatów, jak choćby Google [2].

Taki stan rzeczy powoduje, że być może warto jest próbować poszukiwać w tej dziedzinie innych niekonwencjonalnych rozwiązań, do których zaliczany jest na przykład przekład komputerowy wspomagany przez człowieka, określane angielskojęzycznym terminem Human-Aided Machine Translation (HAMT). Dotychczas ten kierunek badań jest stosunkowo słabo reprezentowany, ponieważ praktycznie cały wysiłek badawczy szedł do tej pory w kierunku rozwoju w pełni automatycznych systemów komputerowego przekładu [3].

Innym powodem dotychczasowego nikłego zainteresowania rozwojem systemów HAMT jest nieco inne ich przeznaczenie w porównaniu z systemami w pełni automatycznymi. Zasadnicza różnica w zasadach działania systemów w pełni automatycznych i systemów HAMT sprowadza się do tego, że w przypadku systemów HAMT nie mamy do czynienia z gotowym tekstem zapisanym w języku źródłowym przekładu, tylko tekst taki tworzony jest niejako w trybie on-line w trakcie interaktywnej współpracy użytkownika z systemem [4].

Głównym obszarem zastosowań systemów HAMT jest tworzenie wielojęzycznych dokumentów w międzynarodowych korporacjach. W takim przypadku użytkownik systemu HAMT tworzy tekst w tzw. języku kontrolowanym, który stanowi pewien podzbiór danego języka naturalnego bądź jest wręcz językiem sztucznym, utworzonym specjalnie do tego celu. W przypadku tego rodzaju języków znacznemu ograniczeniu podlega wszechobecna w każdym języku naturalnym wieloznaczność, która jest zasadniczą przyczyną trudności napotykanych podczas prac nad automatyzacją przekładu [5].

Prowadzone obecnie przez autorów prace stanowią wstęp do planowanego na przyszłość większego interdyscyplinarnego projektu, w realizację którego byliby zaangażowani specjaliści z różnych dziedzin, takich jak językoznawstwo, przekładoznawstwo, filologie poszczególnych języków, informatyka, a zwłaszcza systemy baz danych i inżynieria oprogramowania [6 – 19].

Istotę wnioskowanego przez autorów systemu HAMT stanowi zastosowanie języka pośredniczącego przekładu. Z tego powodu proces translacji przebiega dwuetapowo. Najpierw tekst utworzony przez użytkownika w języku źródłowym przekładu tłumaczony jest na język pośredniczący, a następnie dopiero z języka pośredniczącego generowany jest tekst w języku docelowym przekładu. Tego rodzaju podejście w zdecydowanym stopniu ułatwia budowę systemu wielojęzycznego, w którym istnieje możliwość tłumaczenia w obu kierunkach pomiędzy dowolnie wybraną parą języków.

### 3. Idea zastosowania języka pośredniczącego w tłumaczeniu komputerowym

Obecnie, w dobie powszechnie panującej globalizacji świata, w wielu obszarach działalności ludzkiej zachodzi konieczność jednoczesnego tłumaczenia różnego typu tekstów na wiele języków świata, niekiedy bardzo od siebie odległych w sensie genetycznym, typologicznym i geograficznym.

W ogólnym przypadku, aby móc tłumaczyć w dowolnym kierunku pomiędzy

dowolnie wybraną parą spośród  $N$  języków, należałoby opracować  $N(N - 1)$  programów komputerowej translacji. Tak więc dla  $N = 100$  języków trzeba byłoby opracować aż 9900 programów tłumaczących. Tak wielkie przedsięwzięcie wydaje się niemożliwe do urzeczywistnienia w praktyce, a programy komputerowej translacji dokonujące przekładu na przykład pomiędzy językiem gruzińskim a językiem irlandzkim, czy też pomiędzy językiem koreańskim a słoweńskim prawdopodobnie nie powstaną nigdy, ponieważ zapotrzebowanie na tego typu aplikacje zapewne jest stosunkowo niewielkie, chociaż w dobie powszechnej

globalizacji świata ewentualnej przydatności tego rodzaju oprogramowania nie można przecież z góry całkowicie wykluczyć.

W takim wypadku skutecznym rozwiązaniem wydaje się zastosowanie języka pośredniczącego przekładu, określanego łacińskim terminem *interlingua*. Zastosowanie języka pośredniczącego powoduje, że tłumaczenie pomiędzy językiem A i pomiędzy językiem B przebiega dwuetapowo. W pierwszej kolejności tekst zapisany w języku A przekładany jest na język pośredniczący, a następnie tekst zapisany w języku pośredniczącym przekładany jest dopiero na język B.

W przypadku tego typu rozwiązania dla  $N$  różnych języków świata wystarczy opracować zaledwie  $2N$  programów komputerowej translacji ( $N$  programów tłumaczących na język pośredniczący i  $N$  programów tłumaczących z języka pośredniczącego), aby uzyskać możliwość tłumaczenia w dowolnym kierunku pomiędzy dowolnie wybraną parą języków.

Zalety przyjęcia tego rodzaju rozwiązania wydają się być bezdyskusyjne, jednak oparcie komputerowego przekładu na języku pośredniczącym też nie jest bynajmniej pozbawione wad. W tym wypadku problem polega na tym, że jednostki leksykalne w jednym języku nie mają zwykle swych dokładnych odpowiedników w drugim języku, a jedynie ich pola semantyczne w różnym stopniu się przenikają i częściowo pokrywają. Wprowadzenie dwuetapowego procesu translacji, opartego na wykorzystaniu języka pośredniczącego, dodatkowo całą sprawę jeszcze mocno komplikuje i w efekcie semantyka uzyskanego przekładu może znacznie dobiegać od semantyki tekstu zapisanego w języku źródłowym.

Nie bez znaczenia jest również fakt, że różne języki świata w dość arbitralny sposób dzielą otaczającą nas rzeczywistość na klasy obiektów, którym następnie przypisywane są jednostki leksykalne, takie jak na przykład „drzewa”, „ptaki”, „ryby”, „kamienie”, „szafy”, „krzesła”, „gwiazdy” itp. Istniejące w tym względzie pomiędzy różnymi językami odstępstwa są niekiedy znaczne i z punktu widzenia użytkownika języka polskiego mogą być niekiedy wręcz zaskakujące. Na przykład przetłumaczenie na język japoński polskiego rzeczownika „siostra” nie jest wcale rzeczą oczywistą, ponieważ użytkownik języka japońskiego, z jakichś jemu tylko znanych powodów, musi jeszcze

**Centrum Inteligentnych Systemów Informatycznych** Akademia Górniczo-Hutnicza im. Stanisława Staszica al. Mickiewicza 30, 30-059 Kraków  
budynek C-2 pokój 426 tel.: 12 617 44 53 www.isi.agh.edu.pl [isi@agh.edu.pl](mailto:isi@agh.edu.pl)

koniecznie wiedzieć, czy siostra, o której jest mowa, jest siostrą młodszą czy też siostrą starszą. W pierwszym przypadku użytkownik języka japońskiego posłużyłby się rzeczownikiem „imoto”, a w drugim rzeczownikiem „onesan”.

Podobnie, aby przetłumaczyć na język włoski polski rzeczownik „dziecko” trzeba jeszcze znać płeć tego dziecka, podczas gdy w języku polskim jest to całkowicie nieistotne, ponieważ rzeczownik „dziecko” jest rodzaju nijakiego. Natomiast w języku włoskim, jeśli dziecko jest płci męskiej, należy posłużyć się rzeczownikiem „bambino”, a w przypadku przeciwnym rzeczownikiem „bambina”.

Z kolei arabski rzeczownik „kalam” oznacza cokolwiek, co służy do pisania, na przykład pióro, długopis, ołówek, pisak itp. Zatem jego przekład na język nie jest sprawą łatwą, zwłaszcza gdy nie dysponujemy szerszym kontekstem tłumaczonej wypowiedzi.

Analizując leksykę różnych języków świata tego rodzaju przykładów można wymienić bardzo wiele, a wprowadzenie języka pośredniczącego przekładu dodatkowo całą sprawę jeszcze tylko komplikuje [20].

Z tego powodu na etapie opracowywania koncepcji systemu komputerowego przekładu wspomaganego przez człowieka podjęto decyzję, że w przypadku wnioskowanego systemu komputerowego przekładu typu HMT podstawowymi jednostkami semantycznymi języka, które będą podlegały procesowi przekładu, nie będą w przeważającej liczbie przypadków pojedyncze wyrazy, które są zwykle wieloznaczne, lecz dłuższe frazy składające się z kilku wyrazów, które zwykle już można przełożyć na inny język w sposób jednoznaczny.



## 4. Planowany przebieg interakcji użytkownika z systemem komputerowego przekładu

W niniejszym punkcie opisano szczegółowo, w jaki sposób powinna przebiegać interakcja użytkownika z systemem komputerowego przekładu typu HMT, polegająca na budowaniu poprawnych syntaktycznie zdań w języku źródłowym przekładu (w tzw. języku

kontrolowanym), które następnie są tłumaczone przez komputer na język docelowy. Po uruchomieniu programu komputerowego przekładu typu HMT ekran monitora

powinien przyjąć widok, jak przedstawiono na rys. 1.



Rys. 1. Widok ekranu monitora podczas wyboru języka, którym posługuje się użytkownik programu

Na rozważanym etapie ma miejsce wybór języka, którym posługuje się użytkownik programu. Wszelkie informacje tekstowe początkowo zamieszczone są wyłącznie w języku angielskim, ponieważ język ten jest powszechnie postrzegany jako podstawowy język komunikacji międzynarodowej, w związku z czym można zakładać, że przedstawione na rys. 1 menu wyboru jest zrozumiałe dla każdego z użytkowników programu, posługującego się na co dzień dowolnym z wymienionych na rys. 1 języków. W celu ułatwienia dokonywania wyboru odpowiedniego języka docelowo można umieścić obok angielskich nazw odpowiednich języków także flagi państw, w których dany język pełni funkcje języka urzędowego.

Po zaznaczeniu odpowiedniej opcji wyboru języka użytkownik powinien kliknąć strzałkę z napisem „Next”, aby przejść do kolejnego etapu pracy z programem.

Jeżeli jako język, którym posługuje się użytkownik programu, został wybrany język angielski, wówczas ekran monitora przyjmuje widok, jak przedstawiono to na rys. 2. Na tym etapie użytkownik dokonuje wyboru języka źródłowego przekładu, czyli języka, z którego program będzie tłumaczył, oraz języka docelowego przekładu, czyli języka, na który program będzie tłumaczył. Ponieważ założono, że na etapie wcześniejszym użytkownik dokonał wyboru języka angielskiego jako języka, którym się na co dzień posługuje, dlatego język angielski zaznaczony jest domyślnie również jako język źródłowy przekładu. Oczywiście, użytkownik programu może tę opcję w każdej chwili zmienić i jako język źródłowy przekładu wskazać dowolny inny język znajdujący się w menu wyboru.



Source language	Target language
<input checked="" type="radio"/> English	<input checked="" type="radio"/> English
<input type="radio"/> German	<input type="radio"/> German
<input type="radio"/> Dutch	<input type="radio"/> Dutch
<input type="radio"/> French	<input type="radio"/> French
<input type="radio"/> Italian	<input type="radio"/> Italian
<input type="radio"/> Spanish	<input type="radio"/> Spanish
<input type="radio"/> Esperanto	<input type="radio"/> Esperanto
<input type="radio"/> Polish	<input type="radio"/> Polish

Next

Rys. 2. Widok ekranu monitora na etapie dokonywania wyboru języka źródłowego i języka docelowego w przypadku, gdy użytkownik uprzednio zadeklarował, że posługuje się językiem angielskim

Z kolei w przypadku, gdy użytkownik programu zadeklarował, że jego językiem ojczystym jest język polski, wówczas ekran monitora powinien przyjąć widok, jak przedstawiono na rys. 3.

### Język źródłowy i docelowy

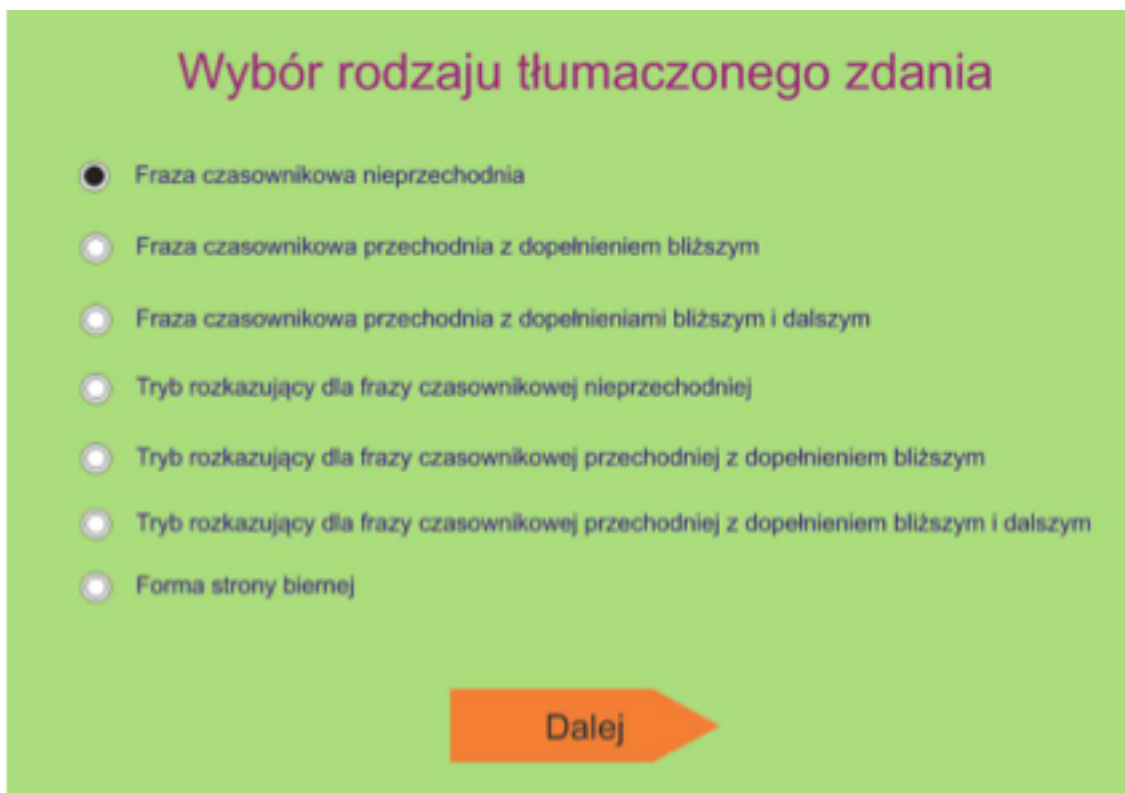
Język źródłowy	Język docelowy
<input type="radio"/> Angielski	<input checked="" type="radio"/> Angielski
<input type="radio"/> Niemiecki	<input type="radio"/> Niemiecki
<input type="radio"/> Niderlandzki	<input type="radio"/> Niderlandzki
<input type="radio"/> Francuski	<input type="radio"/> Francuski
<input type="radio"/> Włoski	<input type="radio"/> Włoski
<input type="radio"/> Hiszpański	<input type="radio"/> Hiszpański
<input type="radio"/> Esperanto	<input type="radio"/> Esperanto
<input checked="" type="radio"/> Polski	<input type="radio"/> Polski



Rys. 3. Widok ekranu monitora na etapie dokonywania wyboru języka źródłowego i języka docelowego w przypadku, gdy użytkownik uprzednio zadeklarował, że posługuje się językiem polskim

Co więcej, począwszy od tego momentu wszystkie kolejne komunikaty ukazujące się na ekranie monitora na dalszych etapach pracy z programem zamieszczane będą już pojawiały się w języku polskim. Reguła ta dotyczy dowolnego z języków, w przypadku którego użytkownik programu zadeklarował, że posługuje się nim na co dzień w sposób biegły. Ponadto w przypadku dokonania przez użytkownika wyboru języka polskiego, język ten automatycznie zostaje wybrany jako język źródłowy przekładu, choć oczywiście można ten wybór później zmienić.

Po kliknięciu przycisku z napisem „Dalej” użytkownik systemu przechodzi do kolejnego etapu, w którym musi podjąć decyzję odnośnie wyboru rodzaju zdania, które będzie chciał przetłumaczyć, co pokazano na rys. 4.



Rys. 4. Widok ekranu monitora na etapie dokonywania wyboru rodzaju zdania podlegającego procesowi automatycznego przekładu w systemie typu HAMT

Na rozważanym etapie użytkownik ma możliwość wyboru pomiędzy zdaniem typu SV (ang. Subject – Verb) z frazą czasownikową nieprzechodnią, a zdaniami typu SVO (ang. Subject – Verb - Object) z frazą czasownikową przechodnią z dopełnieniem bliższym bądź frazą czasownikową przechodnią z dopełnieniem bliższym i dalszym. Dodatkowo użytkownik ma możliwość wyboru zdania w formie trybu rozkazującego, w którym występuje fraza czasownikowa nieprzechodnia bądź fraza czasownikowa przechodnia z dopełnieniem bliższym lub fraza czasownikowa przechodnia z dopełnieniem bliższym i dalszym. Ostatnią z możliwości występujących na rozważanym etapie jest dokonanie wyboru zdania występującego w formie strony biernej.

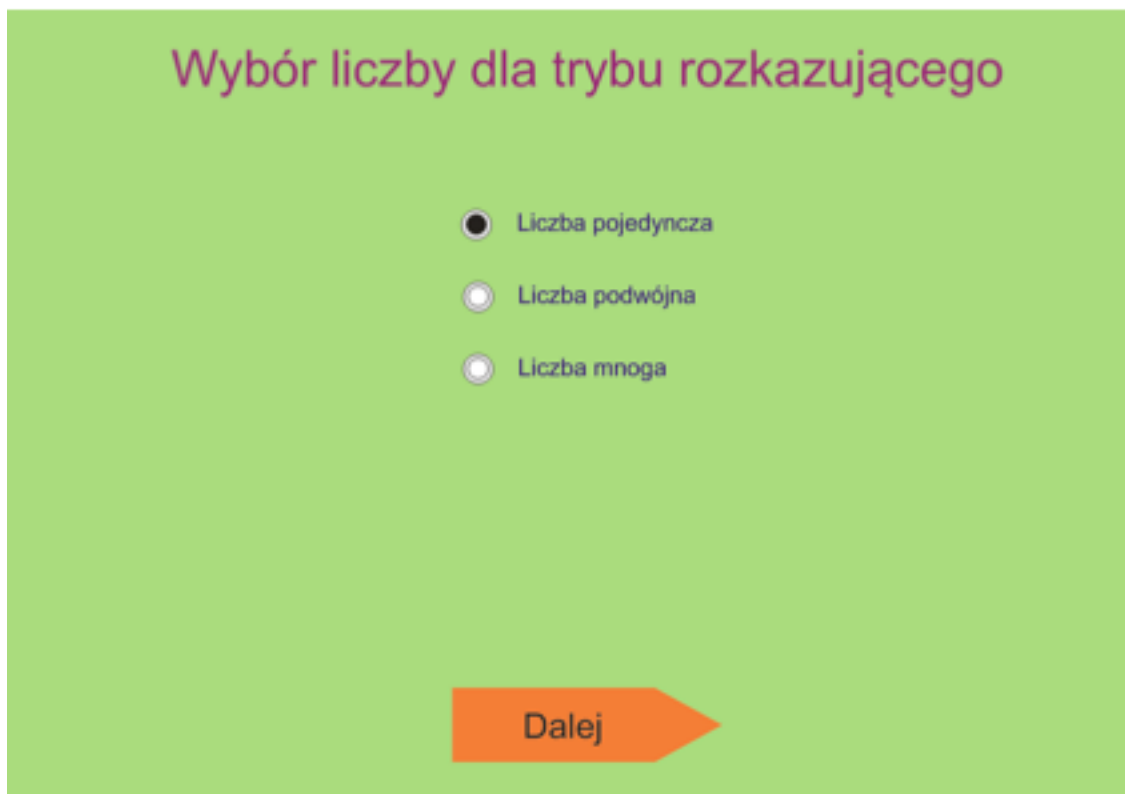
Jak już uprzednio wspomniano, opracowywany system pomyślany jest jako system otwarty syntaktycznie, co oznacza, że w przyszłości może zostać dodatkowo rozbudowany o możliwości tworzenia w nim również innych typów zdań. Jednak na obecnym etapie ograniczono się do uwzględnienia w systemie jedynie wymienionych podstawowych typów zdań.

W przypadku dokonania przez użytkownika wyboru jednej z opcji „Tryb rozkazujący dla frazy czasownikowej nieprzechodniej”, „Tryb rozkazujący dla frazy czasownikowej przechodniej z dopełnieniem bliższym” lub „Tryb rozkazujący dla frazy czasownikowej przechodniej z dopełnieniem bliższym i dalszym”, ekran monitora powinien przyjąć widok, jak pokazano na rys. 5.

Na rozważanym etapie użytkownik musi dokonać wyboru liczby gramatycznej, w której wystąpi fraza czasownikowa budowanego przez niego zdania w trybie rozkazującym. Jak wynika z rys. 5, użytkownik musi dokonać wyboru pomiędzy liczbą pojedynczą, liczbą podwójną i liczbą mnogą. Jak

**Centrum Inteligentnych Systemów Informatycznych** Akademia Górniczo-Hutnicza im. Stanisława Staszica al. Mickiewicza 30, 30-059 Kraków  
budynek C-2 pokój 426 tel.: 12 617 44 53 www.isi.agh.edu.pl [isi@agh.edu.pl](mailto:isi@agh.edu.pl)

wiadomo, uwzględnienie również formy liczby podwójnej (tzw. dualisu) jest konieczne ze względu na takie języki, jak na przykład słoweński i arabski, w których tego rodzaju kategoria gramatyczna nadal dosyć powszechnie występuje [20].



Rys. 5. Widok ekranu monitora na etapie dokonywania wyboru liczby gramatycznej dla frazy czasownikowej występującej w trybie rozkazującym

Jeżeli użytkownik systemu na etapie pokazanym na rys. 4 zdecydował się na wybór opcji „Tryb rozkazujący dla frazy czasownikowej nieprzechodniej”, wówczas za pomocą systemu komputerowego przekładu wspomaganego przez człowieka może dokonać tłumaczenia takich zdań, jak na przykład „Usiądź!”, „Nie rozmawiajcie!”, „Zatańczcie!” itp.

Z kolei w przypadku dokonania wyboru opcji „Tryb rozkazujący dla frazy czasownikowej przechodniej z dopełnieniem bliższym” użytkownik może za pomocą systemu tłumaczyć automatycznie na wybrane języki takie zdania jak na przykład „Ugotuj zupę!”, „Nie kupujcie pomarańczy!”, „Nie deptaj trawy!” itp.

Analogicznie w przypadku dokonania wyboru opcji „Tryb rozkazujący dla frazy czasownikowej przechodniej z dopełnieniem bliższym i dalszym” użytkownik może dokonywać przekładu takich zdań jak na przykład: „Oddaj pieniądze żonie!”, „Nie kupujcie cukierków dzieciom!”, „Nie śpiewaj tej piosenki mojemu bratu!” itp.

**Centrum Inteligentnych Systemów Informatycznych** Akademia Górniczo-Hutnicza im. Stanisława Staszica al. Mickiewicza 30, 30-059 Kraków  
budynek C-2 pokój 426 tel.: 12 617 44 53 www.isi.agh.edu.pl [isi@agh.edu.pl](mailto:isi@agh.edu.pl)

Jednak w celu budowy tego rodzaju zdań użytkownik musi dokonać jeszcze na dalszych etapach interakcji z systemem wyboru fraz stanowiących orzeczenie zdania oraz jego dopełnienie bliższe i ewentualnie dopełnienie dalsze.

Z kolei w przypadku dokonania przez użytkownika na etapie przedstawionym na rys. 4 jednej z następujących opcji: „Fraza czasownikowa nieprzechodnia”, „Fraza czasownikowa przechodnia z dopełnieniem bliższym”, „Fraza czasownikowa przechodnia z dopełnieniem bliższym i dalszym” lub „Forma strony biernej” ekran monitora powinien przyjąć wygląd, jak pokazano na rys. 6, ponieważ w następnej kolejności użytkownik systemu będzie musiał podjąć decyzję odnośnie wyboru podmiotu budowanego przez siebie zdania.

Jak wynika z rys. 6, na rozważanym etapie użytkownik, decydując się na wybór rodzaju podmiotu zdania, ma możliwość wyboru albo odpowiedniego zaimka osobowego, albo jednej z dostępnych w formie rozwijanego menu fraz rzeczownikowych. Jak już uprzednio wspomniano, opracowywany system planowany jest jako system otwarty leksykalnie, co umożliwi systematyczne wzbogacanie jego repertuaru o kolejne dostępne do wyboru frazy rzeczownikowe.



Rys. 6. Widok ekranu monitora na etapie dokonywania wyboru podmiotu tłumaczonego zdania

Jeśli użytkownik zdecydował się na wybór zaimka osobowego w roli podmiotu zdania, wówczas ekran monitora powinien przybrać widok, jak przedstawiono na rys. 7.



Rys. 7. Widok ekranu monitora w przypadku wyboru zaimka osobowego jako podmiotu tłumaczonego zdania

Jak wynika z postaci rys. 7, dokonując wyboru odpowiedniej formy zaimka osobowego użytkownik musi wybrać przede wszystkim osobę gramatyczną (pierwszą, drugą lub trzecią), a następnie liczbę gramatyczną (pojedynczą, podwójną bądź mnogą) i rodzaj gramatyczny (męski lub żeński, a dodatkowo w trzeciej osobie może to być również rodzaj nijaki).

Tego rodzaju, na pozór wyglądający dość skomplikowanie, sposób wyboru formy zaimka osobowego jest jednak konieczny, ponieważ różne języki świata oferują w tym względzie bardzo różne rozwiązania, co zostało poniżej szczegółowo omówione na kilku wybranych przykładach [20].

Jak wiadomo, w języku polskim zaimki osobowe podlegają odmianie przez osoby, liczby i rodzaje gramatyczne. Jednak system zaimków osobowych języka polskiego nie jest bynajmniej w jakiś szczególny sposób nadmiernie rozbudowany w porównaniu z innymi językami świata.

Na przykład nieco bardziej skomplikowany system zaimków osobowych w stosunku do języka polskiego posiada język hiszpański, w przypadku którego prezentuje się on w sposób następujący (tab. 1).

Tab. 1. System zaimków osobowych języka hiszpańskiego

	<b>Liczba pojedyncza</b>	<b>Liczba mnoga</b>
<b>Osoba pierwsza</b>	yo	nosotros (rodzaj męski) nosotras (rodzaj żeński)
<b>Osoba druga</b>	tú	vosotros (rodzaj męski) vosotras (rodzaj żeński)
<b>Osoba trzecia</b>	él (rodzaj męski) ella (rodzaj żeński)	ellos (rodzaj męski) ellas (rodzaj żeński)

Jak widać, swego rodzaju „nowością” w stosunku do języka polskiego jest występowanie w języku hiszpańskim form męskich i żeńskich dla zaimków „my” i „wy”. Wygląda to trochę tak, jak gdyby ktoś mówił „my mężczyźni”, „my kobiety” lub „wy mężczyźni”, „wy kobiety”. Z kolei w języku hiszpańskim brak jest rodzaju nijakiego.

Jeszcze bardziej złożony system zaimków osobowych wykazuje na przykład język słoweński, w przypadku którego przedstawia się on w sposób przedstawiony w tab. 2.

Tab. 2. System zaimków osobowych języka słoweńskiego

	<b>Liczba pojedyncza</b>	<b>Liczba podwójna</b>	<b>Liczba mnoga</b>
<b>Osoba pierwsza</b>	jaz	midva (rodzaj męski) midve (rodzaj żeński)	mi (rodzaj męski) me (rodzaj żeński)
<b>Osoba druga</b>	ti	vidva (rodzaj męski) vidve (rodzaj żeński)	vi (rodzaj męski) ve (rodzaj żeński)
<b>Osoba trzecia</b>	on (rodzaj męski) ona (rodzaj żeński) ono (rodzaj nijaki)	onadva	oni

W przypadku języka słoweńskiego kolejną nowością jest pojawienie się form zaimków osobowych dla liczby podwójnej, których należy używać, gdy w danej sytuacji komunikacyjnej występują dokładnie dwie osoby. Gdyby tego rodzaju formy występowały w języku polskim, wówczas należałoby być może mówić „my dwaj”, „my dwie” lub „wy dwaj”, „wy dwie”.

Jeszcze bardziej skomplikowanego systemu zaimków osobowych można dopatrzeć się w języku arabskim, w którym również istnieje liczba podwójna, a dodatkowo w drugiej osobie zaimki osobowe wykazują różnicowanie ze względu na rodzaj gramatyczny (męski lub żeński), w związku z czym system zaimków osobowych języka arabskiego wygląda, jak przedstawiono w tab. 3, przy czym zastosowano w tym wypadku uproszczoną transkrypcję fonetyczną.



Tab. 3. System zaimków osobowych języka arabskiego

	Liczba pojedyncza	Liczba podwójna	Liczba mnoga
<b>Osoba pierwsza</b>	ana	nahnu	nahnu
<b>Osoba druga</b>	anta (rodzaj męski) anti (rodzaj żeński)	antuma	antum (rodzaj męski) antunna (rodzaj żeński)
<b>Osoba trzecia</b>	huua (rodzaj męski) hija (rodzaj żeński)	huma	hum (rodzaj męski) hunna (rodzaj żeński)

Z kolei znanych jest wiele języków, w przypadku których system zaimków osobowych jest uproszczony do granic możliwości, ponieważ zaimki odmieniają się jedynie przez osoby i występują tylko w formach charakterystycznych dla liczby pojedynczej i mnogiej. Tego rodzaju systemy zaimków osobowych posiadają, między innymi, wszystkie języki należące do rodzin chińsko-tybetańskiej, uralskiej i altajskiej, a także języki takie, jak japoński, koreański, indonezyjski, malajski, tagalski i wietnamski. Spośród języków indoeuropejskich tego rodzaju system zaimków osobowych wykazuje między innymi język perski, dla którego przedstawia się on w sposób przedstawiony w tab. 4 (zastosowano uproszczoną transkrypcję fonetyczną).

Tab. 4. System zaimków osobowych języka perskiego

	Liczba pojedyncza	Liczba mnoga
<b>Osoba pierwsza</b>	man	ma
<b>Osoba druga</b>	to	szoma
<b>Osoba trzecia</b>	o	anha

Ponieważ opracowywany przez autorów system typu Human-Aided Machine Translation jest przewidziany jako system wielojęzyczny z możliwością jego dalszej rozbudowy o kolejne języki, jest rzeczą oczywistą, że musi on uwzględniać wszelkie możliwe niuanse dotyczące systemów zaimków osobowych występujących w różnych językach świata, czego wyrazem jest swoiste menu wyboru formy zaimka osobowego, które zostało przedstawione na rys. 7.

Z kolei w przypadku, gdy użytkownik systemu HAMT podczas wyboru podmiotu zdania zdecyduje się na wybór frazy rzeczownikowej, wówczas ekran monitora powinien przyjąć widok, jak przedstawiono na rys. 8.

W tym wypadku oprócz wyboru odpowiedniej frazy rzeczownikowej z rozwijanego menu, użytkownik może dodatkowo zdecydować, czy fraza ta ma być poprzedzona rodzajnikiem nieokreślonym (jest to opcja domyślna) lub rodzajnikiem określonym, oczywiście w przypadku, gdy w danym języku źródłowym bądź docelowym przekładu stosowanie tego rodzaju rodzajników jest dopuszczalne.

Ponadto użytkownik może poprzedzić wybraną przez siebie frazę rzeczownikową odpowiednim zaimkiem dzierżawczym lub zaimkiem wskazującym, przy czym warty zwrócenia bliższej uwagi jest fakt, że w projektowanym systemie komputerowego przekładu uwzględniono trójstopniowy system zaimków wskazujących, co sprowadza się do tego, że istnieją odrębne formy tego rodzaju zaimków dla obiektów znajdujących się w bezpośredniej bliskości osoby, która mówi, obiektów nieco od niej oddalonych oraz obiektów bardzo odległych. Tego rodzaju trójstopniowy system zaimków wskazujących występuje na przykład w języku japońskim, gdzie spotykane są odpowiednio formy: *kore*, *sore*, *are*, na oznaczenie znajdujących się coraz dalej od mówcy przedmiotów lub osób. Także w języku hiszpańskim

występują trzy formy zaimków wskazujących, które podlegają dodatkowo odmianie przez rodzaj gramatyczny – męski lub żeński: *esto/esta, eso/esa, aquello/aquella*.



Rys. 7. Widok ekranu monitora w przypadku wyboru frazy rzeczownikowej jako podmiotu tłumaczonego zdania

Z rys. 7 wynika ponadto, że użytkownik może jeszcze dodatkowo poprzedzić wybraną przez siebie frazę rzeczownikową liczebnikiem (występującym wyłącznie w zapisie cyfrowym), przy czym warty zwrócenia uwagi jest fakt, że wcześniejszy wybór opcji „Rodzajnik nieokreślony” uniemożliwia dalsze wprowadzenie liczebnika i na odwrót, wybór opcji „Liczebnik” automatycznie blokuje wybór opcji „Rodzajnik nieokreślony”. Natomiast nie ma żadnych przeszkód, aby fraza rzeczownikowa poprzedzona rodzajnikiem określonym, zaimkiem dzierżawczym bądź zaimkiem wskazującym była dodatkowo uzupełniona przez liczebnik.

W roli fraz rzeczownikowych występować mogą zarówno pojedyncze rzeczowniki, jak na przykład „pies”, „kot”, drzewo”, „jezioro”, „autobus” itp., jak również wyrazy wielosegmentowe, takie jak na przykład „okręt wojenny”, „podnośnik hydrauliczny”, „elektrownia atomowa”, „ tranzystor unipolarny”, „silnik prądu stałego”, „pociąg towarowy”, „chłodnia kominowa”, „procesor wielordzeniowy”, „wirusowe zapalenie opon mózgowych”, „przekład symultaniczny” itp. Ponieważ opracowywany system przewidziany jest jako system otwarty leksykalnie, dlatego istnieje możliwość systematycznego wzbogacania jego warstwy leksykalnej o nowe jednostki (zarówno pojedyncze rzeczowniki, jak i wyrazy wielosegmentowe o różnej długości).

W przypadku, gdy użytkownik systemu HAMT zdecyduje się poprzedzić wybraną przez niego frazę rzeczownikową zaimkiem dzierżawczym, ekran monitora powinien przyjąć widok, jak pokazano na rys. 8. W tym miejscu warto zwrócić uwagę, że w przypadku niektórych języków, takich jak na przykład język arabski, zaimki dzierżawcze nie będą poprzedzały wybranej przez użytkownika frazy rzeczownikowej, lecz będą dołączane na zasadzie aglutynacji do ostatniego jej elementu.

Jak wynika z rys. 8, wybór odpowiedniej formy zaimka dzierżawczego odbywa się na identycznej zasadzie jak wybór zaimka osobowego, ponieważ w przypadku różnych języków świata występuje odmienny sposób zróżnicowania tego rodzaju zaimków ze względu na osobę, liczbę i rodzaj gramatyczny.

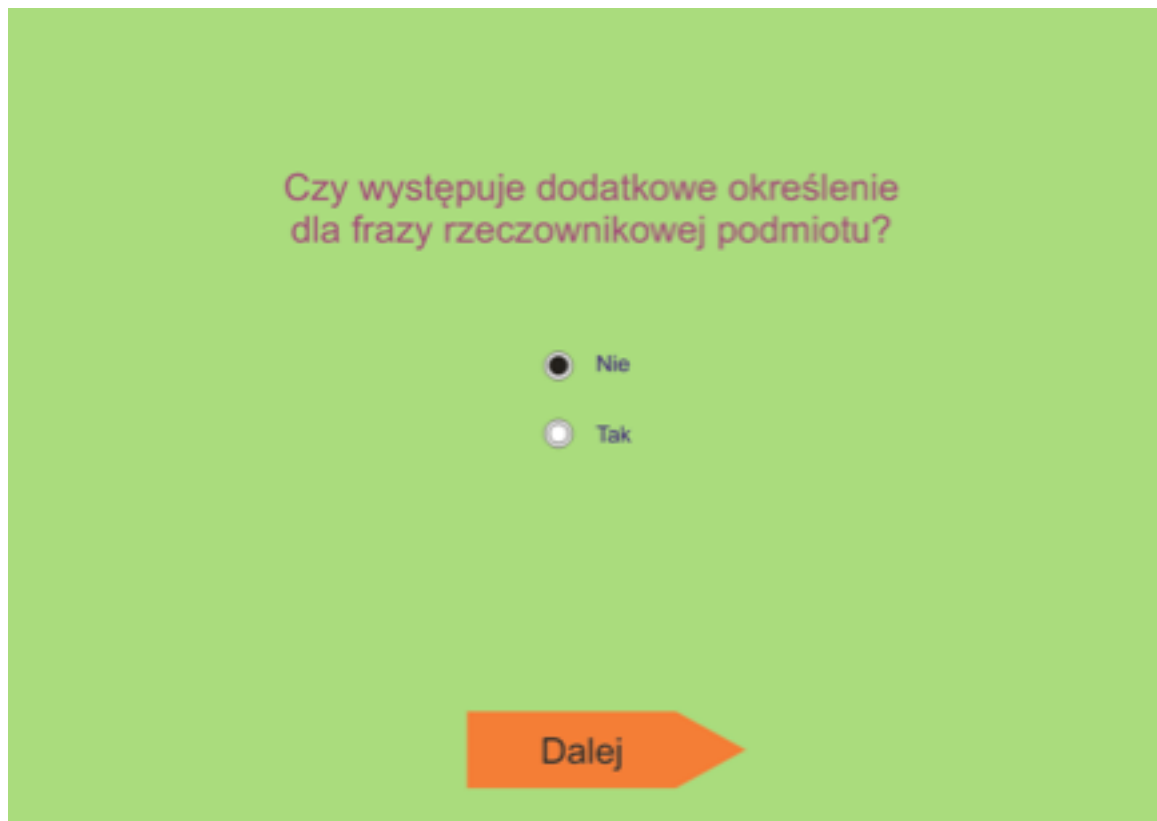
## Zaimki dzierżawcze

	Liczba pojedyncza	Liczba podwójna	Liczba mnoga
Osoba pierwsza	<input checked="" type="radio"/> rodzaj męski <input type="radio"/> rodzaj żeński	<input type="radio"/> rodzaj męski <input type="radio"/> rodzaj żeński	<input type="radio"/> rodzaj męski <input type="radio"/> rodzaj żeński
Osoba druga	<input type="radio"/> rodzaj męski <input type="radio"/> rodzaj żeński	<input type="radio"/> rodzaj męski <input type="radio"/> rodzaj żeński	<input type="radio"/> rodzaj męski <input type="radio"/> rodzaj żeński
Osoba trzecia	<input type="radio"/> rodzaj męski <input type="radio"/> rodzaj żeński <input type="radio"/> rodzaj nijaki	<input type="radio"/> rodzaj męski <input type="radio"/> rodzaj żeński <input type="radio"/> rodzaj nijaki	<input type="radio"/> rodzaj męski <input type="radio"/> rodzaj żeński <input type="radio"/> rodzaj nijaki

**Dalej** 

Rys. 8. Widok ekranu monitora w przypadku poprzedzenia frazy rzeczownikowej zaimkiem dzierżawczym

Po wybraniu przez użytkownika odpowiedniej frazy rzeczownikowej występującej w roli podmiotu zdania ekran monitora przyjmuje widok, jak przedstawiono na rys. 9.



Rys. 9. Widok ekranu monitora po wyborze przez użytkownika odpowiedniej frazy rzeczownikowej występującej w roli podmiotu zdania

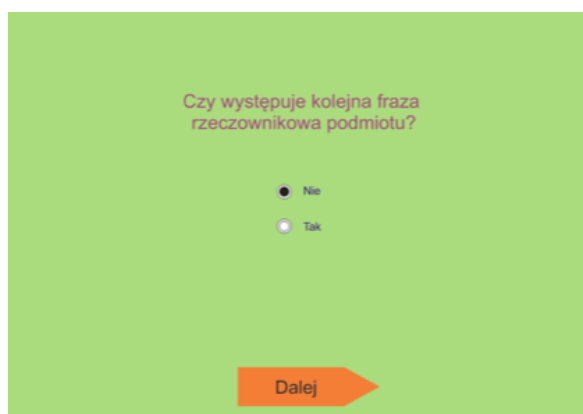
Na rozważanym etapie użytkownik systemu HAMT musi podjąć decyzję, czy wybrana przez niego fraza rzeczownikowa posiada dodatkowe określenie w postaci innej frazy rzeczownikowej. Domyślną odpowiedzią na zamieszczone na rys. 9 pytanie jest „Nie”. Jednak w przypadku, gdy użytkownik wybierze odpowiedź „Tak”, wówczas ekran monitora przyjmie wygląd, jak przedstawiono na rys. 10.

W takim wypadku użytkownik będzie musiał dokonać wyboru kolejnej frazy rzeczownikowej, stanowiącej określenie uprzednio wybranej przez niego frazy rzeczownikowej podmiotu.



Rys. 10. Widok ekranu monitora w przypadku wyboru przez użytkownika frazy rzeczownikowej stanowiącej określenie podmiotu zdania

W etapie kolejnym użytkownik musi podjąć decyzję odnośnie tego, czy w budowanym przez niego zdaniu występuje kolejna fraza rzeczownikowa pełniąca rolę podmiotu zdania. W takim wypadku ekran monitora przyjmuje widok, jak przedstawiono na rys. 11, przy czym domyślną odpowiedzią na tak postawione pytanie jest „Nie”.

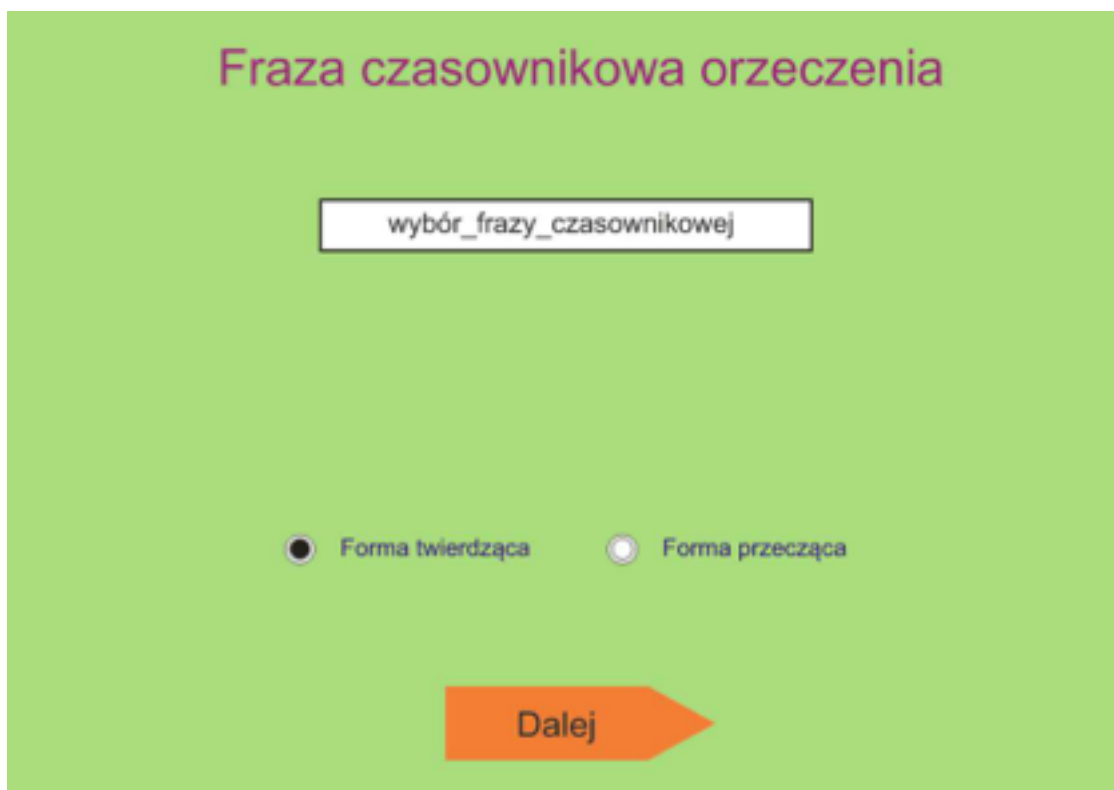


Rys. 11. Widok ekranu monitora w przypadku podejmowania przez użytkownika decyzji odnośnie występowania kolejnej frazy rzeczownikowej stanowiącej podmiot zdania

W przypadku dokonania przez użytkownika wyboru odpowiedzi „Tak” istnieje możliwość budowy zdań, w przypadku których w roli podmiotu występuje cały ciąg fraz rzeczownikowych, przy czym wszystkie elementy takiego ciągu są oddzielane od siebie za pomocą przecinków, natomiast dwa ostatnie jego elementy łączone są przez system domyślnie spójnikiem „i”. W ten sposób w roli podmiotu mogą wystąpić na przykład następujące ciągi fraz: „psy, koty, konie i mały”, „lokomotywy elektryczne, lokomotywy

spalinowe i lokomotywy parowe”, „ tranzystory bipolarne, tranzystory unipolarne, termistory, warystory i diody półprzewodnikowe” itp.

W etapie kolejnym użytkownik systemu HAMT musi dokonać wyboru frazy czasownikowej pełniącej rolę orzeczenia zdania. W przypadku, gdy na etapie pokazanym na rys. 4 użytkownik dokonał wyboru formy trybu rozkazującego, czyli jednej z opcji: „Tryb rozkazujący dla frazy czasownikowej nieprzechodniej”, „Tryb rozkazujący dla frazy czasownikowej przechodniej z dopełnieniem bliższym” lub „Tryb rozkazujący dla frazy czasownikowej przechodniej z dopełnieniem bliższym i dalszym”, wówczas ekran monitora przyjmie widok, jak pokazano na rys. 12.

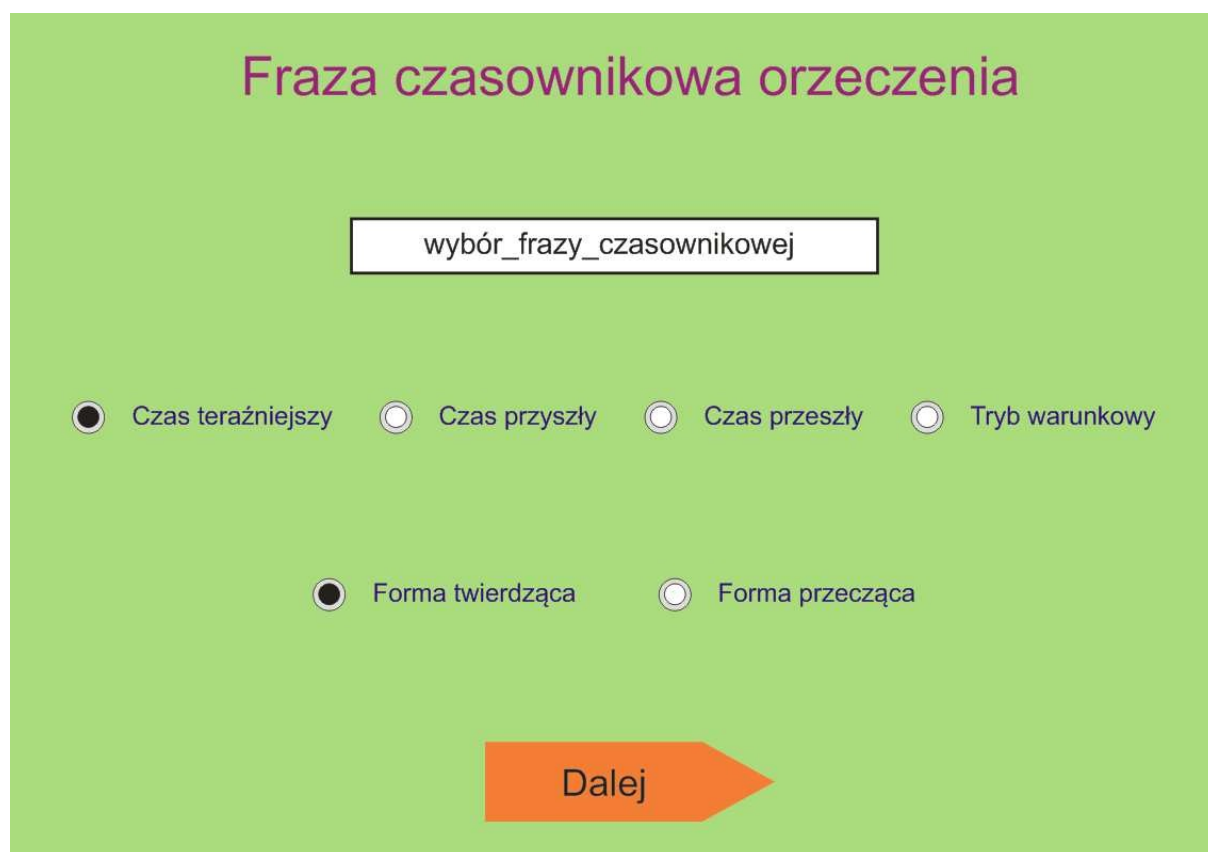


Rys. 12. Widok ekranu monitora w przypadku wyboru frazy czasownikowej pełniącej rolę orzeczenia zdania w formie trybu rozkazującego

Jak widać na rys. 12, na omawianym etapie użytkownik musi dodatkowo określić, czy wybrana przez niego fraza czasownikowa występuje w formie twierdzącej czy przeczącej.

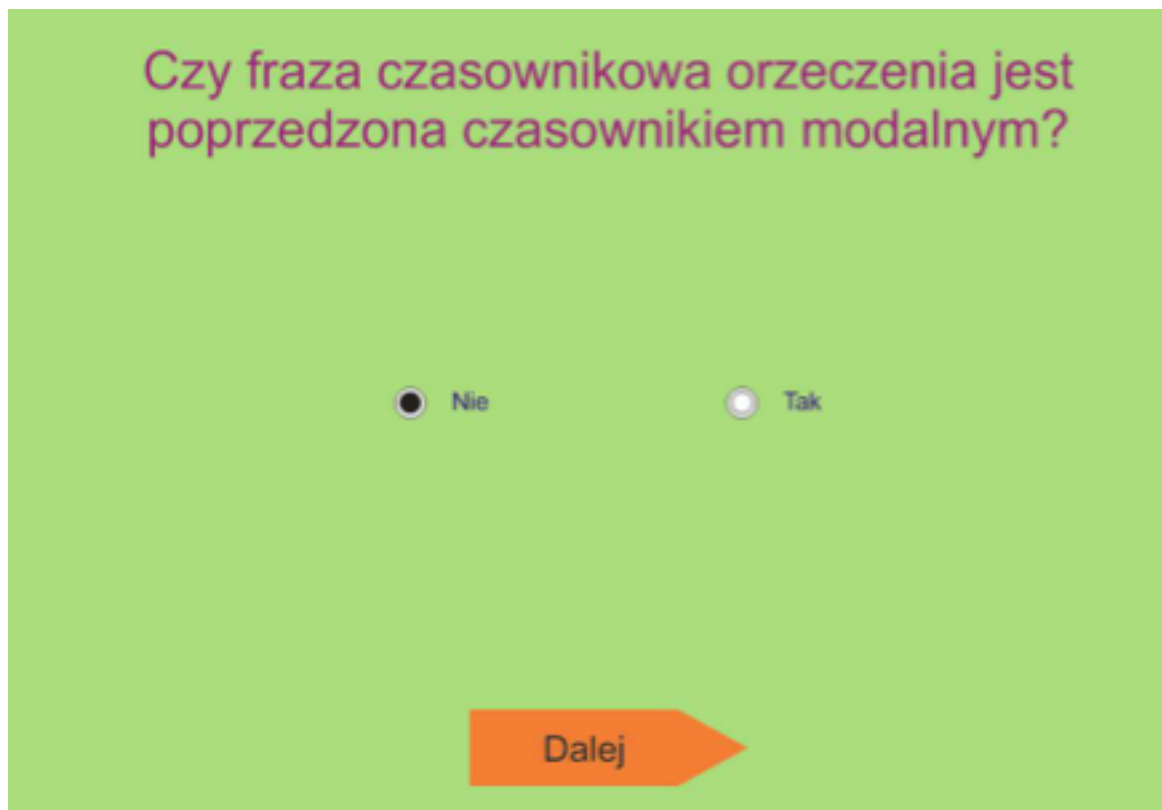
Z kolei w przypadku dokonania przez użytkownika na etapie przedstawionym na rys. 4 wyboru jednej z opcji: „Fraza czasownikowa nieprzechodnia”, „Fraza czasownikowa przechodnia z dopełnieniem bliższym”, „Fraza czasownikowa przechodnia z dopełnieniem bliższym i dalszym” lub „Forma strony biernej” ekran monitora powinien przyjąć wygląd, jak pokazano na rys. 13.

Jak wynika z rys. 13 w rozważanym przypadku użytkownik powinien dokonać jeszcze wyboru czasu, w którym wystąpić ma dana fraza czasownikowa. W przypadku języka polskiego użytkownik może dokonać wyboru pomiędzy formami czasu teraźniejszego, czasu przyszłego i czasu przeszłego, ewentualnie może zdecydować się na wybór formy trybu warunkowego. Oczywiście, w przypadku innych języków świata występujących w roli języka źródłowego przekładu system czasów gramatycznych może być o wiele bardziej rozbudowany. Ponieważ jednak tłumaczenie w systemie dokonywane jest poprzez język pośredniczący przekładu (interlingua), bardziej rozbudowane systemy czasów, które występują w przypadku niektórych języków ulegną redukcji do systemu jedynie sześciu czasów i trybów, do których należy: „czas teraźniejszy trybu orzekającego”, „czas przyszły trybu orzekającego”, „czas przeszły trybu orzekającego”, „czas zaprzeczony trybu orzekającego”, „czas przeszły trybu warunkowego”, a także „czas zaprzeczony trybu warunkowego”.



Rys. 13. Widok ekranu monitora w przypadku wyboru frazy czasownikowej pełniącej rolę orzeczenia zdania z opcja wyboru czasów i trybów gramatycznych

W etapie kolejny użytkownik systemu HAMT musi określić, czy wybrana przez niego fraza czasownikowa pełniąca rolę podmiotu zdania ma być dodatkowo poprzedzona jednym z czasowników modalnych, co pokazano na rys. 14. Jak widać, domyślną odpowiedzią na tak postawione pytanie jest „Nie”.



Rys. 14. Widok ekranu monitora w przypadku podejmowania przez użytkownika decyzji odnośnie poprzedzenia frazy czasownikowej pełniącej funkcję orzeczenia zdania czasownikiem modalnym

W przypadku udzielenia przez użytkownika na zamieszczone na rys. 14 pytanie odpowiedzi „Tak”, ekran monitora powinien przybrać wygląd, jak przedstawiono na rys. 15.



## Wybór czasownika modalnego

- Chcieć       Ośmielać się       Móc
- Potrzebować       Potrafić       Musieć

Dalej

Rys. 15. Widok ekranu monitora w przypadku wyboru czasownika modalnego poprzedzającego frazę czasownikową pełniącą rolę orzeczenia zdania

Jak wynika z rys. 15, na etapie planowania systemu zakłada się, że użytkownik systemu HAMT będzie miał do wyboru jeden z sześciu różnych czasowników modalnych, którymi może poprzedzić wybraną uprzednio frazę czasownikową. Do rozważanych czasowników modalnych należą takie czasowniki, jak: „chcieć”, „potrzebować”, „ośmielać się”, „potrafić”, „móc” oraz „musieć”. Jak już uprzednio wspomniano, opracowywany system przekładu komputerowego wspomaganego przez człowieka jest planowany jako system otwarty składniowo, w związku z czym istnieje możliwość wzbogacenia w przyszłości repertuaru dostępnych czasowników modalnych o nowe jednostki leksykalne.

W dalszej kolejności użytkownik systemu HAMT musi dokonać wyboru formy tłumaczonego zdania, co pokazano na rys. 16.

## Wybór formy tłumaczonego zdania

- Zdanie w formie twierdzącej
- Zdanie w formie pytającej

Dalej

Rys. 16. Widok ekranu monitora w przypadku wyboru formy tłumaczonego zdania

Jak wynika z rys. 15, użytkownik może dokonać wyboru pomiędzy zdaniem w formie twierdzącej bądź pytającej, przy czym domyślną opcją wyboru jest forma twierdząca.

Z kolei w przypadku zaznaczenia przez użytkownika opcji związanej z wyborem zdania w formie pytającej ekran monitora powinien przybrać widok, jak pokazano na rys. 17.

## Wybór rodzaju pytania

- Pytanie ogólne
- Pytanie szczegółowe

Dalej

Rys. 17. Widok ekranu monitora w przypadku wyboru rodzaju pytania

Jak widać na rys. 17, na rozważanym etapie użytkownik musi podjąć decyzję, czy tworzone przez niego zdanie pytające będzie mieściło się w kategorii pytań ogólnych, czy też pytań szczegółowych. Opcją domyślną jest wybór pytania ogólnego, natomiast w przypadku dokonania wyboru pytania szczegółowego ekran monitora przyjmie widok jak na rys. 18.



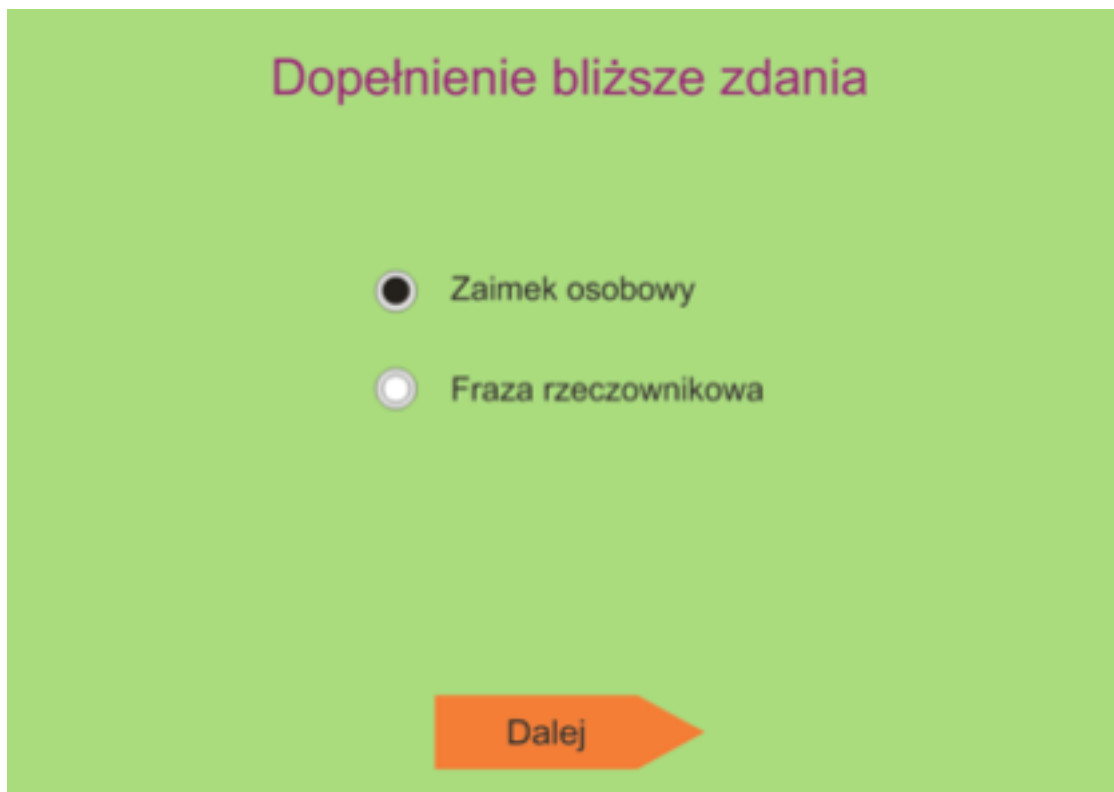
Rys. 18. Widok ekranu monitora w przypadku wyboru rodzaju partykuły pytającej

Jak wynika z rys. 18, podczas budowy pytania szczegółowego użytkownik ma do wyboru sześć różnych partykuł pytających, takich jak: „Kto?”, „Co?”, „Gdzie?”, „Kiedy?”,

„Jak?” i „Dlaczego?”. Za ich pomocą w rozważanym systemie HAMD można zbudować na przykład następujące zdania występujące w formie pytającej: „Kto mieszka w mieście?”, „Co zostało ukradzione?”, „Gdzie twój brat znalazł te klucze?”, „Kiedy oni dostaną pieniądze?”, „Jak samochód został naprawiony?”, „Dlaczego on nie przyjechał do was?” itp.

Jak już wcześniej wspomniano, opracowywany system HAMD pomyślany jest jako system otwarty składniowo, w związku z czym na każdym etapie jego rozwoju istnieje możliwość poszerzenia dostępnego repertuaru partykuł pytających o nowe jednostki leksykalne.

Kolejny etap interakcji użytkownika z systemem HAMD polega na dokonaniu wyboru formy dopełnienia bliższego zdania, jeśli oczywiście uprzednio wybrana przez użytkownika fraza czasownikowa pełniąca rolę orzeczenia zdania wymaga użycia tego rodzaju dopełnienia. Jak wynika z rys. 19 w roli dopełnienia bliższego zdania może wystąpić albo zaimek osobowy, albo fraza rzeczownikowa.



Rys. 19. Widok ekranu monitora w przypadku wyboru formy dopełnienia bliższego zdania

W przypadku wybrania przez użytkownika domyślnej opcji „Zaimek osobowy” ekran monitora powinien przybrać wygląd, jak przedstawiono na rys. 8, dzięki czemu istnieje możliwość wyboru odpowiedniej formy zaimka osobowego ze względu na osobę, liczbę i rodzaj gramatyczny.

Jeśli oba zaimki osobowe występujące w roli zarówno podmiotu, jak i dopełnienia bliższego zdania występują w trzeciej osobie gramatycznej oraz w tej samej liczbie i rodzaju gramatycznym, wówczas ekran monitora powinien przyjąć wygląd, jak pokazano na rys. 20.

W tym wypadku chodzi o rozstrzygnięcie, czy dany zaimek osobowy występujący w roli dopełnienia bliższego zdania nie jest przypadkiem zaimkiem zwrotnym. W ten sposób można wprowadzić rozróżnienie pomiędzy zdaniami takimi jak na przykład: „On widzi jego” i „On widzi siebie” bądź „Oni okłamują ich” i „Oni okłamują siebie”, które w każdym wypadku znaczą przecież coś zgoła odmiennego.

Jak wynika z rys. 20, jako domyślna opcja przyjmowane jest, że wybrany przez użytkownika zaimek osobowy nie jest zaimkiem zwrotnym, czyli w roli podmiotu i dopełnienia bliższego zdania występują fizycznie różne osoby. Natomiast w sytuacji gdy zarówno w roli podmiotu, jak i dopełnienia bliższego zdania ma wystąpić ta sama osoba,

wówczas użytkownik systemu HAMT powinien z menu przedstawionego na rys. 20 wybrać odpowiedź „Tak”.

Czy zaimek osobowy występujący w trzeciej osobie jest zaimkiem zwrotnym?

Nie

Tak

Dalej

Rys. 20. Widok ekranu monitora w przypadku podejmowania przez użytkownika systemu decyzji, czy zaimek osobowy występujący w trzeciej osobie jest zaimkiem zwrotnym

Natomiast w sytuacji, gdy użytkownik zdecyduje się na wybór opcji „Fraza rzeczownikowa”, wówczas ekran monitora powinien przybrać wygląd, jak pokazano na rys. 21.

## Fraza rzeczownikowa dopełnienia bliższego

wybór\_frazy\_rzeczownikowej

Rodzajnik nieokreślony       Rodzajnik określony       Zaimek dzierżawczy

Zaimek wskazujący (obiekt znajdujący się bliżej)       Zaimek wskazujący (obiekt znajdujący się dalej)       Zaimek wskazujący (obiekt bardzo odległy)

Liczebnik

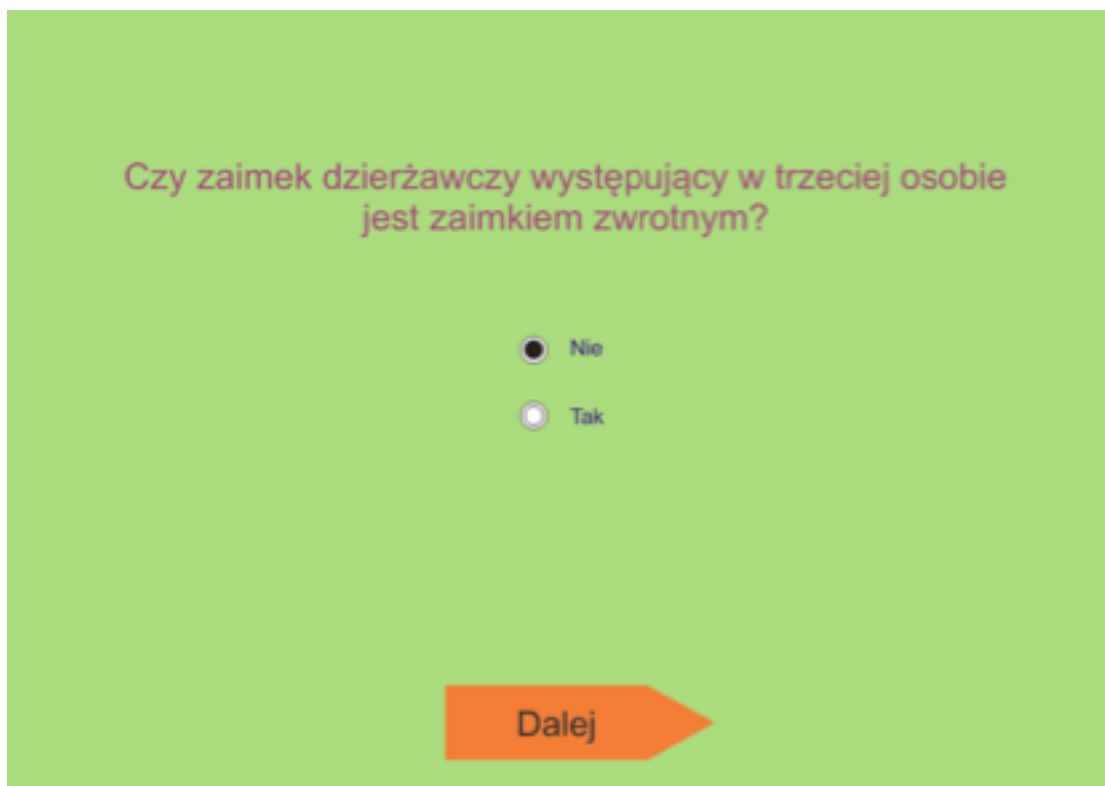
Dalej

Rys. 21. Widok ekranu monitora w przypadku wyboru frazy rzeczownikowej pełniącej rolę dopełnienia bliższego zdania

Jak wynika z rys. 21, podobnie jak miało to miejsce podczas wyboru frazy rzeczownikowej pełniącej rolę podmiotu zdania, również i w tym wypadku użytkownik ma możliwość poprzedzenia wybranej przez siebie z menu frazy rodzajnikiem nieokreślonym lub

rodzajnikiem określonym, oczywiście jeśli w danym języku takie kategorie gramatyczne istnieją. Podobnie fraza rzeczownikowa pełniąca rolę dopełnienia bliższego zdania może zostać poprzedzona zaimkiem dzierżawczym, jednym z trzech możliwych do wyboru zaimków wskazujących i dodatkowo zapisanym w formie cyfrowej liczebnikiem.

Jeżeli użytkownik zdecydował się poprzedzić frazę rzeczownikową pełniącą rolę dopełnienia bliższego zdania zaimkiem dzierżawczym występującym w trzeciej osobie, a w roli podmiotu zdania wystąpił zaimek osobowy również w trzeciej osobie i w tej samej liczbie i rodzaju gramatycznym, wówczas ekran monitora przyjmie wygląd, jak pokazano na rys. 22.



Rys. 22. Widok ekranu monitora w przypadku podejmowania przez użytkownika systemu decyzji, czy zaimek dzierżawczy występujący w trzeciej osobie jest zaimkiem zwrotnym

Jak wynika z rys. 22, opcją domyślną jest odpowiedź „Nie”. Natomiast, gdy rozważany zaimek osobowy ma być zaimkiem zwrotnym, wówczas użytkownik musi zaznaczyć odpowiedź „Tak”.

Dzięki wprowadzeniu tego rodzaju rozróżnienia możliwy jest za pomocą opracowywanego systemu typu HAMA prawidłowy przekład zdań w języku angielskim, takich jak na przykład „He took his car”. W zależności od rodzaju decyzji podjętej przez użytkownika na etapie przedstawionym na rys. 22 rozważane angielskie zdanie zostanie przetłumaczone na język polski albo jako „On wzięł jego samochód”, albo jako „On wzięł swój samochód”.

Jak widać, w rozważanym przypadku język polski jest bardziej szczegółowy w opisie danego fragmentu rzeczywistości w porównaniu z językiem angielskim, stąd na etapie projektowania systemu HAMA konieczne było wprowadzenie tego rodzaju rozróżnienia w zakresie tłumaczenia zaimków dzierżawczych.

W przypadku gdy użytkownik systemu zdecydował się na wybór odpowiedniej frazy rzeczownikowej występującej w roli dopełniania bliższego zdania, wówczas w kroku kolejnym musi podjąć decyzję, czy wybrana przez niego fraza posiada dodatkowe określenie. Tego rodzaju wybór dokonywany jest za pomocą menu przedstawionego na rys. 23.

Na przykład, jeśli użytkownik wybrał frazę „samochód sportowy”, jako jej dodatkowe określenie może następnie wybrać przykładowo frazę „mój starszy brat”, w związku z czym zostanie utworzona

**Centrum Inteligentnych Systemów Informatycznych** Akademia Górniczo-Hutnicza im. Stanisława Staszica al. Mickiewicza 30, 30-059 Kraków  
budynek C-2 pokój 426 tel.: 12 617 44 53 www.isi.agh.edu.pl [isi@agh.edu.pl](mailto:isi@agh.edu.pl)



dłuższa fraza „samochód sportowy mojego starszego brata”, która wystąpi w roli dopełnienia bliższego zdania.



Czy występuje dodatkowe określenie  
dla frazy rzeczownikowej  
dopełnienia bliższego?

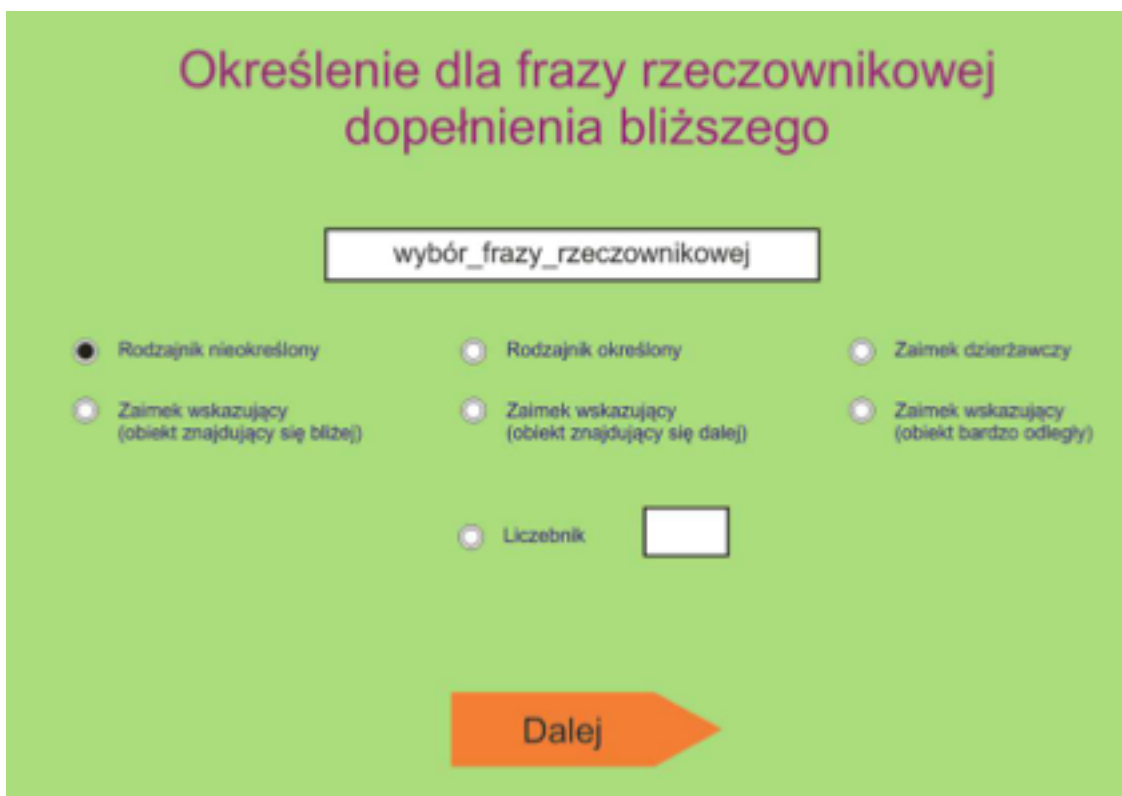
Nie

Tak

Dalej

Rys. 23. Widok ekranu monitora w przypadku podejmowania przez użytkownika systemu decyzji odnośnie występowania dodatkowego określenia dla frazy rzeczownikowej występującej w roli dopełnienia bliższego

Z kolei określenie dla frazy rzeczownikowej dopełnienia bliższego wybierane jest za pomocą menu przedstawionego na rys. 24.



Rys. 24. Widok ekranu monitora w przypadku dokonywania przez użytkownika wyboru określenia dla frazy rzeczownikowej dopełnienia bliższego zdania.

Jak wynika z rys. 24, fraza rzeczownikowa stanowiąca dodatkowe określenie dla frazy rzeczownikowej pełniącej rolę dopełnienia bliższego zdania może zostać poprzedzona rodzajnikiem nieokreślonym, rodzajnikiem określonym, zaimkiem dzierżawczym, jednym z dostępnych w danym języku zaimków wskazujących, a ponadto może być dodatkowo uzupełniona liczebnikiem w postaci zapisu cyfrowego.

W kolejnym etapie użytkownik systemu HAMT musi podjąć decyzję, czy występuje w budowanym przez niego zdaniu kolejna fraza rzeczownikowa pełniąca rolę dopełnienia bliższego zdania, co pokazano na rys. 25.

Czy występuje kolejna fraza rzeczownikowa dopełnienia bliższego?

Nie

Tak

Dalej

Rys. 25. Widok ekranu monitora w przypadku podejmowania przez użytkownika decyzji, czy w budowanym przez niego zdaniu występuje kolejna fraza rzeczownikowa w roli dopełnienia bliższego zdania

Jak wynika z rys. 25, domyślną odpowiedzią na tak postawione pytanie jest „Nie”. Jednak w przypadku, gdy użytkownik wybierze opcję „Tak”, możliwe jest budowanie dłuższych fraz, składających się z wielu członów, przy czym dwa ostatnie z nich łączone są automatycznie przez system spójnikiem „i”. Postępując w ten sposób użytkownik systemu HMT może utworzyć na przykład następujące frazy: „grzyby niejadalne i grzyby trujące”, „elektrownie wodne, elektrownie z turbinami gazowymi i elektrownie szczytowo-pompowe”, „języki programowania wysokiego poziomu, języki skryptowe i języki assemblerowe” itp.

## Dopełnienie dalsze zdania

- Zaimek osobowy
- Fraza rzeczownikowa

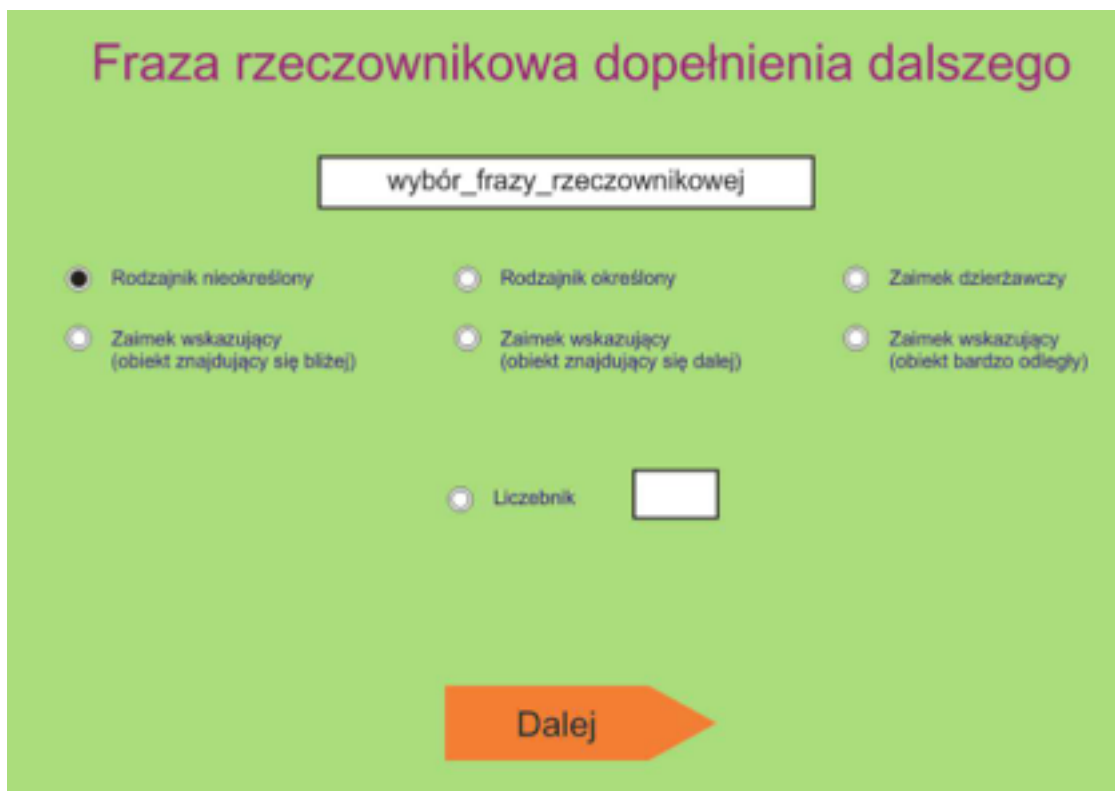
Dalej

Rys. 26. Widok ekranu monitora w przypadku dokonywania przez użytkownika wyboru formy dopełnienia dalszego zdania

W przypadku, gdy użytkownik systemu HAMT na etapie przedstawionym na rys. 4 zdecydował się na wybór w roli orzeczenia zdania frazy czasownikowej wymagającej zarówno dopełnienia bliższego, jak i dalszego, wówczas ekran monitora powinien przybrać wygląd, jak pokazano na rys. 26.

Za pomocą przedstawionego na rys. 26 menu, użytkownik może dokonać wyboru, czy dopełnieniem dalszym zdania ma być fraza rzeczownikowa czy zaimek osobowy.

Fraza rzeczownikowa dopełnienia dalszego zdania wybierana jest za pomocą menu przedstawionego na rys. 27.



Rys. 27. Widok ekranu monitora w przypadku dokonywania przez użytkownika wyboru frazy rzeczownikowej występującej w roli dopełnienia dalszego zdania

Jak widać na rys. 27, fraza rzeczownikowa pełniąca rolę dopełnienia dalszego zdania może zostać poprzedzona rodzajnikiem nieokreślonym, rodzajnikiem określonym, zaimkiem dzierżawczym, jednym z dostępnych w danym języku zaimków wskazujących, a także może zostać dodatkowo uzupełniona liczebnikiem w postaci zapisu cyfrowego.

Jeżeli użytkownik systemu zdecydował się na wybór zaimka dzierżawczego w trzeciej osobie, a podmiot zdania również wystąpił w trzeciej osobie oraz w tej samej liczbie i rodzaju gramatycznym, wówczas ekran monitora przyjmie widok, jak pokazano na rys. 22, ponieważ użytkownik będzie musiał rozstrzygnąć, czy dany zaimek dzierżawczy jest zaimkiem zwrotnym.

W kroku kolejnym użytkownik systemu HAMA musi rozstrzygnąć, czy wybrana przez niego fraza rzeczownikowa pełniąca rolę dopełnienia dalszego zdania posiada dodatkowe określenie w postaci kolejnej frazy rzeczownikowej. Domyślną odpowiedzią na tak postawione pytanie jest „Nie”, a wybór dokonywany jest za pośrednictwem menu przedstawionego na rys. 28.

Czy występuje dodatkowe określenie dla frazy rzeczownikowej dopełnienia dalszego?

Nie

Tak

Dalej

Rys. 28. Widok ekranu monitora w przypadku podejmowania przez użytkownika decyzji odnośnie tego, czy w budowanym przez niego zdaniu występuje dodatkowe określenie dla frazy rzeczownikowej pełniącej rolę dopełnienia dalszego

W przypadku gdy użytkownik systemu podjął decyzję, że fraza okolicznikowa dopełniania dalszego zdania ma posiadać dodatkowe określenie w postaci kolejnej frazy rzeczownikowej, wówczas ekran monitora powinien przybrać wygląd, jak pokazano na rys. 29.

Określenie dla frazy rzeczownikowej  
dopełnienia dalszego

wybór\_frazy\_rzeczownikowej

Rodzajnik nieokreślony

Rodzajnik określony

Zaimek dzierżawczy

Zaimek wskazujący (obiekt znajdujący się bliżej)

Zaimek wskazujący (obiekt znajdujący się dalej)

Zaimek wskazujący (obiekt bardzo odległy)

Liczebnik

Dalej

Rys. 29. Widok ekranu monitora w przypadku wyboru przez użytkownika systemu dodatkowego określenia dla frazy rzeczownikowej pełniącej rolę dopełnienia dalszego zdania

W etapie kolejnym użytkownik systemu musi podjąć decyzję odnośnie tego, czy występuje kolejna fraza rzeczownika pełniąca rolę dopełnienia dalszego budowanego przez niego zdania. W tym wypadku ekran monitora powinien przyjąć widok, jak pokazano na rys. 30.

Czy występuje kolejna fraza  
rzeczownikowa dopełnienia dalszego?

Nie

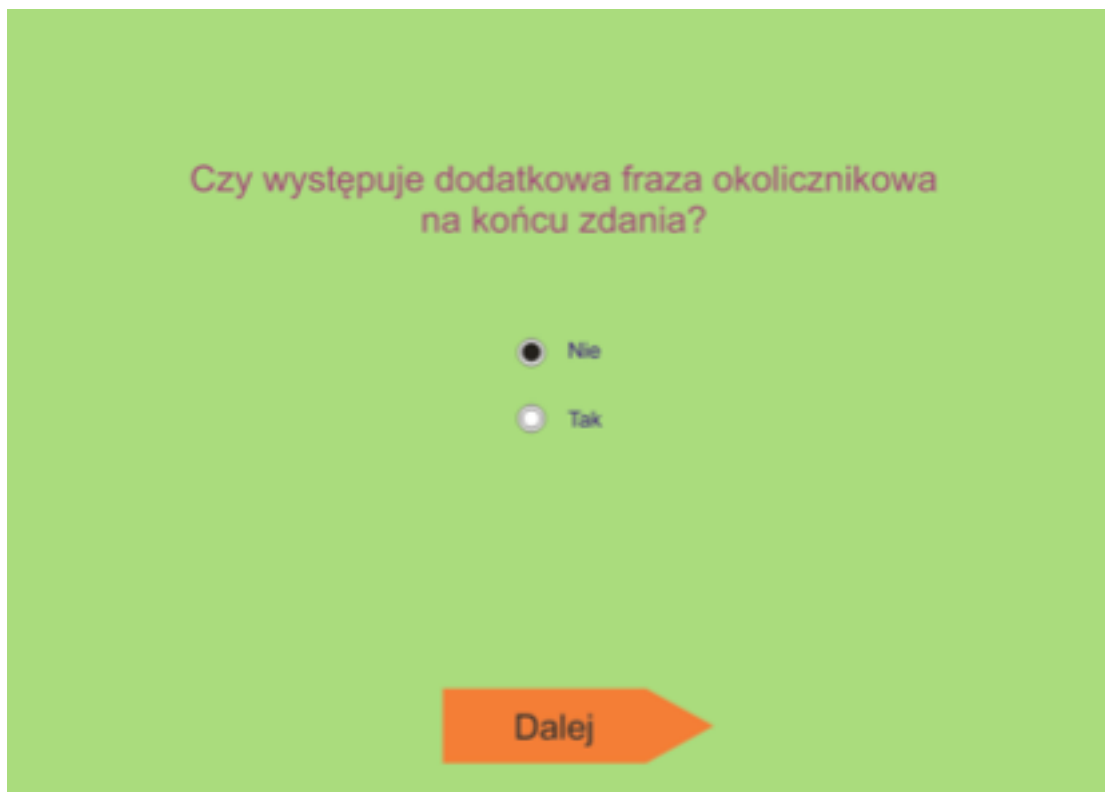
Tak

Dalej

Rys. 30. Widok ekranu monitora w przypadku podejmowania przez użytkownika systemu decyzji odnośnie występowania kolejnej frazy rzeczownikowej w roli dopełnienia dalszego zdania

W etapie kolejnym użytkownik musi wypowiedzieć się odnośnie tego, czy na końcu budowanego przez niego zdania występuje jeszcze jakaś dodatkowa fraza okolicznikowa. Wybór właściwej odpowiedzi dokonywany jest za pośrednictwem menu przedstawionego na rys. 31.





Rys. 31. Widok ekranu monitora w przypadku podejmowania przez użytkownika systemu decyzji odnośnie występowania dodatkowej frazy okolicznikowej na końcu budowanego przez niego zdania

Jak wynika z rys. 31, domyślną odpowiedzią na tak postawione pytanie jest „Nie”. Natomiast w przypadku dokonania przez użytkownika wyboru odpowiedzi „Tak” ekran monitora powinien przybrać wygląd, jak pokazano na rys. 32.

W tym wypadku użytkownik dokonuje wyboru właściwej frazy okolicznikowej za pomocą rozwijanego menu, w którym dostępne są wszelkie frazy okolicznikowe, które zostały uwzględnione w systemie HAMT.

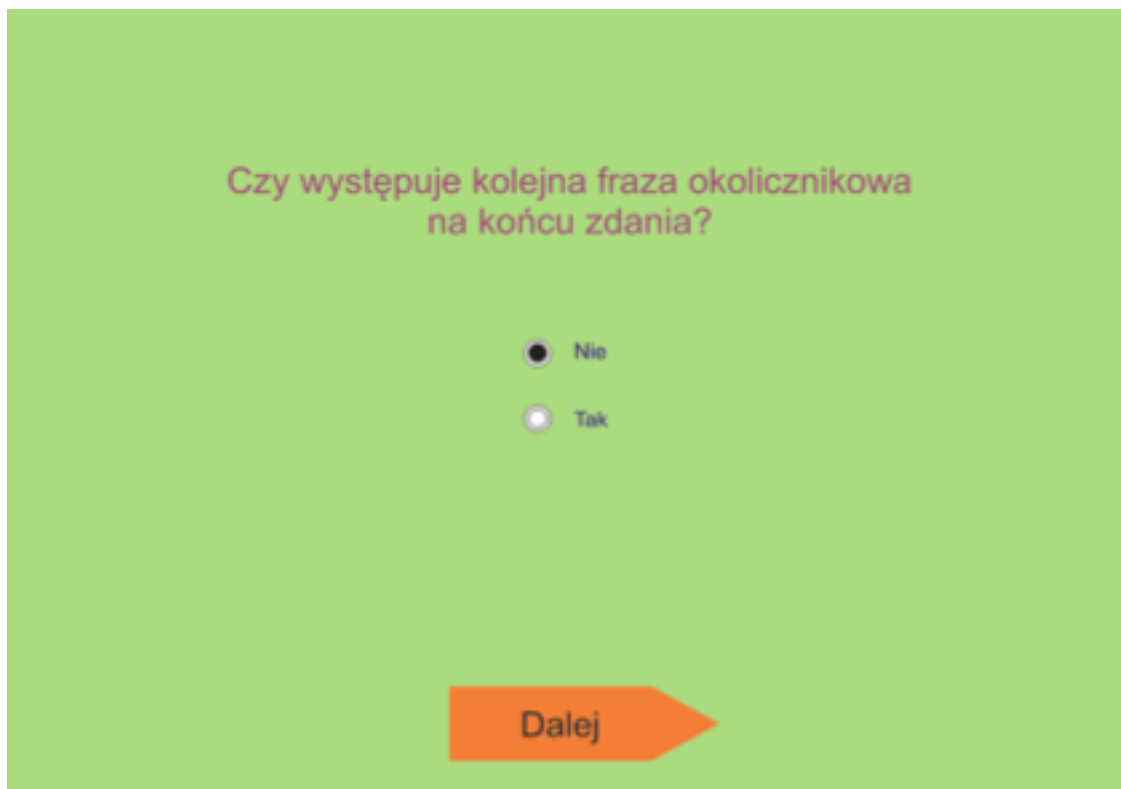
## Fraza okolicznikowa

wybór\_frazy\_okolicznikowej

Dalej

Rys. 32. Widok ekranu monitora w przypadku dokonywania przez użytkownika wyboru frazy okolicznikowej

Po wybraniu odpowiedniej frazy okolicznikowej użytkownik musi jeszcze podjąć decyzję, czy w budowanym przez niego zadaniu występuje kolejna fraza okolicznikowa, co pokazano na rys. 33.



Rys. 33. Widok ekranu monitora w przypadku podejmowania przez użytkownika decyzji, czy na końcu budowanego przez niego zdania występują kolejne frazy okolicznikowe

Jak wynika z rys. 33, domyślną odpowiedzią na tak postawione pytanie jest „Nie”, natomiast w przypadku, gdy użytkownik systemu pragnie wzbogacić tworzone przez siebie zdanie o kolejną frazę okolicznikową, wówczas musi zaznaczyć opcję „Tak”. W takim wypadku ekran monitora powinien przyjąć wygląd, jak pokazano na rys. 32, aby umożliwić wybór kolejnej frazy okolicznikowej.

Jak widać, budowane przez użytkownika zdanie może zostać zakończone ciągiem fraz okolicznikowych o dowolnej długości. W roli fraz okolicznikowych mogą wystąpić różnego typu wyrażenia stanowiące okoliczniki miejsca, czasu, sposobu, przyczyny, celu itp. Jako

przykład tego rodzaju fraz okolicznikowych można podać: „nad jeziorem”, „w lesie”, „w domku letniskowym”, „późnym wieczorem”, „w grudniu zeszłego roku”, „za dwa lata”, „w ogromnym pośpiechu”, „z zażenowaniem”, „z wielką przyjemnością” itp.

Po dokonaniu przez użytkownika wyboru ostatniej frazy okolicznikowej ekran monitora powinien przybrać wygląd, jak pokazano na rys. 34.

Czy występuje kolejne zdanie złożone  
współrzędnie lub podrzędnie?

Nie

Tak

Dalej

Rys. 34. Widok ekranu monitora w przypadku podejmowania przez użytkownika decyzji, czy występuje kolejne zdanie złożone współrzędnie lub podrzędnie.

Na rozważanym etapie użytkownik systemu musi podjąć decyzję, czy budowane przez niego zdanie jest zdaniem prostym, czy też występuje kolejne zdanie złożone z nim współrzędnie lub podrzędnie.

## Wybór elementu łączącego zdania

wybór\_frazy\_łącznikowej

Dalej

Rys. 35. Widok ekranu monitora w przypadku dokonywania przez użytkownika wyboru elementu łączącego zdania złożone współrzędnie bądź podrzędnie

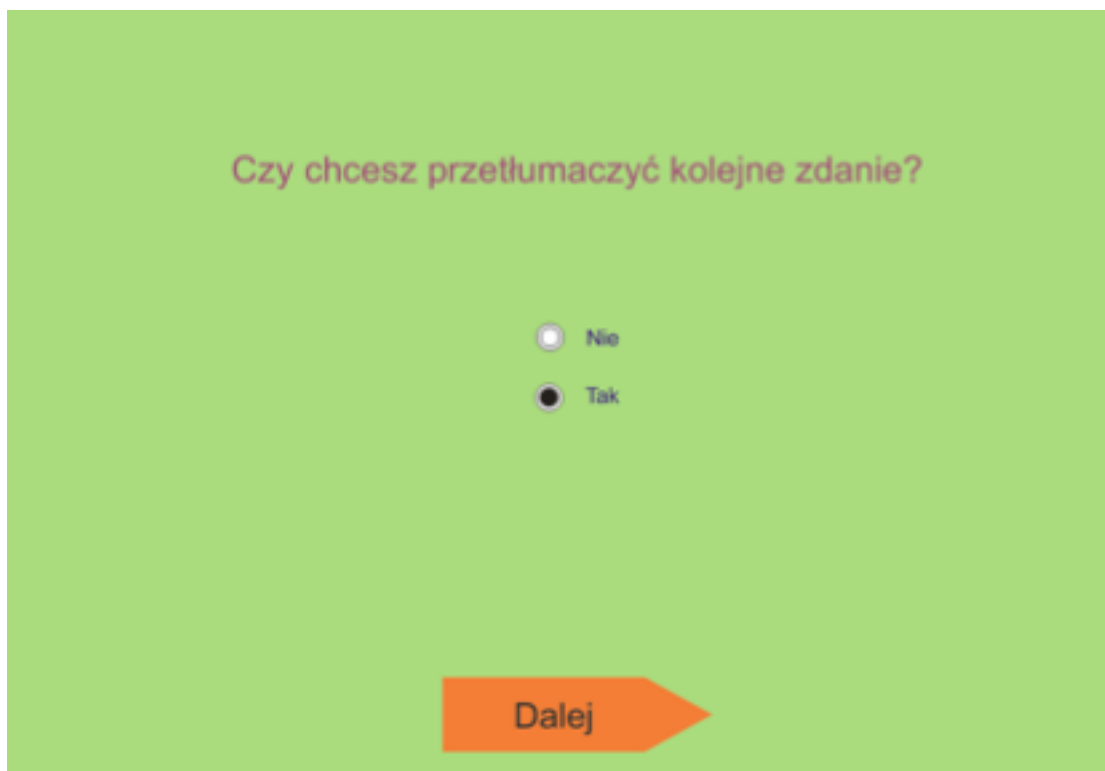
Jak wynika z rys. 34, domyślną odpowiedzią na tak postawione pytanie jest „Nie”. Natomiast w przypadku zaznaczenia przez użytkownika opcji „Tak” ekran monitora powinien przybrać wygląd, jak przedstawiono na rys. 35.

W takim wypadku użytkownik powinien dokonać wyboru z rozwijanego menu elementu łączącego oba zdania. Jako przykład tego rodzaju łączników można wymienić: „i”, „lub”, „albo”, „a”, „ale”, „lecz”, „ponieważ”, „zanim”, „podczas, gdy”, „po tym, jak”, „po to, aby” itp.

Po dokonaniu wyboru elementu łączącego oba zdania użytkownik przechodzi do etapu pokazanego na rys. 4, aby móc zbudować kolejne zdanie złożone współrzędnie bądź podrzędnie ze zbudowanym dotychczas przez niego zdaniem.

Jak widać, postępując w opisany sposób użytkownik ma możliwość utworzenia dowolnej długości ciągu zdań prostych złożonych ze sobą współrzędnie lub podrzędnie.

Po wprowadzeniu wszystkich elementów ostatniego z budowanych zdań złożonych współrzędnie bądź podrzędnie ekran monitora przyjmuje widok, jak pokazano na rys. 36.



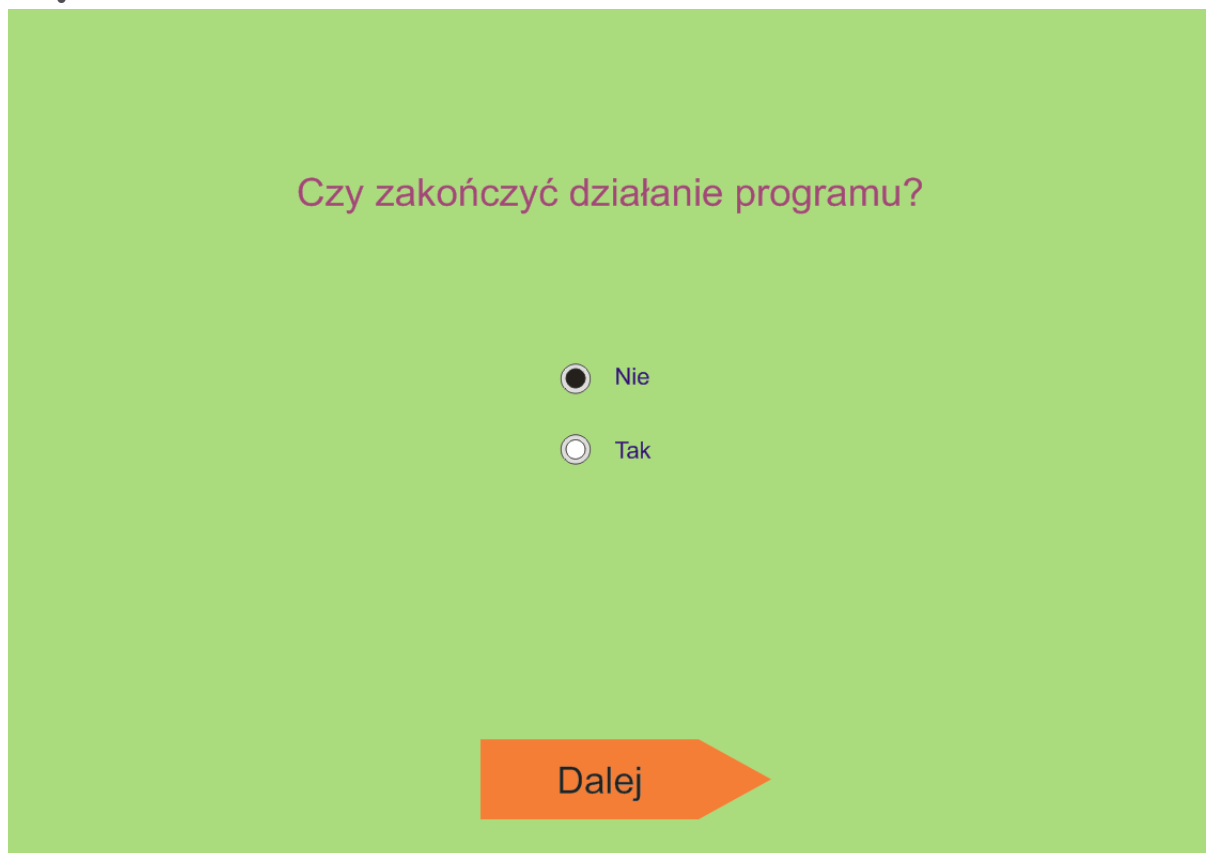
Rys. 36. Widok ekranu monitora w przypadku podejmowania przez użytkownika decyzji, czy za pomocą systemu HAMI chce przetłumaczyć kolejne zdanie proste bądź złożone

Na rozważanym etapie użytkownik musi podjąć decyzję, czy za pomocą systemu HAMI chce przetłumaczyć kolejne zdanie proste bądź złożone. Domyślną odpowiedzią na tak sformułowane pytanie jest „Tak”. W takim wypadku ekran monitora powinien przyjąć wygląd, jak pokazano na rys. 4.

Natomiast, gdy użytkownik nie chce już tłumaczyć za pomocą systemu HAMI kolejnego zdania, wówczas musi dokonać wyboru opcji „Nie”. W takim wypadku ekran monitora przyjmie widok, jak pokazano na rys. 37.

Na rozważanym etapie użytkownik systemu HAMI musi podjąć decyzję, czy chce zakończyć w tym momencie działanie programu. Domyślną odpowiedzią na tak postawione pytanie jest „Nie”. Wówczas nastąpi przejście do etapu pokazanego na rys. 3, gdzie istnieje możliwość dokonania zmiany języka źródłowego przekładu bądź języka docelowego (języków docelowych) przekładu.

Natomiast w przypadku dokonania przez użytkownika wyboru opcji „Tak” system HAMI powinien zakończyć swoje działanie.



Rys. 37. Widok ekranu monitora w przypadku podejmowania przez użytkownika decyzji, czy działanie systemu HAMT ma zostać w tym momencie zakończone

## 5. Mechanizmy leżące u podstaw funkcjonowania systemu HAMT

Jak już we wstępie wspomniano, w przypadku projektowanego systemu komputerowego przekładu typu HAMT tłumaczenie dokonywane jest dwuetapowo. W pierwszej kolejności zbudowane przez użytkownika systemu zdanie w języku kontrolowanym (proste bądź złożone współrzędnie lub podrzędnie) tłumaczone jest na język pośredniczący (tzw. interlingua). Otrzymany przekład wprowadzonego przez użytkownika systemu na jego wejście zdania jest jedynie wewnętrzną sprawą systemu, w związku z czym tekst w języku pośredniczącym przekładu jest dla użytkownika całkowicie nie widoczny, a służy on jedynie do wygenerowania przekładów w wybranych przez użytkownika językach docelowych.

Tak naprawdę użytkownik systemu nie musi być nawet świadomy istnienia języka pośredniczącego przekładu, ponieważ tworzone przez komputer teksty w języku pośredniczącym przekładu są dla użytkownika niedostępne.

Dzięki zastosowaniu koncepcji języka pośredniczącego przekładu opracowywany przez autorów system jest systemem otwartym na dodawanie do niego nowych języków. W tym celu dla każdego z

nowododanych języków należy opracować po jednym programie tłumaczącym z niego na język pośredniczący oraz po jednym programie tłumaczącym z języka pośredniczącego na wybrany język.

Jak już uprzednio wspomniano, istotne podniesienie jakości przekładu można uzyskać poprzez nałożenie swego rodzaju więzów na język źródłowy, poprzez dopuszczanie do użytku jedynie wybranych struktur składniowych i precyzyjne określenie możliwych do użytku zasobów leksykalnych danego języka. Z tego powodu opracowywany przez autorów system jest systemem tłumaczącym teksty zapisane w językach kontrolowanych, które stanowią jedynie pewien podzbiór języków naturalnych, na których są wzorowane. Jest przy tym ważne, aby podzbiór ten był dostatecznie mocny, aby w rozważanych językach kontrolowanych można było wyrazić wszelkie potrzebne treści dotyczące dziedziny, w ramach której dokonywany będzie przekład tekstów.

Opracowywany przez autorów system jest pomyślany jako system otwarty zarówno leksykalnie, jak i składniowo, co oznacza, że zasoby słownikowe systemu mogą być systematycznie uzupełniane o nowe jednostki leksykalne.

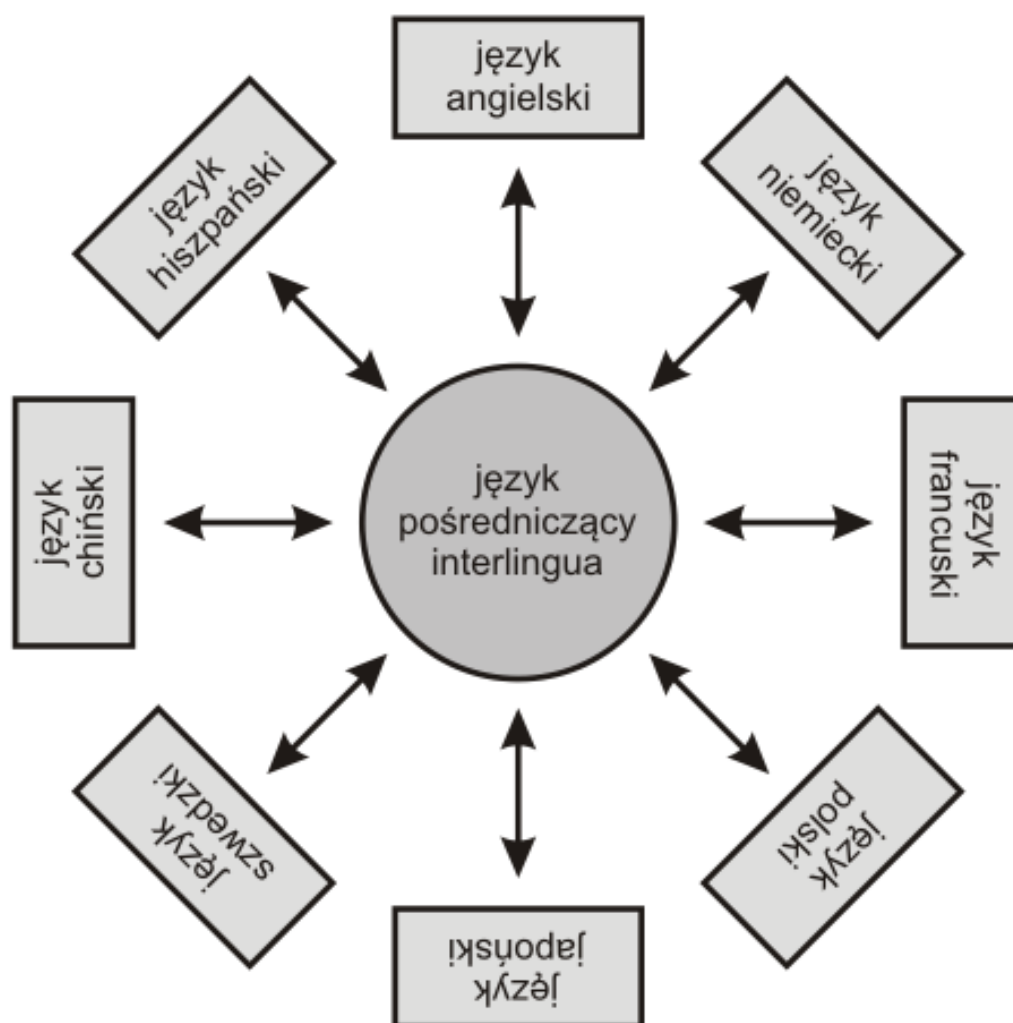
Oczywiście, proces wprowadzania nowych jednostek leksykalnych musi za każdym razem uwzględniać jednocześnie wszystkie języki występujące w systemie. Co więcej, wprowadzenie nowej jednostki leksykalnej może niekiedy wymuszać konieczność modyfikacji leksyki języka pośredniczącego, tak aby przekład z takiego języka był zawsze leksykalnie jednoznaczny z punktu widzenia wszystkich występujących w systemie języków naturalnych.

Chodzi w tym wypadku o to, że niekiedy przekład leksemu jednego języka na drugi jest możliwy jedynie poprzez wykorzystanie hiponimów tego leksemu, w sytuacji gdy język docelowy przekładu jest w tym wypadku bardziej szczegółowy w opisie danego fragmentu rzeczywistości. Na przykład angielski rzeczownik „pipeline” (rurociąg) może zostać przetłumaczony na język francuski jedynie poprzez użycie jednego z dwóch jego hiponimów: „oléoduc” (ropociąg) bądź „gazoduc” (gazociąg), ponieważ ich hiperonim w języku francuskim w ogóle nie występuje. Tego rodzaju przykładów zróżnicowanego poziomu szczegółowości opisu świata w przypadku różnych języków naturalnych można wymienić bardzo wiele, co jest przyczyną sporych trudności występujących podczas prowadzenia prac nad zasobami leksykalnymi języka pośredniczącego przekładu [6, 7].

Z kolei otwarcie składniowe projektowanego przez autorów systemu polega na możliwości wprowadzania dodatkowych struktur składniowych zarówno do języka pośredniczącego przekładu, jak i do poszczególnych języków kontrolowanych występujących w roli języków źródłowych. W swym początkowym etapie rozwoju opracowywany system komputerowego przekładu wspomaganego przez człowieka jest przewidziany jedynie do tłumaczenia zdań prostych typu SVO (ang. Subject – Verb – Object) występujących zarówno w stronie czynnej, jak i biernej. Jednak docelowo reguły składniowe systemu muszą być rozbudowane w ten sposób, aby umożliwić również automatyczny przekład różnego typu zdań złożonych (współrzędnie, podrzędnie, warunkowych, czasowych itp.).



Rys. 38. Ilustracja zasad funkcjonowania wielojęzycznego systemu komputerowego przekładu opartego na języku pośredniczącym (interlingua)



Reasumując, dzięki wprowadzeniu koncepcji języka pośredniczącego przekładu możliwą jest budowa systemu wielojęzycznego komputerowego tłumacza, za pośrednictwem którego można będzie dokonywać przekładów w obu kierunkach pomiędzy dowolnie wybraną parą języków. Idea wykorzystania języka pośredniczącego przekładu została przedstawiona schematycznie na rys. 37.

Planowane jest oparcie leksyki języka pośredniczącego przekładu głównie na jednostkach leksykalnych języka polskiego, które w niektórych wypadkach mogą zostać rozbudowane w razie konieczności o stosowne komentarze doprecyzowujące ich znaczenie. Taka sytuacja będzie występowała każdorazowo, gdy w danym przypadku któryś z języków uwzględnionych dotychczas w systemie HAMT będzie bardziej szczegółowy w opisie rozważanego fragmentu rzeczywistości niż język polski.

Wiedza lingwistyczna potrzebna do funkcjonowania planowanego przez autorów systemu komputerowego przekładu typu HAMT zapisana jest w odpowiednich plikach tekstowych.

**Centrum Inteligentnych Systemów Informatycznych** Akademia Górniczo-Hutnicza im. Stanisława Staszica al. Mickiewicza 30, 30-059 Kraków  
budynek C-2 pokój 426 tel.: 12 617 44 53 www.isi.agh.edu.pl isi@agh.edu.pl

Przewidziane są odrębne pliki dla fraz rzeczownikowych, fraz czasownikowych i fraz typu okolicznikowego, przy czym istnieją odrębne pliki potrzebne do tłumaczenia wymienionych rodzajów fraz z danego języka naturalnego na język pośredniczący i z języka pośredniczącego na dany język naturalny.

W przypadku tłumaczenia fraz rzeczownikowych z języka polskiego na język pośredniczący potrzebna do tego wiedza lingwistyczna zapisana jest w pliku:

### Polish\_INT\_Noun.txt

Przykładowy rekord danych zapisanych w rozważanym pliku może przyjąć następującą postać:

```
*drzewo owocowe$drzewo owocowe$drzewa owocowego$drzewu owocowemu$ drzewo owocowe  
$drzewem owocowym$drzewie owocowym%_drzewo_owocowe&1&3#
```

Każdy tego typu rekord danych zapisany w pliku Polish\_INT\_Noun.txt rozpoczyna się znakiem „\*”, po którym występuje fraza rzeczownikowa, którą użytkownik może wybrać z rozwijanego menu. W rozpatrywanym tutaj przypadku tego typu frazą jest fraza rzeczownikowa „drzewo owocowe”.

Następnie po kolejnych znakach „\$” występują odpowiednie formy gramatyczne rozważanej frazy rzeczownikowej, odpowiadające sześciu kolejnym przypadkom języka polskiego, którymi są odpowiednio: mianownik, dopełniacz, celownik, biernik, narzędnik i miejscownik.

W dalszej kolejności po znaku „%” występuje tłumaczenie rozważanej frazy na język pośredniczący przekładu. W rozpatrywanym przypadku tłumaczeniem tym jest wyrażenie „\_drzewo\_owocowe”. Jak już uprzednio wspomniano, warstwa leksykalna języka pośredniczącego została oparta głównie na jednostkach leksykalnych języka polskiego, przy czym w niektórych przypadkach jednostki te muszą zostać uzupełnione o stosowne komentarze doprecyzowujące ich znaczenie, jeśli któryś z języków uwzględnionych w systemie HAMT tego rodzaju doprecyzowania wymaga.

Dalej po znaku „&” zamieszczona jest informacja dotycząca liczby gramatycznej, w której występuje dana fraza rzeczownikowa języka polskiego, przy czym w tym miejscu mogą wystąpić jedynie dwie różne wartości: 1 lub 2. Pojawienie się tam jedynki oznacza, że dana fraza rzeczownikowa występuje w liczbie pojedynczej, co ma miejsce w rozpatrywanym przykładzie. Natomiast pojawienie się na rozpatrywanej pozycji dwójki oznacza, że dana

fraza rzeczownikowa występuje w liczbie mnogiej. Tego rodzaju informacja jest niezbędna do zachowania związku zgody pomiędzy podmiotem a orzeczeniem zdania.

Po kolejnym znaku „&” występuje informacja dotycząca rodzaju gramatycznego danej frazy rzeczownikowej. Na rozpatrywanym polu mogą pojawić się jedynie trzy różne wartości: 1, 2 lub 3, które odpowiadają rodzajowi męskiemu, żeńskiemu i nijakiemu. Jak widać w rozpatrywanym przypadku w polu tym występuje trójka, ponieważ fraza rzeczownikowa „drzewo owocowe” jest rodzaju nijakiego.

Znak „#” kończy zapis rekordu danych dla rozpatrywanej frazy rzeczownikowej. W przypadku, gdy dana fraza rzeczownikowa może wystąpić także w liczbie mnogiej, w pliku Polish\_INT\_Noun.txt powinny pojawić się jeszcze dwa kolejne rekordy danych:

```
*drzewa owocowe [dwa]$drzewa owocowe$drzew owocowych$drzewom owocowym$ drzewa  
owocowe$drzewami owocowymi$drzewom owocowym%_drzewo_owocowe_D&2&2#
```

\*drzewa owocowe [więcej niż dwa]\$drzewa owocowe\$drzew owocowych\$drzewom owocowym\$  
drzewa owocowe\$drzewami owocowymi\$drzewom owocowym%\_drzewo\_owocowe\_P&2&2#

W związku z powyższym użytkownik systemu HAMT może z rozwijanego menu wybrać frazę rzeczownikową „drzewa owocowe [dwa]” bądź frazę rzeczownikową „drzewa owocowe [więcej niż dwa]”, w zależności od tego, ile tych drzew owocowych ma aktualnie na myśli. Jak już wspomniano, wprowadzenie tego rodzaju rozróżnienia jest konieczne ze względu na języki takie, jak na przykład arabski, sanskryt, słoweński, języki łużyckie (górnołużycki i dolnołużycki), w przypadku których w deklinacji rzeczowników występują odrębne formy fleksyjne dla liczby podwójnej.

W obu rozważanych przykładach po pierwszym znaku „&” występuje dwójka, co wynika z konieczności użycia w języku polskim form fleksyjnych charakterystycznych dla liczby mnogiej, aby pomiędzy frazą stanowiącą podmiot zdania a frazą stanowiącą jego orzeczenie spełniony był związek zgody. Ponadto po drugim znaku „&” również występuje dwójka, co oznacza rodzaj gramatyczny niemęskoosobowy.

Warto jest zwrócić uwagę, że do jednej jednostki leksykalnej w języku pośredniczącym przekładu może prowadzić więcej niż jedno hasło, które użytkownik wybiera z rozwijanego menu, co wyjaśniono na następujących przykładach:

\*śmigłowiec\$śmigłowiec\$śmigłowca\$śmigłowcu\$śmigłowiec\$śmigłowcem\$ śmigłowcu  
%\_śmigłowiec&1&1#

\*śmigłowce [dwa]\$śmigłowce\$śmigłowców\$śmigłowcom\$śmigłowce\$śmigłowcami\$ śmigłowcach  
%\_śmigłowce\_D&2&2#

\*śmigłowce [więcej niż dwa]\$śmigłowce\$śmigłowców\$śmigłowcom\$śmigłowce\$śmigłowcami\$  
śmigłowcach %\_śmigłowce\_P&2&2#

\*helikopter\$helikopter\$helikoptera\$helikopterowi\$helikopter\$helikopterem\$ helikopterze  
%\_śmigłowce&1&1#

\*helikoptery [dwa]\$helikoptery\$helikopterów\$helikopterom\$helikoptery\$helikopterami\$ helikopterach  
%\_śmigłowce\_D&2&2#

\*helikoptery [więcej niż dwa]\$helikoptery\$helikopterów\$helikopterom\$helikoptery\$helikopterami\$  
helikopterach %\_śmigłowce\_P&2&2#

Jak wynika z zamieszczonych powyżej przykładów, do jednostki leksykalnej w języku pośredniczącym przekładu „\_śmigłowiec” mogą prowadzić dwa hasła dostępne w rozwijanym menu wyboru: „śmigłowiec” i „helikopter”.

Podobnie do jednostki leksykalnej „\_śmigłowiec\_D” prowadzi zarówno hasło „śmigłowce [dwa]”, jak i „helikoptery [dwa]”. Także do jednostki leksykalnej „\_śmigłowiec\_P” prowadzą dwa hasła: „śmigłowce [więcej niż dwa]” oraz „helikoptery [więcej niż dwa]”.

W niektórych wypadkach znajdujące się w menu wyboru hasła w języku polskim muszą zostać uzupełnione stosownym komentarzem, który doprecyzowuje ich znaczenia. Dzieje się tak zawsze, gdy dane hasło dostępne z poziomu rozwijanego menu wyboru jest wieloznaczne z punktu widzenia leksyki języka pośredniczącego przekładu. Sytuacje taką przedstawiono na podstawie następujących przykładów:

**Centrum Inteligentnych Systemów Informatycznych** Akademia Górniczo-Hutnicza im. Stanisława Staszica al. Mickiewicza 30, 30-059 Kraków  
budynek C-2 pokój 426 tel.: 12 617 44 53 www.isi.agh.edu.pl [isi@agh.edu.pl](mailto:isi@agh.edu.pl)

\*dziadek (ze strony ojca)\$dziadek ze strony ojca\$dziadka ze strony ojca\$dziadkowi ze strony ojca  
\$dziadka ze strony ojca\$dziadkiem ze strony ojca\$dziadkowi ze strony ojca  
%\_dziadek\_ze\_strony\_ojca&1&1#

\*dziadkowie (ze strony ojców) [dwaj]\$dziadkowie ze strony ojców\$dziadków ze strony ojców\$  
dziadkom ze strony ojców\$dziadków ze strony ojców\$dziadkami ze strony ojców  
\$dziadkach ze strony ojców%\_dziadek\_ze\_strony\_ojca\_D&2&1#

\*dziadkowie (ze strony ojców) [więcej niż dwaj]\$dziadkowie ze strony ojców  
\$dziadków ze strony ojców\$dziadkom ze strony ojców\$dziadków ze strony ojców \$dziadkami ze  
strony ojców\$dziadkach ze strony ojców%\_dziadek\_ze\_strony\_ojca\_P&2&1#

\*dziadek (ze strony matki)\$dziadek ze strony matki \$dziadka ze strony matki \$dziadkowi ze strony  
matki \$dziadka ze strony matki \$dziadkiem ze strony matki \$dziadkowi ze strony matki  
%\_dziadek\_ze\_strony\_matki&1&1#

\*dziadkowie (ze strony matek) [dwaj]\$dziadkowie ze strony matek\$dziadków ze strony matek\$  
dziadkom ze strony matek\$dziadków ze strony matek\$dziadkami ze strony matek  
\$dziadkach ze strony matek%\_dziadek\_ze\_strony\_matki\_D&2&1#

\*dziadkowie (ze strony matek) [więcej niż dwaj]\$dziadkowie ze strony matek  
\$dziadków ze strony matek \$dziadkom ze strony matek \$dziadków ze strony matek \$dziadkami ze  
strony matek \$dziadkach ze strony matek%\_dziadek\_ze\_strony\_matki\_P&2&1#

Jak wynika z zamieszczonych powyżej przykładów hasło języka polskiego „dziadek” byłoby wieloznaczne z punktu widzenia języka pośredniczącego przekładu, w związku z czym należy wprowadzić dwa odrębne hasła: „dziadek ze strony ojca” oraz „dziadek ze strony matki”. Dzieje się tak za sprawą języka szwedzkiego, w którym nie istnieje jeden rzeczownik na oznaczenie obu rodzajów dziadków, lecz zamiast tego są dwa rzeczowniki: „farfar” – oznaczający dziadka ze strony ojca i „morfar” – oznaczający dziadka ze strony matki.

Jak widać, opracowanie siatki haseł dla języka pośredniczącego przekładu nie jest bynajmniej sprawą prostą, ponieważ należy uwzględnić odpowiedni poziom szczegółowości opisu danego fragmentu rzeczywistości dla wszystkich języków uwzględnionych w systemie. Taki stan rzeczy powoduje, że w przypadku dodania do systemu HAMT nowego języka w wielu przypadkach siatka haseł języka pośredniczącego przekładu będzie musiała zostać odpowiednio zmodyfikowana.

W związku z powyższym język pośredniczący przekładu nie może być w żadnym wypadku postrzegany jako pewien stabilny twór natury statycznej, lecz jako obiekt dynamiczny, podlegający nieustannym przekształceniom, co nader często będzie miało

miejsce wraz z dodawaniem do zasobów leksykalnych systemu HAMT nowych jednostek, a zwłaszcza podczas wprowadzania do niego nowego języka. W tym ostatnim wypadku język pośredniczący przekładu będzie musiał zwykle podlegać dość gruntownej przebudowie.

W przypadku tłumaczenia jednostek leksykalnych języka pośredniczącego pełniących rolę fraza rzeczownikowych na język polski system HAMT również korzysta z odpowiedniego pliku. Potrzebna do przeprowadzenia tego procesu wiedza lingwistyczna zapisana jest w pliku:

## INT\_Polish\_Noun.txt

Centrum Inteligentnych Systemów Informatycznych Akademia Górniczo-Hutnicza im. Stanisława Staszica al. Mickiewicza 30, 30-059 Kraków  
budynek C-2 pokój 426 tel.: 12 617 44 53 www.isi.agh.edu.pl isi@agh.edu.pl

Przykładowy rekord danych zapisanych w powyższym pliku może przyjąć następującą postać:

\*\_transformator\_wysokiego\_napięcia\$transformator wysokiego napięcia \$transformatora wysokiego napięcia\$transformatorowi wysokiego napięcia \$transformator wysokiego napięcia\$transformatorem wysokiego napięcia \$transformatorze wysokiego napięcia&1&1#

W podanym powyżej przykładzie fraza rzeczownikowa należąca do języka pośredniczącego przekładu „\_transformator\_wysokiego\_napięcia”, występująca bezpośrednio po znaku „\*”, tłumaczona jest na język polski. Z tego powodu po kolejnych znakach „\$” występują przekłady rozważanej frazy na język polski, które podlegają odmianie w kolejnych przypadkach języka polskiego (mianowniku, dopełniaczu, celowniku, bierniku, narzędniku i miejscowniku).

Następnie po pierwszym znaku „&” występuje informacja o liczbie gramatycznej polskiej frazy rzeczownikowej „transformator wysokiego napięcia”. W podanym przykładzie na rozważanej pozycji znajduje się jedynek, ponieważ fraza rzeczownikowa „transformator wysokiego napięcia” występuje w liczbie pojedynczej. W rozważanym miejscu może pojawić się jeszcze dwójka, w przypadku gdyby dana fraza rzeczownikowa występowała w liczbie mnogiej.

Analogicznie, po drugim ze znaków „&” również pojawia się jedynek, ponieważ fraza rzeczownikowa „transformator wysokiego napięcia” występuje w rodzaju męskim. Dopuszczalnymi wartościami liczbowymi, które mogą wystąpić w rozważanym miejscu, są jeszcze dwójka i trójka – odpowiednio dla rodzaju żeńskiego i nijakiego. Natomiast gdy dana fraza rzeczownikowa występuje w liczbie mnogiej, wówczas dopuszczalnymi wartościami są jedynie jedynek (dla rodzaju męskoosobowego) i dwójka (dla rodzaju niemęskoosobowego).

Podanie tego rodzaju informacji jest konieczne ze względu na występowanie w języku polskim związku zgody pomiędzy frazami pełniącymi funkcję podmiotu i orzeczenia zdania.

Zapis rekordu danych związanych z jednostką leksykalną języka pośredniczącego przekładu „\_transformator\_wysokiego\_napięcia” zakończony jest znakiem „#”.

Poniżej podano jeszcze przykłady rekordów danych związanych z frazą języka pośredniczącego przekładu „\_transformator\_wysokiego\_napięcia” występująca w liczbie podwójnej i mnogiej:

\*\_transformator\_wysokiego\_napięcia\_D\$transformatory wysokiego napięcia \$transformatorów wysokiego napięcia\$transformatorom wysokiego napięcia \$transformatory wysokiego napięcia \$transformatorami wysokiego napięcia \$transformatorach wysokiego napięcia&2&2#

\*\_transformator\_wysokiego\_napięcia\_P\$transformatory wysokiego napięcia \$transformatorów wysokiego napięcia\$transformatorom wysokiego napięcia \$transformatory wysokiego napięcia \$transformatorami wysokiego napięcia

\$transformatorach wysokiego napięcia&2&2#

Jak wynika z podanych powyżej przykładów, w obu przypadkach po pierwszym ze znaków „&” występuje dwójka, ponieważ zarówno liczba podwójna, jak i liczba mnoga języka pośredniczącego tłumaczone są na formy liczby mnogiej języka polskiego.

Podobnie po drugim ze znaków „&” występuje również dwójka, co związane jest z faktem, że fraza rzeczownikowa „transformatory wysokiego napięcia” występuje w rodzaju niemęskoosobowym.

Z kolei wiedza lingwistyczna potrzebna do tłumaczenia fraz czasownikowych z języka pośredniczącego przekładu (interlingua) na język polski zapisana jest w trzech różnych plikach, związanych odpowiednio z formami charakterystycznymi dla pierwszej, drugiej i trzeciej osoby gramatycznej.

W przypadku fraz czasownikowych przechodnich potrzebna do funkcjonowania systemu HAMT wiedza lingwistyczna zapisana jest w następujących plikach, zawierających formy fleksyjne odpowiednio dla pierwszej, drugiej i trzeciej osoby gramatycznej:

### **Polish\_INT\_VerbTrans\_P1.txt Polish\_INT\_VerbTrans\_P2.txt Polish\_INT\_VerbTrans\_P3.txt**

Przykładowy rekord danych znajdujący się w pliku „Polish\_INT\_VerbTrans\_P1.txt” może mieć następującą postać:

```
*wprowadzać w błąd$wprowadzam w błąd$wprowadzamy w błąd
$wprowadzę w błąd$wprowadzimy w błąd
$wprowadziłem w błąd$wprowadziłam w błąd
$wprowadziliśmy w błąd$wprowadziłyśmy w błąd
$wprowadziłbym w błąd$wprowadziłabym w błąd
$wprowadzilibyśmy w błąd$wprowadziłybyśmy w błąd%_wprowadzać_w_błąd&4#
```

Jak widać, zapis rozważanego rekordu danych rozpoczyna się od znaku „\*”, po którym występuje polskojęzyczna fraza czasownikowa „wprowadzać w błąd”. Następnie po kolejnych znakach „\$” występują kolejne formy fleksyjne rozważanej frazy, charakterystyczne odpowiednio dla:

- czasu teraźniejszego liczby pojedynczej,
- czasu teraźniejszego liczby mnogiej,
- czasu przyszłego liczby pojedynczej,
- czasu przyszłego liczby mnogiej,
- czasu przeszłego liczby pojedynczej (rodzaj męski),
- czasu przeszłego liczby pojedynczej (rodzaj żeński),
- czasu przeszłego liczby mnogiej (rodzaj męskoosobowy),
- czasu przeszłego liczby mnogiej (rodzaj niemęskoosobowy),
- trybu warunkowego liczby pojedynczej (rodzaj męski),
- trybu warunkowego liczby pojedynczej (rodzaj żeński),
- trybu warunkowego liczby mnogiej (rodzaj męskoosobowy),

- trybu warunkowego liczby mnogiej (rodzaj niemęskoosobowy).  
Następnie po znaku „%” występuje tłumaczenie rozważanej frazy czasownikowej

języka polskiego „wprowadzać w błąd” na język pośredniczący przekładu

„\_wprowadzać\_w\_błąd”. Jak już uprzednio wspomniano, każda jednostka leksykalna należąca do języka pośredniczącego przekładu rozpoczyna się od znaku „\_”, a kolejne jej człony są również rozdzielone znakami „\_”.

Dalej po znaku „&” występuje czwórka, ponieważ polskojęzyczna fraza czasownikowa „wprowadzać w błąd” wymaga użycia jej dopełnienia bliższego właśnie w czwartym przypadku języka polskiego, czyli w bierniku. Dopuszczalnymi wartościami liczbowymi, które mogą wystąpić w tym polu są: 1, 2, 3, 4, 5 i 6, odpowiednio dla mianownika, dopełniacza, celownika, biernika, narzędnika i miejscownika.

Dodatkowo warto zwrócić uwagę, że siódmy przypadek języka polskiego, czyli wołacz, nie jest na tym etapie uwzględniany, ponieważ jest on przypadkiem niezależnym i nie wchodzi w związku wyrazowe z innymi elementami zdania.

Analogicznie, przykładowy rekord danych znajdujący się w „Polish\_INT\_VerbTrans\_P2.txt” może przybrać następującą postać:

```
*wprowadzać w błąd$wprowadzasz w błąd$wprowadzacie w błąd $wprowadzisz w błąd
$wprowadzicie w błąd
$wprowadziłeś w błąd$wprowadziłaś w błąd
$wprowadziliście w błąd$wprowadziłyście w błąd
```

```
$wprowadziłbyś w błąd$wprowadziłabyś w błąd
$wprowadzilibyście w błąd$wprowadziłybyście w błąd%_wprowadzać_w_błąd&4#
```

pliku

Znaczenie poszczególnych pól zamieszczonego powyżej rekordu danych jest identyczne jak w omówionym uprzednio przypadku. Jedyna różnica sprowadza się do tego, że odpowiednie formy fleksyjne danej frazy czasownikowej charakterystyczne są dla drugiej osoby gramatycznej.

Podobnie, przykładowy rekord danych znajdujący się w „Polish\_INT\_VerbTrans\_P3.txt” może przybrać następującą postać:

```
*wprowadzać w błąd$wprowadza w błąd$wprowadzają w błąd $wprowadzi w błąd$wprowadzą w
błąd
$wprowadził w błąd$wprowadziła w błąd$wprowadziło w błąd $wprowadzili w błąd$wprowadziły w
błąd
```

```
$wprowadziłby w błąd$wprowadziłaby w błąd$wprowadziłoby w błąd $wprowadziliby w błąd
$wprowadziłyby w błąd%_wprowadzać_w_błąd&4#
```

pliku

W rozważanym przypadku po kolejnych znakach „\$” występują kolejne formy fleksyjne rozważanej frazy czasownikowej, charakterystyczne odpowiednio dla:

- czasu teraźniejszego liczby pojedynczej,
- czasu teraźniejszego liczby mnogiej,
- czasu przyszłego liczby pojedynczej,
- czasu przyszłego liczby mnogiej,
- czasu przeszłego liczby pojedynczej (rodzaj męski),
- czasu przeszłego liczby pojedynczej (rodzaj żeński),
- czasu przeszłego liczby pojedynczej (rodzaj nijaki);
- czasu przeszłego liczby mnogiej (rodzaj męskoosobowy),
- czasu przeszłego liczby mnogiej (rodzaj niemęskoosobowy),
- trybu warunkowego liczby pojedynczej (rodzaj męski),
- trybu warunkowego liczby pojedynczej (rodzaj żeński),
- trybu warunkowego liczby pojedynczej (rodzaj nijaki),
- trybu warunkowego liczby mnogiej (rodzaj męskoosobowy),
- trybu warunkowego liczby mnogiej (rodzaj niemęskoosobowy).

Z kolei w przypadku tłumaczenia z języka pośredniczącego przekładu na język polski potrzebna do funkcjonowania systemu HAMT wiedza lingwistyczna zapisana jest w następujących plikach, zawierających formy fleksyjne odpowiednio dla pierwszej, drugiej i trzeciej osoby gramatycznej:

#### **INT\_Polish\_VerbTrans\_P1.txt INT\_Polish\_VerbTrans\_P2.txt INT\_Polish\_VerbTrans\_P3.txt**

Przykładowy rekord danych znajdujący się w pliku „INT\_Polish\_VerbTrans\_P1.txt” może mieć następującą postać:

\*\_wprowadzać\_w\_błąd\$wprowadzam w błąd\$wprowadzamy w błąd \$wprowadzę w błąd  
\$wprowadzimy w błąd  
\$wprowadziłem w błąd\$wprowadziłam w błąd  
\$wprowadziliśmy w błąd\$wprowadziłyśmy w błąd

\$wprowadziłbym w błąd\$wprowadziłabym w błąd \$wprowadzilibyśmy w błąd\$wprowadziłybyśmy w błąd&4#

W zamieszczonym powyżej rekordzie danych po rozpoczynającym go znaku „\*” występuje fraza czasownikowa należąca do języka pośredniczącego przekładu „\_wprowadzać\_w\_błąd”. Następnie

**Centrum Inteligentnych Systemów Informatycznych** Akademia Górniczo-Hutnicza im. Stanisława Staszica al. Mickiewicza 30, 30-059 Kraków  
budynek C-2 pokój 426 tel.: 12 617 44 53 www.isi.agh.edu.pl [isi@agh.edu.pl](mailto:isi@agh.edu.pl)



po kolejnych znakach „\$” występują tłumaczenia tej frazy na język polski w formach fleksyjnych charakterystycznych dla:

- czasu teraźniejszego liczby pojedynczej,
  - czasu teraźniejszego liczby mnogiej,
  - czasu przyszłego liczby pojedynczej,
  - czasu przyszłego liczby mnogiej,
  - czasu przeszłego liczby pojedynczej (rodzaj męski),
  - czasu przeszłego liczby pojedynczej (rodzaj żeński),
  - czasu przeszłego liczby mnogiej (rodzaj męskoosobowy),
  - czasu przeszłego liczby mnogiej (rodzaj niemęskoosobowy),
  - trybu warunkowego liczby pojedynczej (rodzaj męski),
  - trybu warunkowego liczby pojedynczej (rodzaj żeński),
  - trybu warunkowego liczby mnogiej (rodzaj męskoosobowy),
  - trybu warunkowego liczby mnogiej (rodzaj niemęskoosobowy).
- Z kolei po znaku „&” występuje czwórka, ponieważ rozważana fraza czasownikowa wymaga użycia jej dopełnienia bliższego właśnie w czwartym przypadku języka polskiego, czyli w bierniku.

Podobnie, przykładowy rekord danych znajdujący się w „INT\_Polish\_VerbTrans\_P2.txt” może mieć następującą postać:

\*\_wprowadzać\_w\_błąd\$wprowadzasz w błąd\$wprowadzacie w błąd \$wprowadzisz w błąd  
\$wprowadzicie w błąd  
\$wprowadziłeś w błąd\$wprowadziłaś w błąd  
\$wprowadziliście w błąd\$wprowadziłyście w błąd

\$wprowadziłbyś w błąd\$wprowadziłabyś w błąd \$wprowadzilibyście w błąd\$wprowadziłybyście w błąd&4#

pliku

Analogicznie jak uprzednio po kolejnych znakach „\$” występują tłumaczenia rozważanej frazy czasownikowej na język polski w drugiej osobie gramatycznej w formach fleksyjnych charakterystycznych dla:

- czasu teraźniejszego liczby pojedynczej,
- czasu teraźniejszego liczby mnogiej,
- czasu przyszłego liczby pojedynczej,
- czasu przyszłego liczby mnogiej,
- czasu przeszłego liczby pojedynczej (rodzaj męski),
- czasu przeszłego liczby pojedynczej (rodzaj żeński),
- czasu przeszłego liczby mnogiej (rodzaj męskoosobowy),
- czasu przeszłego liczby mnogiej (rodzaj niemęskoosobowy),
- trybu warunkowego liczby pojedynczej (rodzaj męski),
- trybu warunkowego liczby pojedynczej (rodzaj żeński),
- trybu warunkowego liczby mnogiej (rodzaj męskoosobowy),
- trybu warunkowego liczby mnogiej (rodzaj niemęskoosobowy).

Z kolei przykładowy rekord danych znajdujący się w „INT\_Polish\_VerbTrans\_P3.txt” może mieć następującą postać:

\*\_wprowadzać\_w\_błąd\$wprowadza w błąd\$wprowadzają w błąd \$wprowadzi w błąd\$wprowadzą w błąd  
\$wprowadził w błąd\$wprowadziła w błąd\$wprowadziło w błąd \$wprowadzili w błąd\$wprowadziły w błąd

\$wprowadziłby w błąd\$wprowadziłaby w błąd\$wprowadziłoby w błąd \$wprowadziliby w błąd  
\$wprowadziłyby w błąd&4#

pliku

Podobnie jak uprzednio po kolejnych znakach „\$” występują tłumaczenia rozważanej frazy czasownikowej na język polski w trzeciej osobie gramatycznej w formach fleksyjnych charakterystycznych dla:

- czasu teraźniejszego liczby pojedynczej,
- czasu teraźniejszego liczby mnogiej,
- czasu przyszłego liczby pojedynczej,

**Centrum Inteligentnych Systemów Informatycznych** Akademia Górniczo-Hutnicza im. Stanisława Staszica al. Mickiewicza 30, 30-059 Kraków  
budynek C-2 pokój 426 tel.: 12 617 44 53 www.isi.agh.edu.pl [isi@agh.edu.pl](mailto:isi@agh.edu.pl)

- czasu przyszłego liczby mnogiej,
  - czasu przeszłego liczby pojedynczej (rodzaj męski),
  - czasu przeszłego liczby pojedynczej (rodzaj żeński),
  - czasu przeszłego liczby pojedynczej (rodzaj nijaki),
  - czasu przeszłego liczby mnogiej (rodzaj męskoosobowy),
  - czasu przeszłego liczby mnogiej (rodzaj niemęskoosobowy),
  - trybu warunkowego liczby pojedynczej (rodzaj męski),
  - trybu warunkowego liczby pojedynczej (rodzaj żeński),
  - trybu warunkowego liczby pojedynczej (rodzaj nijaki),
  - trybu warunkowego liczby mnogiej (rodzaj męskoosobowy),
  - trybu warunkowego liczby mnogiej (rodzaj niemęskoosobowy).
- Z kolei w przypadku tłumaczenia z języka polskiego na język pośredniczący przekładu fraz czasownikowych występujących w formie strony biernej potrzebna do przeprowadzenia tego rodzaju procesu wiedza lingwistyczna zapisana jest w następujących plikach:

### Polish\_INT\_VerbPassiv\_P1.txt

### Polish\_INT\_VerbPassive\_P2.txt Polish\_INT\_VerbPassive\_P3.txt

Przykładowy rekord danych zapisany w pliku „Polish\_INT\_VerbPassive\_P1, zawierający formy fleksyjne charakterystyczne dla pierwszej osoby gramatycznej, może mieć następującą postać:

```
*oszukać$jestem oszukiwany$jestem oszukiwana  
$jesteśmy oszukiwani$jesteśmy oszukiwane  
$zostanę oszukany$zostanę oszukana  
$zostaniemy oszukani$zostaniemy oszukane  
$zostałem oszukany$zostałam oszukana  
$zostaliśmy oszukani$zostaliśmy oszukane  
$zostałbym oszukany$zostałabym oszukana  
$zostalibyśmy oszukani$zostałybyśmy oszukane%_oszukać#
```

Podobnie, przykładowy rekord danych „Polish\_INT\_VerbPassive\_P2, zawierający formy fleksyjne charakterystyczne dla drugiej osoby gramatycznej, może przyjąć następującą postać:

```
*oszukać$jesteś oszukiwany$jesteś oszukiwana  
$jesteście oszukiwani$jesteście oszukiwane  
$zostaniesz oszukany$zostaniesz oszukana
```

**Centrum Inteligentnych Systemów Informatycznych** Akademia Górniczo-Hutnicza im. Stanisława Staszica al. Mickiewicza 30, 30-059 Kraków  
budynek C-2 pokój 426 tel.: 12 617 44 53 www.isi.agh.edu.pl [isi@agh.edu.pl](mailto:isi@agh.edu.pl)

\$zostaniecie oszukani\$zostaniecie oszukane  
\$zostałeś oszukany\$zostałaś oszukana  
\$zostaliście oszukani\$zostaliście oszukane  
\$zostałbyś oszukany\$zostałabyś oszukana  
\$zostalibyście oszukani\$zostałybyście oszukane%\_oszukać#

Analogicznie, przykładowy rekord danych „Polish\_INT\_VerbPassive\_P3, zawierający formy fleksyjne charakterystyczne dla trzeciej osoby gramatycznej, może przyjąć następującą postać:

\*oszukać\$jest oszukiwany\$jest oszukiwana\$jest oszukiwane \$są oszukiwani\$są oszukiwane  
\$zostanie oszukany\$zostanie oszukana\$zostanie oszukane \$zostaną oszukani\$zostaną oszukane

\$został oszukany\$została oszukana\$zostało oszukane \$zostali oszukani\$zostały oszukane  
\$zostałby oszukany\$zostałaby oszukana\$zostałoby oszukane \$zostaliby oszukani\$zostałyby oszukane%\_oszukać#

Również w przypadku tłumaczenia z języka pośredniczącego przekładu na język polski fraz czasownikowych występujących w formie strony biernej potrzebna do przeprowadzenia tego rodzaju procesu wiedza lingwistyczna zapisana jest w następujących plikach:

**INT\_Polish\_VerbPassiv\_P1.txt INT\_Polish\_VerbPassive\_P2.txt INT\_Polish\_VerbPassive\_P3.txt**

zapisany w

pliku

zapisany w

pliku

Przykładowy rekord danych zapisany w pliku „INT\_Polish\_VerbPassive\_P1, zawierający formy fleksyjne charakterystyczne dla pierwszej osoby gramatycznej, może mieć następującą postać:

\*\_oszukać\$jestem oszukiwany\$jestem oszukiwana \$jesteśmy oszukiwani\$jesteśmy oszukiwane  
\$zostanę oszukany\$zostanę oszukana \$zostaniemy oszukani\$zostaniemy oszukane \$zostałem oszukany\$zostałam oszukana \$zostaliśmy oszukani\$zostaliśmy oszukane \$zostałbym oszukany  
\$zostałabym oszukana \$zostalibyśmy oszukani\$zostałybyśmy oszukane#

Podobnie, przykładowy rekord danych „INT\_Polish\_VerbPassive\_P2, zawierający formy fleksyjne charakterystyczne dla drugiej osoby gramatycznej, może przyjąć następującą postać:

\*\_oszukać\$jesteś oszukiwany\$jesteś oszukiwana \$jesteście oszukiwani\$jesteście oszukiwane  
\$zostaniesz oszukany\$zostaniesz oszukana \$zostaniecie oszukani\$zostaniecie oszukane \$zostałeś oszukany\$zostałaś oszukana \$zostaliście oszukani\$zostaliście oszukane \$zostałbyś oszukany  
\$zostałabyś oszukana \$zostalibyście oszukani\$zostałybyście oszukane#

Analogicznie, przykładowy rekord danych „INT\_Polish\_VerbPassive\_P3, zawierający formy fleksyjne charakterystyczne dla trzeciej osoby gramatycznej, może przyjąć następującą postać:

\*\_oszucać\$jest oszukiwany\$jest oszukiwana\$jest oszukiwane \$są oszukiwani\$są oszukiwane  
\$zostanie oszukany\$zostanie oszukana\$zostanie oszukane \$zostaną oszukani\$zostaną oszukane

\$został oszukany\$została oszukana\$zostało oszukane \$zostali oszukani\$zostały oszukane  
\$zostałby oszukany\$zostałaby oszukana\$zostałoby oszukane \$zostaliby oszukani\$zostałyby oszukane#

W przypadku tłumaczenia fraz czasownikowych przechodnich wymagających użycia dopełnienia bliższego z języka pośredniczącego przekładu na język polski potrzebna do funkcjonowania systemu HAMT wiedza lingwistyczna zapisana jest w pliku:

### Polish\_INT\_Verblmp.txt

Z kolei przykładowy rekord danych zapisanych w rozważanym pliku „Polish\_INT\_Verblmp.txt” może przybrać następującą postać:

\*przyglądać się\$przyglądaj się\$przyglądajcie się%\_przyglądać\_się&3#

W zamieszczonym powyżej rekordzie danych po znaku „\*” występuje polskojęzyczna fraza czasownikowa „przyglądać się”. Następnie po kolejnych znakach „\$” występują formy trybu rozkazującego drugiej osoby gramatycznej, odpowiednio w liczbie pojedynczej i

zapisany w

pliku

zapisany w

pliku

mnoziej. Z kolei po znaku „%” występuje tłumaczenie rozważanej frazy czasownikowej na język pośredniczący przekładu (interlingua), czyli „\_przyglądać\_się”. Podobnie, po znaku „&” występuje trójka, ponieważ występuje polskojęzyczna fraza czasownikowa „przyglądać się” wymaga użycia dopełnienia bliższego w trzecim przypadku języka polskiego, czyli w celowniku.

Z kolei tłumaczenie fraz czasownikowych należących do języka pośredniczącego przekładu na język polski wymaga od systemu posiłkowania się następującym plikiem, w którym jest zawarta odpowiednia wiedza lingwistyczna:

### INT\_Polish\_Verblmp.txt

Przykładowy rekord danych zapisanych w powyższym pliku może przyjąć następującą postać:

\*\_przyglądać\_się\$przyglądaj się\$przyglądajcie się&3#

W tym wypadku bezpośrednio po znaku „\*” występuje fraza czasownikowa należąca do języka pośredniczącego przekładu „\_przyglądać\_się”. Analogicznie, po kolejnych znakach „\$” występują polskojęzyczne tłumaczenia wymienionej frazy na formę trybu rozkazującego drugiej osoby gramatycznej, odpowiednio w liczbie pojedynczej i mnogiej. Również po znaku „&” występuje trójka, ponieważ występuje polskojęzyczna fraza czasownikowa „przyglądać się” wymaga użycia dopełnienia bliższego w trzecim przypadku języka polskiego, czyli w celowniku.

**Centrum Inteligentnych Systemów Informatycznych** Akademia Górniczo-Hutnicza im. Stanisława Staszica al. Mickiewicza 30, 30-059 Kraków  
budynek C-2 pokój 426 tel.: 12 617 44 53 www.isi.agh.edu.pl [isi@agh.edu.pl](mailto:isi@agh.edu.pl)

W przypadku gdy wybrana przez użytkownika fraza czasownikowa poprzedzona jest jednym z możliwych do wyboru w systemie HAMI czasowników modalnych, wówczas musi ona wystąpić w formie bezokolicznikowej. W związku z powyższym system musi zostać wyposażony jeszcze w następujące pliki:

#### **Polish\_INT\_Verblnf.txt INT\_Polish\_Verblnf.txt**

Przykładowy rekord danych zawarty w pliku „Polish\_INT\_Verblnf.txt” może mieć następująca postać:

\*wprowadzać w błąd\$wprowadzać w błąd%\_wprowadzać\_w\_błąd&4#

Analogicznie, Przykładowy rekord danych zawarty w pliku „INT\_Polish\_Verblnf.txt”

może przybrać następująca postać: \*\_wprowadzać\_w\_błąd\$wprowadzać w błąd&4#

W obu wymienionych powyżej przypadkach po znaku „&” występuje czwórka, ponieważ polskojęzyczna fraza „wprowadzać w błąd” wymaga użycia dopełnienia bliższego w bierniku.

Także tłumaczenia nieodmiennych fraz typu okolicznikowego zapisane są w odpowiednich plikach noszących nazwy:

#### **Polish\_INT\_Phrase.txt INT\_Polish\_Phrase.txt**

W przypadku tłumaczenia frazy okolicznikowej z języka polskiego na język pośredniczący przekładu (interlingua) przykładowy rekord danych zapisany w pliku „Polish\_INT\_Phrase.txt” może przybrać następującą postać:

\*późnym wieczorem%\_późnym\_wieczorem#

Zaraz po znaku „\*” występuje polskojęzyczna fraza okolicznikowa (okolicznik czasu) „późnym wieczorem”, a po znaku „%” jej tłumaczenie na język pośredniczący przekładu „\_późnym\_wieczorem”.

Z kolei w przypadku tłumaczenia z języka pośredniczącego przekładu na język polski przykładowy rekord danych z pliku „INT\_Polish\_Phrase.txt” może mieć następująca postać:

\*\_późnym\_wieczorem\$późnym wieczorem#

Podobnie jak uprzednio każdy rekord danych rozpoczyna się znakiem „\*”, a kończy znakiem „#”. Bezpośrednio po znaku „\*” występuje fraza okolicznikowa języka pośredniczącego przekładu „\_późnym\_wieczorem”, pełniąca funkcję okolicznika czasu i rozpoczynająca się znakiem „\_”. Natomiast po znaku „\$” występuje przekład rozważanej frazy języka pośredniczącego (interlingua) na język polski, czyli fraza „późnym wieczorem”.

## 6. Podsumowanie

W niniejszym opracowaniu omówiono podstawowe idee leżące u zasad funkcjonowania planowanego wielojęzycznego systemu komputerowego przekładu opartego na wykorzystaniu idei języka pośredniczącego przekładu (interlingua).

Jak już wspomniano, rozważany system komputerowego przekładu dokonuje automatycznych tłumaczeń tekstów zapisanych w językach kontrolowanych, czyli takich, które stanowią jedynie pewien podzbiór języków naturalnych, na których są pierwotnie wzorowane. W zamierzeniu autorów opracowywany system komputerowego przekładu ma być systemem otwartym zarówno semantycznie, jak i składniowo.

Otwarcie semantyczne polega na tym, że słowniki wykorzystywane w systemie mogą być przez cały czas uzupełniane nowymi jednostkami leksykalnymi, związanymi tematycznie z dziedziną, dla której opracowywany jest wielojęzyczny system komputerowego przekładu.

Z kolei otwarcie składniowe systemu polega na tym, że gramatyki języków kontrolowanych mogą być uzupełniane przez nowe reguły pozwalające na tworzenie poprawnych składniowo zdań.

Na zakończenie należy nadmienić, że znajdujący się obecnie fazie analiz wykonalności projekt informatyczny jest pomyślany przez autorów jako zakrojone na szeroką skalę przedsięwzięcie, w realizacji którego będzie docelowo uczestniczyć większa liczba osób, reprezentujących zarówno kierunki informatyczne, jak i filologiczne oraz lingwistyczne.

Zwłaszcza ze strony informatycznej konieczna będzie współpraca osób specjalizujących się w programowaniu systemów sieciowych, w tym także w pisaniu aplikacji dla urządzeń mobilnych. Wiele pracy będzie także do wykonania dla specjalistów od inżynierii systemów, inżynierii oprogramowania, baz danych i graficznych interfejsów użytkownika. Natomiast przedstawiciele nauk lingwistycznych i kierunków filologicznych będą niezbędni podczas uzupełniania leksykonów systemu i lingwistycznych baz danych o nowe jednostki leksykalne wraz z ich przekładami na język pośredniczący i poszczególne języki docelowe przekładu.

Reasumując, wnioskowany system komputerowego przekładu wspomagany przez człowieka HAMT oparty na języku pośredniczącym (interlingua) jest w pełni oryginalną i unikatową koncepcją, ponieważ zgodnie z najlepszą wiedzą autorów na świecie nie istnieją do tej pory tego rodzaju systemy tłumaczące.

Ponadto podstawowym atutem wnioskowanego systemu jest jego wielojęzyczność, która osiągnięta jest dzięki zastosowaniu koncepcji języka pośredniczącego przekładu. Tego rodzaju rozwiązanie stwarza możliwość względnie łatwej rozbudowy systemu poprzez dodawanie do niego kolejnych języków. Ponieważ system może tłumaczyć w obu kierunkach pomiędzy dowolnie wybraną parą języków (spośród języków w nim występujących), dzięki temu obszar jego potencjalnych zastosowań nie jest ograniczony jedynie do terytorium naszego kraju, ale obejmuje także i inne kraje Europy i świata.

Jako potencjalne obszary zastosowań wnioskowanego systemu można wymienić przede wszystkim wbudowane w urządzenie mobilne narzędzie przeznaczone do umożliwienia wzajemnej komunikacji osób nie znających żadnego wspólnego im języka.

Ponadto po odpowiedniej rozbudowie lingwistycznych baz danych o wymagane jednostki leksykalne wnioskowany system może być wykorzystywany w globalnych korporacjach, w przypadku których zachodzi konieczność sporządzania różnego rodzaju dokumentacji (instrukcje użytkownika, książki serwisowe, materiały reklamowe itp.) w wielu językach świata [21- 26]. W takim wypadku wnioskowany system mógłby posłużyć do jednoczesnego wytwarzania tego rodzaju dokumentacji w wielu wybranych przez użytkownika językach.

Innym obszarem zastosowań wnioskowanego systemu jest nauka języków obcych. W tym wypadku za pomocą systemu HAMT użytkownik może uczyć się, jak budować poprawne składniowo zdania w języku, którego się aktualnie uczy. Za pomocą systemu HAMT może on także weryfikować nabytą dotychczas przez siebie wiedzę.

### Bibliografia:

1. Hutchins W. J., Machine translation – past, present, future, London, Ellis Horwood Series in Computers and their Applications, 1986
2. Arnold D., Balkan L., Meijer S., Humphreys R. L., Sadler L., Machine translation: an introductory guide, London, NCC Blackwell, 1994
3. Allen J. F., Natural language understanding, New York, The Benjamin/Cummings Publishing Company, 1995
4. Whitelock P., Kilby K., Linguistic and computational techniques in machine translation system design, London, UCL Press, 1995
5. Gajer M., The implementation of the example-based machine translation technique for Dutch-to-Polish automatic translation system, Bulletin of the Polish Academy of Sciences – Technical Sciences, vol. 51, no. 1, (2003), 31-57
6. Gajer M., The implementation of the example-based machine translation technique for Greek-to-Polish automatic translation system, Foundation of Computing and Decision Sciences, vol. 28, no. 2, (2003), 83-93
7. Gajer M., The implementation of the example-based machine translation technique for Norwegian-to-Polish automatic translation system, Control and Cybernetics, vol. 34, no. 2, (2004), 357-386
8. Gajer M., The pattern-based French-to-Polish machine translation system, Machine Translation Review, no. 13, (2002), 7-41
9. Gajer M., Towards a fully-automatic high-quality machine translation system for unrestricted text, Machine Translation Review, no. 12, (2001), 24-35
10. Gajer M., Specialized fully automatic machine translation system delivering high quality of translated texts, Task Quarterly, vol. 13, no. 4, (2009), 347-354



11. Gajer M., Analiza możliwości i ograniczeń systemów translacji automatycznej wspomaganiej przez człowieka na przykładzie systemu tłumaczącego z języka szwedzkiego na polski, Kwartalnik Elektroniki i Telekomunikacji, tom 49, zeszyt 1, (2003), 47-105
12. Gajer M., System translacji automatycznej oparty na uogólnionych przykładach translacyjnych, Kwartalnik Elektroniki i Telekomunikacji, tom 49, zeszyt 3, (2003), 405-427
13. Gajer M., Zastosowanie algorytmów ewolucyjnych w lingwistyce komputerowej celem oceny stopnia pokrewieństwa języków naturalnych, Kwartalnik Elektroniki i Telekomunikacji, tom 51, zeszyt 2, (2005), 277-290
14. Gajer M., The implementation of the example-based machine translation technique for Persian-to-Polish automatic translation system, Folia Societatis Scientiarum Lublinensis, vol. 11, (2002), 117-134
15. Gajer M., The implementation of the example-based machine translation technique for French-to-Polish automatic translation system, Task Quarterly, vol. 6, no. 3, (2002), 523-544
16. Gajer M., The implementation of the example-based machine translation technique for German-to-Polish automatic translation system, Informatica, vol. 13, no. 4, (2002), 417-440
17. Gajer M., The proposition of using the example-based machine translation technique for Italian-to-Polish automatic translation system, Studia Informatica, vol. 23, no. 4, (2002), 239-254
18. Gajer M., Zastosowanie metody Example-based machine translation w specjalizowanych systemach komputerowego przekładu, Przegląd Elektrotechniczny, rok LXXXVII, no. 2, (2011), 173-178
19. Szczepaniak L., Królikowski Z., Kontrolowane języki naturalne – przegląd rozwiązań i zastosowań, Pro Dialog, nr 11/2000, 47- 68
20. Handzel Z., Gajer M., Dybiec-Gajer J., Zagadnienia związane z przekładem zaimków osobowych w systemie tłumaczenia komputerowego typu Human-Aided Machine Translation, Elektronika: Konstrukcje – Technologie – Zastosowania, nr 4/2014, 70- 72
21. Canals R., Esteve A., Garrido A.: InterNOSTRUM: a Spanish-Catalan machine translation system, Machine Translation Review, no. 11, 2000, 21-25
22. Fukutomi O.: Report on commercial machine translation in a manufacturing industry domain, Machine Translation Review, no. 10, 1999, 16-25
23. Guidère M.: Toward corpus-based machine translation for standard Arabic, Translation Journal, vol. 6, no. 1, 2002
24. Somers H.: EBMT seen as case-based reasoning, MT Summit VIII, Santiago de Compostela, 2001



25. Sumita E., Iida H.: Experiments and prospects of example-based machine translation, Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics, 1991, 185-192
26. Turcato D., Popowich F.: What is example-based machine translation?, MT Summit VIII, Santiago de Compostela, 2001

**Centrum Inteligentnych Systemów Informatycznych** Akademia Górniczo-Hutnicza im. Stanisława Staszica al. Mickiewicza 30, 30-059 Kraków  
budynek C-2 pokój 426 tel.: 12 617 44 53 [www.isi.agh.edu.pl](http://www.isi.agh.edu.pl) [isi@agh.edu.pl](mailto:isi@agh.edu.pl)

