Klasyfikacja dziedzinowa wg
JEL: L10, C3, C6

Zbigniew Handzel[1]
Mirosław Gajer[2]

Wielojęzyczne generatory gramatyczne występujące w roli narzędzi informatycznych typu Machine-Aided Human Translation

Multilingual generators of syntactic structures playing the role of computer tools for Machine-Aided Human Translation

Streszczenie: W artykule autorzy przedstawili propozycję budowy wielojęzycznych generatorów struktur syntaktycznych, które mogą być postrzegane jako narzędzia wspomagające proces przekładu. Opracowany przez autorów system ma za zadanie umożliwić nie tylko generowanie odpowiednich form fleksyjnych czasowników, ale ma także stanowić swego rodzaju repozytorium związków frazeologicznych, w które mogą wchodzić poszczególne czasowniki podlegające odmianie przez czasy, tryby, osoby, liczby i rodzaje gramatyczne. Niezwykle istotną cechą opracowywanego systemu jest jego wielojęzyczność, która została osiągnięta dzięki wprowadzeniu koncepcji języka pośredniczącego przekładu.

Summary: In the paper the authors propose to build multilingual generators of syntactic structures. Such generators can be regard as computer tools that can be used to facilitate the process of translation. The aim of the system proposed by the authors is not only to generate the proper flexion forms of the verbs, but also it can be respected as a repository of phrases with their flexion forms using tenses, modes, persons, numbers, and grammatical

[1] Dr inż., Wyższa Szkoła Ekonomii i Informatyki w Krakowie, Zakład Informatyki i Metod ilościowych.

[2] Dr inż., AGH Akademia Górniczo-Hutnicza im. S. Staszica w Krakowie Wydział Elektrotechniki, Automatyki, Informatyki i Inżynierii Biomedycznej Katedra Informatyki Stosowanej.

genders. The main merit of the system is its multilingualism that was achieved thanks to the concept of intermediate language.

Słowa kluczowe: przekład komputerowy, przekład wspomagany, generatory struktur gramatycznych

1. Wprowadzenie

W procesie dydaktyki języków obcych bez wątpienia duże trudności osobom uczącym się sprawia czynne opanowanie fleksji czasownikowej. W przypadku bardzo wielu współczesnych języków indoeuropejskich fleksja czasownikowa jest w dosyć wysokim stopniu rozbudowana, ponieważ od podstawowej formy bezokolicznika można utworzyć nawet kilkadziesiąt form pochodnych, związanych z odmianą danego czasownika przez osoby, liczby i rodzaje gramatyczne. Dodatkowo w językach indoeuropejskich czasowniki odmieniają się jeszcze przez rozbudowane systemy czasów i trybów gramatycznych [1].

Uwaga powyższa dotyczy także wielu obecnie już martwych języków indoeuropejskich, które odegrały niegdyś dużą rolę na przestrzeni historii, takich jak na przykład język łaciński, starogrecki, staro-cerkiewno-słowiański, czy też sanskryt. W przypadku wymienionych języków fleksja czasownikowa była nawet o wiele bardziej rozbudowana niż ma to miejsce w przypadku języków nam współczesnych. Jednym z powodów takiego stanu rzeczy było powszechne występowanie w dawnych językach indoeuropejskich liczby podwójnej (tzw. dualis), co pociągało za sobą konieczność tworzenia odrębnych form fleksyjnych w sytuacji, gdy była mowa o dokładnie dwóch osobach lub przedmiotach [2].

Z bogactwem fleksji czasownikowej mamy do czynienia zresztą nie tylko w przypadku języków należących do wielkiej rodziny języków indoeuropejskich. Bowiem analogiczne prawidłowości są obserwowane praktycznie w przypadku wszystkich języków zaliczanych do grupy tzw. języków nostratycznych, co do których istnieją liczne hipotezy dotyczące ich bardzo odległego pokrewieństwa. W tym miejscu mowa jest przede wszystkim o językach należących do rodzin afroazjatyckiej, uralskiej, ałtajskiej i drawidyjskiej, a także o językach takich jak koreański, japoński, riukiuański oraz języki południowego Kaukazu [3]. W przypadku wszystkich wymienionych języków mamy również do czynienia z wysoce złożonymi systemami odmiany czasowników, których opanowanie podczas nauki danego języka jako języka obcego bez wątpienia nastręcza wielu trudności i wymaga poświęcania na to dłuższego czasu i włożenia w cały proces sporego wysiłku.

Jednak nawet osoby, które już dość dobrze opanowały w sposób czynny dany język obcy i na co dzień zajmują się na przykład działalnością translatorską, również niekiedy mogą mieć problemy z poprawnym utworzeniem danej formy fleksyjnej czasownika, zwłaszcza w sytuacji, gdy jest to forma tworzona w sposób nieregularny, co w przypadku wielu języków jest nader częstym zjawiskiem.

Swoistym remedium na rozważane powyżej bolączki może być opracowanie odpowiednich narzędzi informatycznych, których zadaniem jest wspieranie językoznawców, filologów, przekładoznawców, tłumaczy i osób uczących się języków obcych podczas poszukiwania właściwych w danym kontekście form fleksyjnych czasowników. Tego rodzaju programy komputerowe tworzą szerszą klasę narzędzi informatycznych typu MAHT (ang. Machine-Aided Human Translation) [4–29].

Głównym zadaniem systemów typu MAHT jest wspieranie tłumacza podczas jego działalności translatorskiej. Do tego typu narzędzi informatycznych zaliczane są przede wszystkim różnorodne programy komputerowe pełniące rolę na przykład elektronicznych słowników, glosariuszy i leksykonów. Narzędzia typu MAHT to także tzw. pamięci translacyjne, czyli programy komputerowe, za pośrednictwem których pracujący nad przekładem pewnego tekstu tłumacz może wykorzystać swój wcześniejszy dorobek translatorski, co dodatkowo sprzyja ujednoliceniu terminologii i frazeologii tworzonych przez niego tekstów.

Do szeroko rozumianych narzędzi typu MAHT można zaliczyć także tzw. generatory gramatyczne, czyli programy komputerowe, które po podaniu na ich wejście formy podstawowej danej części mowy (na przykład czasownika, rzeczownika, przymiotnika bądź zaimka) potrafią utworzyć od niej dowolną inną formę fleksyjną.

Proponowany przez autorów generator form fleksyjnych czasowników jest generatorem wielojęzycznym, który pozwala na utworzenie wybranej przez użytkownika formy fleksyjnej w dowolnym dostępnym za jego pośrednictwem języku, a następnie dokonuje jej przekładu na inny wybrany przez użytkownika język. Tego rodzaju narzędzie informatyczne typu MAHT może okazać się nieocenioną pomocą dla tłumaczy pracujących nad przekładem tekstów na języki obce, ponieważ oprócz utworzenia poprawnej formy fleksyjnej danego czasownika pozwala stosować w sposób prawidłowy związki frazeologiczne z tym czasownikiem związane.

Zgodnie z najlepszą wiedzą autorów proponowany przez nich wielojęzyczny generator gramatyczny form fleksyjnych czasowników jest rozwiązaniem

nowatorskim i unikatowym, które dotychczas nie ma swojego odpowiednika w dziedzinie narzędzi komputerowych typu MAHT. Co prawda, istnieją powszechnie dostępne, chociażby za pośrednictwem Internetu, liczne generatory form fleksyjnych czasowników, ale zawsze związane są one z pewnym konkretnym językiem i przez to nie dają możliwości tłumaczenia utworzonych za ich pomocą form fleksyjnych czasowników na inny wybrany przez użytkownika język [4–29].

W przypadku proponowanego przez autorów rozwiązania sytuacja przedstawia się zgoła przeciwnie, ponieważ projektowany generator form fleksyjnych czasowników ma być programem obsługującym wiele języków i zapewniającym automatyczny przekład danych form fleksyjnych czasownika również na inne języki.

2. Wielojęzyczne generowanie form fleksyjnych czasowników

Ze wspomnianą już we wstępie koncepcją wielojęzyczności generatora form fleksyjnych czasowników wiąże się jeden zasadniczy problem. Mianowicie, praktycznie we wszystkich językach wyrwany z kontekstu czasownik jest wyrazem w bardzo wysokim stopniu wieloznacznym. Dopiero umieszczenie czasownika w szerszym kontekście w sposób jednoznaczny doprecyzowuje jego znaczenie.

W celu ilustracji prowadzonych rozważań wystarczy chociażby przyjrzeć się polskiemu czasownikowi „kopać", który w zależności od kontekstu występowania może posiadać wiele różnorodnych znaczeń. Z kolei ściśle sprecyzowane znaczenia posiadają dopiero frazy czasownikowe, w których czasownik „kopać" występuje w roli orzeczenia zdania. Jako przykład tego rodzaju fraz czasownikowych można wskazać przykładowo frazy: „kopać piłkę", „kopać ziemniaki", „kopać rów" czy też „kopać leżącego" bądź „kopać pod kimś dołki". W przypadku każdej z wymienionych fraz czasownik kopać ma całkowicie odmienne znaczenie.

Wniosek wypływający z zamieszczonych powyżej rozważań jest taki, że chcąc zbudować wielojęzyczny generator form fleksyjnych czasowników nie można ograniczyć się jedynie do odmiany izolowanych czasowników, lecz należy generować odpowiednie formy fleksyjne dla dłuższych fraz czasownikowych, zawierających wybrany przez użytkownika czasownik przechodni wraz z frazą rzeczownikową, która stanowi jego dopełnianie bliższe.

W związku z powyższym w opracowywanym przez autorów systemie wielojęzycznego generatora form fleksyjnych czasowników odmianie przez osoby,

liczby, rodzaje, czasy i tryby gramatyczne podlegają frazy czasownikowe VP (ang. verb phrase), których budowa może zostać opisana w kategoriach gramatyki transformacyjno-generatywnej w sposób następujący:

$$VP \rightarrow V + NP \quad (1)$$

W regule gramatyki transformacyjno-generatywnej (1) symbolem V (ag. verb) oznaczono czasownik przechodni (łac. verbum transitivum), a symbolem NP (ang. noun phrase) oznaczono frazę rzeczownikową stanowiącą dopełnienie bliższe danego czasownika przechodniego.

Rozważane frazy czasownikowe VP w projektowanym przez autorów systemie wielojęzycznego generatora form fleksyjnych czasowników są następnie tłumaczone na wybrany przez użytkownika język obcy z zachowaniem w miarę możliwości tej samej osoby, liczby i rodzaju gramatycznego. W języku docelowym przekładu wybierany jest także czas i tryb gramatyczny najbardziej zbliżony do tego, który wskazany został przez użytkownika w języku źródłowym przekładu.

Jako przykład frazy czasownikowej, która może podlegać odmianie a następnie procesowi przekładu w projektowanym systemie wielojęzycznego generatora form fleksyjnych czasowników, można podać frazę „jeść jabłko". Jej budowę opisują następujące reguły gramatyki generatywno-transformacyjnej:

$VP \rightarrow V + NP$
$V \rightarrow$ jeść
$NP \rightarrow NOUN$
$NOUN \rightarrow$ jabłko

Innym rodzajem tego typu frazy może być na przykład fraza czasownikowa „generować energię elektryczną". W tym wypadku do przeanalizowania budowy rozważanej frazy czasownikowej należy zastosować następujące reguły gramatyki generatywno-transformacyjnej:

$VP \rightarrow V + NP$
$V \rightarrow$ generować
$NP \rightarrow NOUN + ADJ$
$NOUN \rightarrow$ energię
$ADJ \rightarrow$ elektryczną

Jeszcze inną budowę wykazuje fraza czasownikowa „mieszkać w mieście", w przypadku której zbiór reguł gramatyki generatywno-transformacyjnej przedstawia się następująco:

VP → V + NP
V → mieszkać
NP → PREP + NOUN
PREP → w
NOUN → mieście

Z kolei nieco bardziej złożoną budowę wykazuje fraza „pracować na pełnym etacie", w przypadku której zbiór reguł gramatyki generatywno-transformacyjnej opisujących jej składnię przedstawia się następująco:

VP → V + NP
V → pracować
NP → PREP + NP
PREP → na
NP v ADJ + NOUN
ADJ v pełnym
NOUN v etacie

Jak już wspomniano, projektowany generator form fleksyjnych fraz czasownikowych w swym założeniu ma być generatorem wielojęzycznym, umożliwiającym dokonywanie automatycznego przekładu z wybranego przez użytkownika języka na dowolny inny język dostępny w systemie.

W celu wprowadzenia w systemie wielojęzyczności autorzy zdecydowali się na wprowadzenie języka pośredniczącego przekładu. Aby w projektowanym wielojęzycznym systemie generatora form fleksyjnych fraz czasownikowych nie stawiać w pozycji uprzywilejowanej żadnego z występujących w nim języków naturalnych, postanowiono w roli języka pośredniczącego przekładu zastosować język sztuczny o charakterze aposteriorycznym, który przewidziany jest do pełnienia roli komunikacji międzynarodowej.

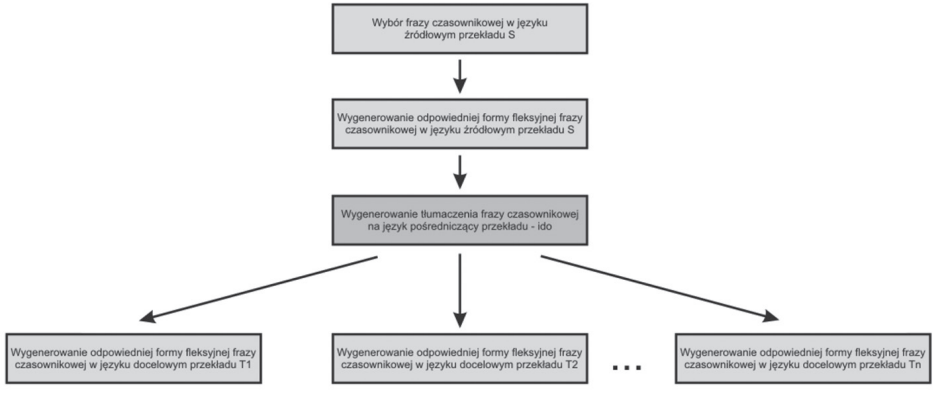
Spośród wielu znanych z literatury propozycji w zakresie języków sztucznych przeznaczonych do komunikacji międzynarodowej autorzy zdecydowali się na wybór języka ido, który stanowi ulepszoną i zreformowaną wersję wcześniejszego języka esperanto. Projekt języka ido został ogłoszony w 1908 roku przez francuskiego lingwistę, Luisa Couturata, który postawił sobie za

cel wyeliminowanie pewnych istotnych wad języka esperanto, a także uproszczenie formy jego graficznego zapisu (eliminacja znaków diakrytycznych) oraz ułatwienia wymowy poszczególnych wyrazów, tak aby język ten mógł dobrze pełnić funkcję języka przeznaczonego do komunikacji międzynarodowej.

W związku z powyższym projektowany system wielojęzycznego generatora form fleksyjnych fraz czasownikowych działa trójetapowo. W etapie pierwszym wybrana przez użytkownika fraza czasownikowa jest odmieniana przez wybrana osobę, liczbę i rodzaj gramatyczny a także występuje w odpowiednim czasie i trybie gramatycznym. W kroku kolejnym rozpatrywana fraza podlega automatycznemu przekładowi na język pośredniczący, którym jest w tym wypadku sztuczny język ido. W etapie końcowym generowana jest fraza czasownikowa w wybranym przez użytkownika języku docelowym przekładu, przy czym fraza ta występuje w odpowiedniej osobie, liczbie, rodzaju, czasie i trybie gramatycznym.

Idea funkcjonowania wielojęzycznego generatora form fleksyjnych fraz czasownikowych została zilustrowana na rys 1.

Rys. 1. Ilustracja sposobu funkcjonowania wielojęzycznego generatora form fleksyjnych fraz czasownikowych



Jak już zostało na wstępie wspomniane, użytkownik systemu wybiera najpierw frazę czasownikową w pewnym języku S, następnie system dokonuje odmiany tej frazy przez wybrane przez użytkownika czas, tryb, osobę, liczbę i rodzaj gramatyczny, po czym dokonuje jej przykładu na sztuczny język ido, pełniący w systemie role języka pośredniczącego przekładu. W etapie kolejnym na podstawie tłumaczenia rozpatrywanej frazy czasownikowej na język

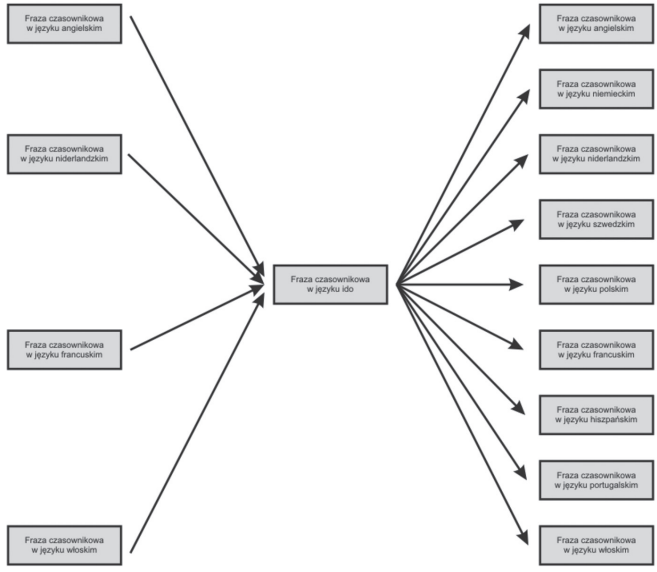
ido dokonywany jest jej przekład na wybrane przez użytkownika języki docelowe T1, T2, ... , Tn.

Jak uprzednio napisano, w roli języka pośredniczącego przekładu postanowiono zastosować sztuczny język ido, dzięki czemu żaden z języków naturalnych nie jest w jakikolwiek sposób faworyzowany w opracowywanym wielojęzycznym systemie generatora form fleksyjnych fraz czasownikowych. Jednak język pośredniczący przez cały czas działania systemu pozostaje jak gdyby w ukryciu i tłumaczenia fraz czasownikowych na język ido nie są bezpośrednio dostępne dla użytkownika systemu. W związku z powyższym język ido odgrywa w projektowanym systemie rolę czysto techniczną. Oczywiście, zawsze istnieje możliwość dobudowania odpowiedniego modułu, dzięki któremu język ido stałby się również jednym z dostępnych do wyboru języków docelowych przekładu. Jednak ze względu na całkowitą regularność tego języka, brak w nim jakichkolwiek wyjątków i uproszczony system czasów i trybów gramatycznych umieszczanie go w roli jednego z języków docelowych przekładu nie wydaje się zbytnio interesujące, aczkolwiek autorzy systemu nie wykluczają w przyszłości takiej możliwości. Zresztą uwaga powyższa dotyczy także i innych języków sztucznych typu naturalistycznego (tzw. aposteriorycznych), pełniących rolę języków komunikacji międzynarodowej, takich jak na przykład esperanto, interlingua (latina sine flexione), occidental, novial, glossa i wiele innych.

Projektując wielojęzyczny generator form fleksyjnych fraz czasownikowych trzeba mieć na uwadze fakt, że każda zapisana w lingwistycznej bazie danych fraza czasownikowa w języku ido musi mieć swoje odpowiedniki we wszystkich pozostałych językach uwzględnionych w systemie. W szczególności nie można w żaden sposób dopuścić do powstania sytuacji, że wybrana przez użytkownika fraza czasownikowa w języku A zostanie przetłumaczona na język ido, a następnie nie będzie możliwości jej przetłumaczenia na jeden z wybranych przez użytkownika języków docelowych, ponieważ w lingwistycznej bazie danych brakowało będzie stosownych rekordów.

Przy okazji prowadzonych rozważań warto jeszcze zwrócić uwagę na fakt, że odwrotna zależność nie jest bynajmniej wymagana do poprawnego funkcjonowania systemu wielojęzycznego generatora form fleksyjnych czasowników. Powyższe stwierdzenie oznacza, że możliwość przejścia z danego języka naturalnego na język ido nie musi występować w przypadku każdej frazy czasownikowej języka ido zapisanej w lingwistycznej bazie danych.

Rys. 2. Ilustracja sposobu dochodzenia do frazy czasownikowej w języku docelowym wychodząc od różnych języków źródłowych



Powyższa prawidłowość została zilustrowana za pomocą schematu przedstawionego na rys. 2. Dla przykładu założono, że w projektowanym systemie wielojęzycznego generatora form fleksyjnych fraz czasownikowych występuje dziewięć języków naturalnych, takich jak angielski, niemiecki, niderlandzki, szwedzki, polski, francuski, hiszpański, portugalski i włoski. Po dokonaniu wyboru w nich dowolnej frazy czasownikowej następuje przejście do odpowiadającej jej frazy czasownikowej w sztucznym języku ido. Następnie na podstawie danej frazy czasownikowej zapisanej w języku ido może nastąpić wygenerowanie odpowiedniej frazy czasownikowej w dowolnym z rozważanych języków naturalnych. Jednak, jak wynika z postaci rys. 2, dojście do rozpatrywanej frazy czasownikowej zapisanej w języku ido może mieć miejsce jedynie za pośrednictwem języków, takich jak angielski, niderlandzki, francuski i włoski. W przypadku pozostałych języków uwzględnionych dotychczas w systemie dojście do rozpatrywanej frazy czasownikowej sztucznego języka ido nie jest możliwe. Jednak taki stan rzeczy, po pierwsze, nie powoduje żadnych zakłóceń odnośnie poprawnego funkcjonowania systemu wielojęzycznego generatora form fleksyjnych, ponieważ jeżeli użytkownik wybierze dowolna frazę w którymś z dostępnych języków, to zawsze fraza ta zostanie przetłumaczona na język ido, a z następnie z tego języka na każdy z języków uwzględnionych w systemie. Po drugie, lingwistyczne bazy danych dla każdego języka mogą być systematycznie rozbudowywane, w związku z czym, jeśli możliwość dojścia

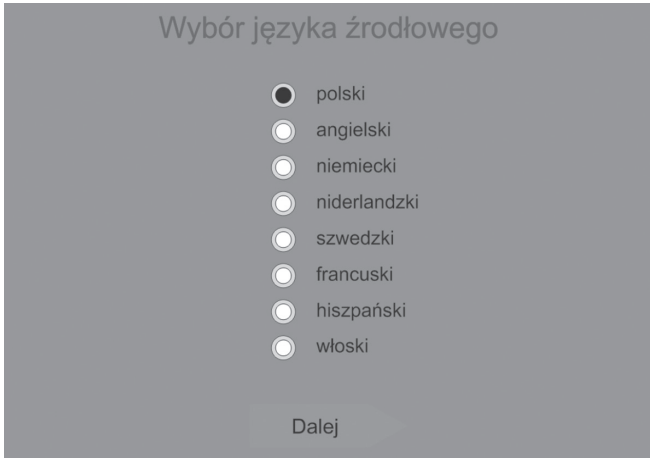
do jakiejś frazy czasownikowej języka ido dotychczas nie występowała, to nie oznacza to wcale, że taka opcja nie pojawi się kiedyś w przyszłości po dodaniu do lingwistycznej bazy danych stosowanych rekordów.

3. Projekt graficznego interfejsu użytkownika

W niniejszym punkcie przedstawiono koncepcję budowy i zasad funkcjonowania graficznego interfejsu użytkownika, za pośrednictwem którego można dokonywać wyboru fraz czasownikowych w dostępnych w systemie językach, a następnie zadecydować, w jakim czasie, trybie, osobie, liczbie i rodzaju gramatycznym wybrana fraza czasownikowa powinna wystąpić. W dalszej kolejności za pośrednictwem graficznego interfejsu użytkownika można dokonać wyboru języka bądź jednocześnie kilku języków, na które dana fraza czasownikowa ma zostać przetłumaczona. Warto zauważyć, że w szczególnym przypadku istnieje możliwość przetłumaczenia frazy czasownikowej zapisanej w pewnym języku L również na ten sam język L. Postępując w ten sposób użytkownik otrzyma zwykle dokładnie tę samą frazę czasownikową, którą wybrał uprzednio. Jednak w niektórych wypadkach może otrzymać również i inną frazę czasownikową, stanowiącą parafrazę frazy czasownikowej uprzednio wybranej. Sytuacja taka może mieć miejsce w wypadku, gdy do danej frazy czasownikowej zapisanej w sztucznym języku ido można będzie dojść wychodząc od dwóch różnych fraz czasownikowych zapisanych w języku źródłowym. Z kolei przejście od frazy czasownikowej zapisanej w języku ido do frazy czasownikowej zapisanej w którymkolwiek z dostępnych w systemie języków musi być zawsze jednoznaczne.

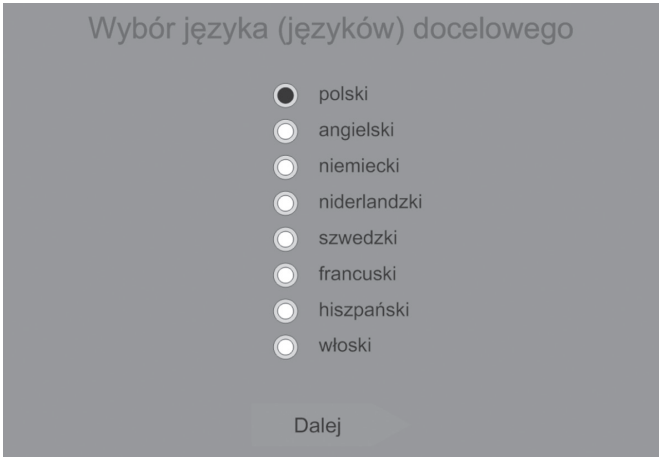
W pierwszym etapie interakcji z systemem wielojęzycznego generatora form fleksyjnych czasowników jego użytkownik musi dokonać wyboru języka, w którym będą podlegały odmianie przez czasy, tryby, osoby, liczby i rodzaje gramatyczne wybierane przez niego frazy czasownikowe, co zostało pokazane na rys. 3.

Rys. 3. Projektowany wygląd graficznego interfejsu użytkownika podczas etapu wyboru języka źródłowego



Postępując w sposób analogiczny użytkownik systemu musi dokonać wyboru języka docelowego, na który tłumaczone będą wybierane przez niego frazy czasownikowe, co przedstawiono na rys. 4.

Rys. 4. Projektowany wygląd graficznego interfejsu użytkownika podczas etapu wyboru języka docelowego

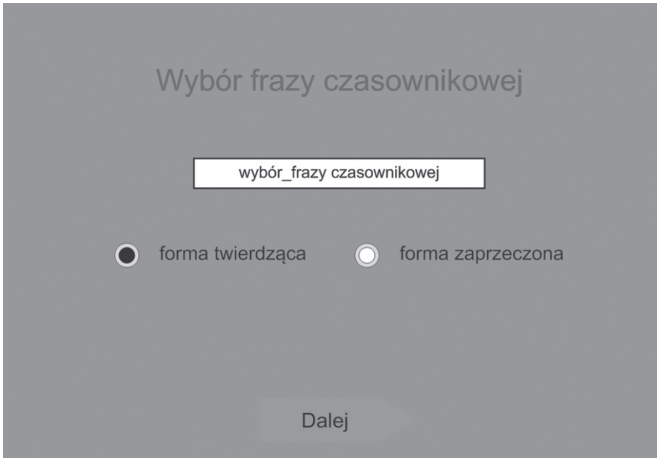


Jak wynika z postaci rys. 3 i 4 w obu przypadkach założono, że wyborem domyślnym jest język polski, ale użytkownik może to oczywiście zmienić

zgodnie ze swoimi preferencjami. Ponadto w przypadku wyboru języka docelowego użytkownik może zaznaczyć jednocześnie kilka wybranych przez niego języków, a w szczególnym przypadku mogą to być nawet i wszystkie dostępne w systemie języki. W takim wypadku wybierane przez użytkownika frazy czasownikowe będą tłumaczone na wszystkie języki zaznaczone przez niego na etapie przedstawionym na rys. 4.

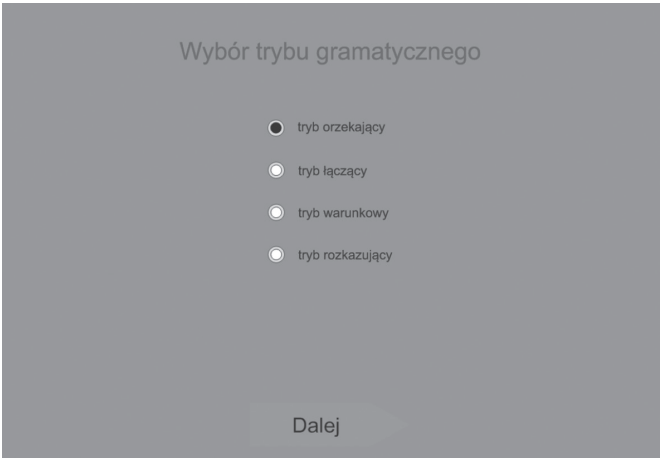
W następnym etapie użytkownik systemu dokonuje wyboru frazy czasownikowej, która podlegać będzie odmianie przez tryby, czasy, osoby, liczby i rodzaje gramatyczne, a następnie będzie tłumaczona na wybrane języki docelowe, co przedstawiono na rys. 5.

Rys. 5. Projektowany wygląd graficznego interfejsu użytkownika podczas etapu wyboru frazy czasownikowej



W kroku kolejnym użytkownik musi dokonać wyboru trybu, w którym wystąpić ma wybrana uprzednio przez niego fraza czasownikowa.

Rys. 6. Projektowany wygląd graficznego interfejsu użytkownika podczas etapu wyboru trybu gramatycznego

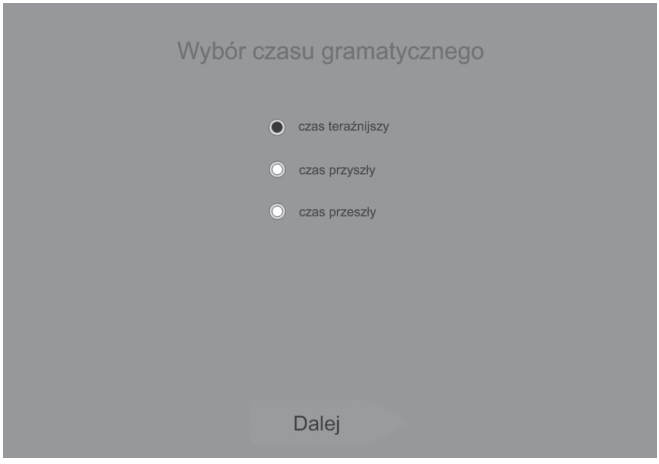


Jak wynika z rys. 6, na rozważanym etapie użytkownik ma do wyboru następujące tryby gramatyczne: orzekający, łączący, warunkowy i rozkazujący. W tym miejscu należy zaznaczyć, że formy fleksyjne charakterystyczne dla trybu łączącego nie występują bynajmniej we wszystkich językach należących do wielkiej rodziny języków indoeuropejskich. Tego rodzaju kategoria gramatyczna jest na przykład spotykana we współczesnych językach należących do grupy języków romańskich. W języku polskim odrębne formy fleksyjne czasowników charakterystyczne dla trybu łączącego nie występują.

W projektowanym systemie wielojęzycznego generatora form fleksyjnych czasowników przyjęto zasadę, w myśl której, jeśli w danym języku tryb łączący nie występuje, wówczas w przypadku dokonania przez użytkownika jego wyboru, system zachowuje się tak, jak gdyby wybrano tryb orzekający.

W kolejnym etapie użytkownik systemu musi dokonać wyboru czasu gramatycznego, w którym ma wystąpić dana fraza czasownikowa, co pokazano na rys. 7.

Rys. 7. Projektowany wygląd graficznego interfejsu użytkownika podczas etapu wyboru czasu gramatycznego



W przypadku wskazania języka polskiego jako języka źródłowego użytkownik systemu ma do wyboru jedynie trzy czasy gramatyczne: teraźniejszy, przyszły i przeszły. Natomiast w przypadku wybrania jako języków źródłowych innych języków należących do rodziny indoeuropejskiej liczba dostępnych opcji może być znacznie większa. Uwaga powyższa dotyczy zwłaszcza języków należących w ramach wielkiej rodziny indoeuropejskiej do grupy germańskiej i romańskiej, w przypadku których istnieje znacznie więcej czasów gramatycznych niż ma to miejsce na przykład w przypadku języków zaliczanych do grupy słowiańskiej.

W kolejnym etapie użytkownik systemu wielojęzycznego generatora form fleksyjnych musi dokonać wyboru osoby, liczby i rodzaju gramatycznego, w których wystąpić ma dana fraza czasownikowa. W istocie użytkownik systemu dokonuje wyboru osoby, liczby i rodzaju gramatycznego zaimka osobowego, który pełni rolę podmiotu zdania, a którego dana fraza czasownikowa jest dopełnieniem bliższym.

Jak wynika z postaci rys. 8 użytkownik dokonuje wyboru osoby, liczby i rodzaju gramatycznego uprzednio wskazanej przez niego frazy czasownikowej za pomocą specyficznego szablonu, poprzez zaznaczenie jednej z wielu możliwych opcji. Ponieważ opracowywany przez autorów system generatora form fleksyjnych fraz czasownikowych ma być z założenia systemem wielojęzycznym, dlatego przyjęto, że zaimki osobowe pełniące rolę podmiotu mogą także w przypadku pewnych języków wystąpić w także w liczbie podwójnej. Na

przykład liczba podwójna wciąż funkcjonuje w takich językach słowiańskich, takich jak słoweński oraz języki dolno- i górnołużycki. Ponadto formy liczby podwójnej występują także w literackim języku arabskim.

Dodatkowo uwzględniono fakt, że w różnych językach świata każdy z zaimków osobowych w pierwszej i drugiej osobie może przyjmować odrębne formy dla rodzaju męskiego i żeńskiego, a w osobie trzeciej mogą się także pojawić odrębne formy charakterystyczne dla rodzaju neutralnego

Rys. 8. Projektowany wygląd graficznego interfejsu użytkownika podczas etapu wyboru osoby, liczby i rodzaju gramatycznego

Po kliknięciu przez użytkownika przycisku „Dalej" na ekranie monitora ukazuje się uprzednio wskazana fraza czasownikowa występująca we właściwej formie fleksyjnej, ze względu na wybrany przez użytkownika czas i tryb gramatyczny, a także uwzględniwszy wybraną przez niego osobę, liczbę i rodzaj gramatyczny. Dodatkowo na ekranie monitora ukazują się tłumaczenia rozważanej frazy czasownikowej na wybrane przez użytkownika języki docelowe.

Jak już uprzednio wspomniano, różne języki świata mogą posiadać całkowicie odmienne systemy czasów i trybów gramatycznych. Z tego powodu uzyskiwane za pomocą projektowanego systemu przekłady odpowiednich form fleksyjnych fraz czasownikowych mają jedynie przybliżony charakter.

Zamieszczone powyżej rozważania dotyczyły przypadku, gdy użytkownik systemu wielojęzycznego generatora form fleksyjnych czasowników na etapie zilustrowanym na rys. 6 dokonał wyboru trybu orzekającego lub łączącego. Natomiast w przypadku dokonania wyboru trybu warunkowego etap zilustro-

wany na rys. 7 jest pomijany i następuje od razu przejście do etapu przedstawionego na rys. 8.

Z kolei w przypadku dokonania przez użytkownika wyboru trybu rozkazującego następuje przejście do etapu zilustrowanego za pomocą rys. 9, podczas którego ma miejsce dokonanie wyboru liczby (pojedynczej, podwójnej bądź mnogiej) oraz rodzaju gramatycznego (męskiego lub żeńskiego).

Jak wynika z postaci rys. 9, na rozważanym etapie domyślną opcją wyboru jest liczba pojedyncza i rodzaj męski, w związku z czym utworzona za pomocą wielojęzycznego generatora form fleksyjnych czasowników formuła rozkazująca adresowana jest do jednego osobnik płci męskiej. W przypadku gdyby rozpatrywana formuła rozkazująca była kierowana do kobiety, wówczas należałoby zaznaczyć opcję „rodzaj żeński". Oczywiście, w przypadku języka polskiego nie stanowi to żadnej różnicy, ale znane są języki świata, w przypadku których istnieją odrębne formy czasowników w drugiej osobie gramatycznej dla rodzaju męskiego i żeńskiego. Do tego rodzaju języków należą na przykład języki arabski i hebrajski.

Rys. 9. Projektowany wygląd graficznego interfejsu użytkownika podczas etapu wyboru liczby i rodzaju gramatycznego dla trybu rozkazującego

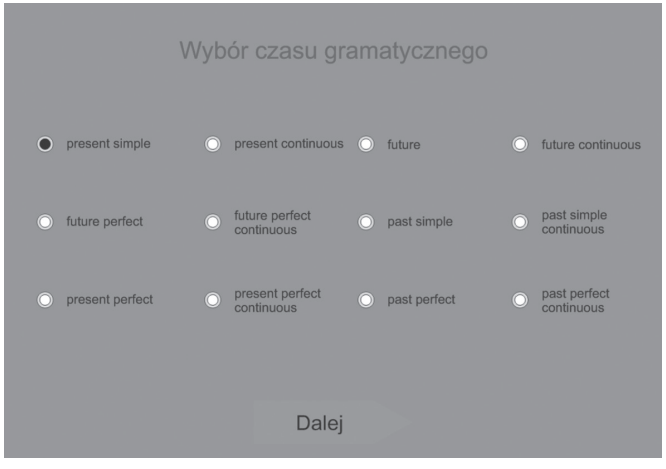
Jak już poprzednio wspomniano, w przypadku wielu języków świata system czasów i trybów gramatycznych może być o wiele bogatszy niż ma to miejsce w wypadku języka polskiego. W celu uwydatnienia ewentualnych różnic i ukazania ich wpływu na sposób funkcjonowania systemu w kolejnym punkcie omówiono bliżej kilka wybranych przypadków.

4. Analiza przypadków wybranych języków

W przypadku współczesnego języka angielskiego, podobnie jak ma to miejsce w języku polskim, tryb łączący czasowników nie występuje, chociaż w przypadku pewnych skostniałych związków frazeologicznych zachowały się jeszcze pewne jego relikty, jak na przykład w wyrażeniu „God save the king!". W każdym razie w projektowanym przez autorów systemie wielojęzycznego generatora form fleksyjnych czasowników trybu łączącego w przypadku języka angielskiego się nie uwzględnia, traktując go jako swoisty archaizm językowy. Podobnie zresztą jak w przypadku języka polskiego nie wprowadza się form czasu zaprzeszłego, który we współczesnej polszczyźnie wyszedł juz całkowicie z użycia.

Gdy użytkownik systemu po selekcji języka angielskiego jako języka źródłowego na etapie zilustrowanym na rys. 6 wybrał tryb orzekający lub łączący, wówczas ekran monitora powinien przybrać widok, jak pokazano na rys. 10.

Rys. 10. Projektowany wygląd graficznego interfejsu użytkownika podczas etapu wyboru czasu gramatycznego dla języka angielskiego



Z kolei gdy użytkownik zdecydował się na wybór trybu warunkowego, wówczas ekran monitora powinien przybrać wygląd, jak pokazano na rys. 11.

Rys. 11. Projektowany wygląd graficznego interfejsu użytkownika podczas etapu wyboru rodzaju trybu warunkowego dla języka angielskiego

Natomiast w przypadku dokonania przez użytkownika wyboru trybu rozkazującego ekran monitora przybiera wygląd identyczny, jak przedstawiono na rys. 9.

Z kolei w przypadku zdecydowania się użytkownika na użycie języka włoskiego w roli języka źródłowego i dokonania wyboru trybu orzekającego ekran monitora powinien przybrać wygląd, jak pokazano na rys. 12.

Rys. 12. Projektowany wygląd graficznego interfejsu użytkownika podczas etapu wyboru czasu gramatycznego w trybie orzekającym dla języka włoskiego

Natomiast w przypadku języka włoskiego, gdyby użytkownik zdecydował się na wybór trybu łączącego, wówczas ekran monitora powinien przybrać widok, jak przedstawiono na rys. 13.

Rys. 13. Projektowany wygląd graficznego interfejsu użytkownika podczas etapu wyboru czasu gramatycznego w trybie łączącym dla języka włoskiego

Z kolei w przypadku wyboru trybu warunkowego dla języka włoskiego ekran monitora przybiera widok, jak pokazano na rys. 14.

Rys. 14. Projektowany wygląd graficznego interfejsu użytkownika podczas etapu wyboru rodzaju trybu warunkowego dla języka włoskiego

Analogicznie, w przypadku wyboru języka portugalskiego jako języka źródłowego ekran monitora powinien przybrać widok, jak przedstawiono na rys. 15

Rys. 15. Projektowany wygląd graficznego interfejsu użytkownika podczas etapu wyboru czasu gramatycznego w trybie orzekającym dla języka portugalskiego

Podobnie, po wybraniu trybu łączącego dla języka portugalskiego ekran monitora przybiera wygląd, jak przedstawiono na rys. 16.

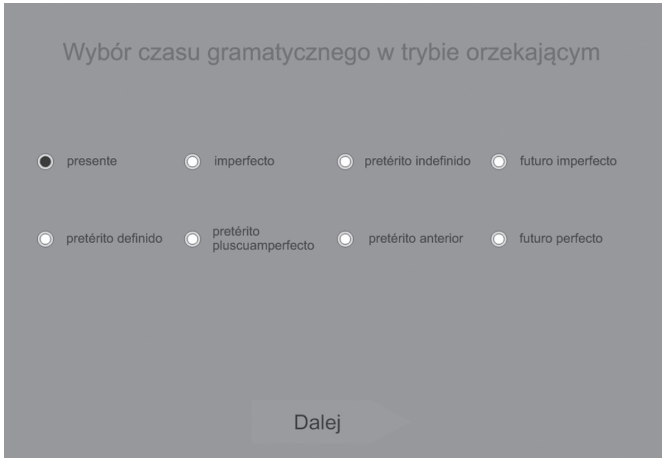
Rys. 16. Projektowany wygląd graficznego interfejsu użytkownika podczas etapu wyboru czasu gramatycznego w trybie łączącym dla języka portugalskiego

Natomiast po wybraniu trybu warunkowego dla języka portugalskiego ekran monitora przybiera wygląd, jak przedstawiono na rys. 17.

Rys. 17. Projektowany wygląd graficznego interfejsu użytkownika podczas etapu wyboru rodzaju trybu warunkowego dla języka portugalskiego

Podobnie w przypadku języka hiszpańskiego, gdy użytkownik wybrał tryb orzekający, wówczas ekran monitora powinien przybrać wygląd, jak przedstawiono na rys. 18.

Rys. 18. Projektowany wygląd graficznego interfejsu użytkownika podczas etapu wyboru czasu gramatycznego w trybie orzekającym dla języka hiszpańskiego

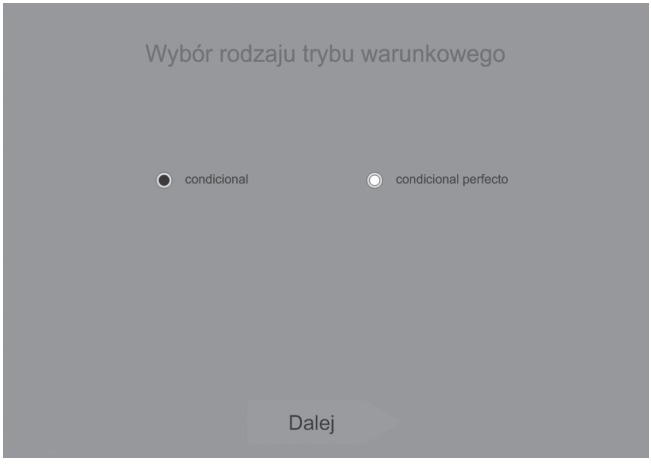


Z kolei w przypadku wybrania przez użytkownika trybu łączącego dla języka hiszpańskiego ekran monitora powinien przybrać widok, jak przedstawiono na rys. 19.

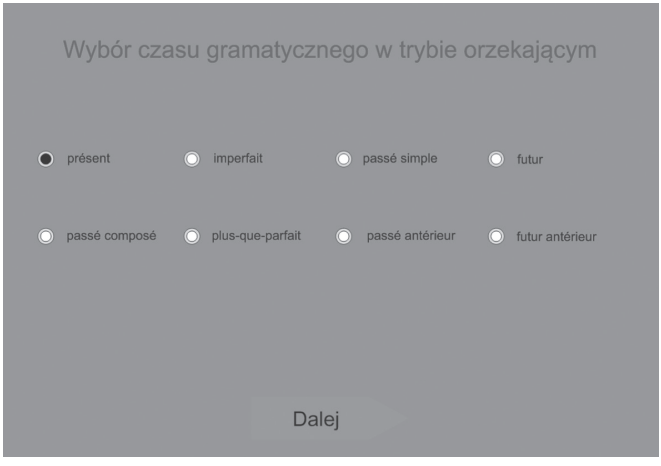
Rys. 19. Projektowany wygląd graficznego interfejsu użytkownika podczas etapu wyboru czasu gramatycznego w trybie łączącym dla języka hiszpańskiego

Analogicznie, w przypadku wyboru trybu warunkowego dla języka hiszpańskiego, ekran monitora przybiera widok, jak przedstawiono na rys. 20.

Rys. 20. Projektowany wygląd graficznego interfejsu użytkownika podczas etapu wyboru rodzaju trybu warunkowego dla języka hiszpańskiego



Z kolei dla języka francuskiego użytkownik w trybie orzekającym ma do wyboru czasy gramatyczne, jak przedstawiono na rys. 21.

Rys. 21. Projektowany wygląd graficznego interfejsu użytkownika podczas etapu wyboru czasu gramatycznego w trybie orzekającym dla języka francuskiego

Natomiast, gdy użytkownik wybrał dla języka francuskiego tryb łączący, wówczas dostępny wybór czasów gramatycznych jest taki, jak ukazano na rys. 22.

Rys. 22. Projektowany wygląd graficznego interfejsu użytkownika podczas etapu wyboru czasu gramatycznego w trybie łączącym dla języka francuskiego

Analogicznie, w przypadku wyboru dla języka francuskiego trybu warunkowego ekran monitora powinien przybrać wygląd, jak pokazano na rys. 23.

Rys. 23. Projektowany wygląd graficznego interfejsu użytkownika podczas etapu wyboru rodzaju trybu warunkowego dla języka francuskiego

Jak wynika z zamieszczonych powyżej przykładów, każdy język naturalny ma swoją specyfikę odnośnie występującego w nim systemu czasów i trybów. Z tego powodu każdy nowo dodany do systemu język musi zostać potrakto-

wany w pewien wyjątkowy sposób, co znajduje bezpośrednie odzwierciedlenie w wyglądzie powiązanego z nim graficznego interfejsu użytkownika.

W punkcie kolejnym zostaną bliżej omówione mechanizmy leżące u podstaw funkcjonowania wielojęzycznego generatora form fleksyjnych fraz czasownikowych.

4. Opis zasad funkcjonowania systemu

Funkcjonowanie opracowywanego przez autorów systemu wielojęzycznego generatora fraz czasownikowych oparte jest na systemie atrybutów, które mogą przyjmować wartości liczbowe całkowite.

Tego rodzaju atrybuty powiązane są, między innymi, z wyborem osoby, liczby i rodzaju gramatycznego. Na przykład atrybut „OsobaGramatyczna" może przyjąć jedną z dopuszczalnych wartości: 1, 2 lub 3, odpowiednio dla pierwszej, drugiej i trzeciej osoby gramatycznej.

Podobnie atrybut „LiczbaGramatyczna" może przyjąć jedną z dopuszczalnych wartości: 1, 2 lub 3, odpowiednio dla liczby pojedynczej, podwójnej i mnogiej.

Także atrybut „RodzajGramatyczny" może przyjąć jedną z dopuszczalnych wartości: 1, 2 lub 3, odpowiednio dla rodzaju męskiego, żeńskiego i nijakiego.

W tym miejscu warto zwrócić uwagę, że bynajmniej nie wszystkie kombinacje wartości atrybutów: „OsobaGramatyczna", „LiczbaGramatyczna" i „RodzajGramatyczny" są dopuszczalne. W szczególności atrybut „RodzajGramatyczny" może przyjąć wartość równą 3 jedynie w sytuacji, gdy atrybut „OsobaGramatyczna" przyjął również wartość równą 3, ponieważ rodzaj nijaki może pojawić się jedynie w przypadku trzeciej osoby gramatycznej.

Z kolei z wyborem przez użytkownika trybu gramatycznego wyselekcjonowanej przez niego frazy czasownikowej wiążą się takie atrybuty, jak: „TrybOrzekający", „TrybŁączący", „TrybWarunkowy" i „TrybRozkazujący". Wymienione atrybuty mogą przyjmować jedynie wartości binarne, czyli zero bądź jeden, przy czym przyjęcie przez jeden z nich wartości jeden automatycznie wyklucza taką możliwość w przypadku pozostałych atrybutów, zatem muszą one w takiej sytuacji posiadać wartość zero. Przykładowo, gdy użytkownik za pomocą menu przedstawionego na rys. 6 wybrał tryb warunkowy, wówczas rozważane atrybuty: „TrybOrzekający", „TrybŁączący", „TrybWarunkowy" i „TrybRozkazujący" przyjmą następujące wartości binarne:

TrybOrzekający = 0

TrybŁączący = 0

TrybWarunkowy = 1

TrybRozkazujący = 0

Podobnie, w związku z wyborem przez użytkownika odpowiedniego czasu gramatycznego wiążą się następujące atrybuty przyjmujące jedynie wartości binarne: „CzasTeraźniejszy", „CzasPrzyszły", „CzasPrzeszły" i „CzasZaprzeszły". Podobnie jak w przypadku wyboru trybu gramatycznego i tutaj obowiązuje zasada, że przyjęcie przez jeden z atrybutów: „CzasTeraźniejszy", „CzasPrzyszły", „CzasPrzeszły" lub „CzasZaprzeszły" wartości równej jeden automatycznie wyklucza taką możliwość w przypadku pozostałych atrybutów, które wówczas bezwzględnie muszą mieć wartość równą zero. Na przykład, gdyby użytkownik w przypadku języka francuskiego wybrał czas „pus-que-parfait", wówczas wymienione atrybuty „CzasTeraźniejszy", „CzasPrzyszły", „CzasPrzeszły" i „CzasZaprzeszły" przyjęłyby następujące wartości binarne:

CzasTeraźniejszy = 0

CzasPrzyszły = 0

CzasPrzeszły = 0

CzasZaprzeszły = 1

W tym miejscu warto poczynić małą dygresję. Otóż, w języku sztucznym ido, pełniącym w wypadku rozważanego systemu wielojęzycznego generatora gramatycznego fraz czasownikowych rolę języka pośredniczącego czas zaprzeszły w ogóle nie występuje. Jednak autorzy niniejszego opracowania postanowili dokonać swoistego rozszerzenia systemu czasów gramatycznych języka ido właśnie o czas zaprzeszły, ponieważ tego rodzaju kategoria gramatyczna występuje jeszcze w wielu językach indoeuropejskich i jej pominięcie prowadziłoby do istotnego zubożenia funkcjonalności projektowanego systemu.

Ponadto jeden z atrybutów „CzasTeraźniejszy", „CzasPrzyszły", „CzasPrzeszły" i „CzasZaprzeszły" może zawsze przyjąć wartość jeden w sytuacji, gdy użytkownik wybrał wcześniej tryb orzekający lub tryb łączący. Natomiast w przypadku wyboru trybu warunkowego wartość równą jeden może przyjąć jedynie atrybut „CzasPrzeszły" albo „CzasZaprzeszły".

Z kolei w przypadku dokonania przez użytkownika wyboru trybu rozkazującego wszystkie atrybuty „CzasTeraźniejszy", „CzasPrzyszły", „CzasPrzeszły"

i „CzasZaprzeszły" przyjmują wartość równą zero, ponieważ w takim wypadku czas gramatyczny danej frazy czasownikowej jest w ogóle nieistotny.

Cała potrzebna do funkcjonowania systemu wiedza lingwistyczna zawarta jest w odpowiednich lingwistycznych bazach danych, przy czym poszczególne jednostki leksykalne mają postać rekordów zapisanych w odpowiednich plikach. Na przykład w przypadku generowania form fleksyjnych fraz czasownikowych języka polskiego i ich późniejszego tłumaczenia na język pośredniczący przykładu, czyli sztuczny język ido, wiedza lingwistyczna niezbędna do funkcjonowania systemu wielojęzycznego generatora form fleksyjnych fraz czasownikowych zawarta jest w trzech plikach tekstowych, z których każdy związany jest z inną osobą gramatyczną.

Z pierwsza osobą gramatyczną związany jest plik noszący nazwę „Polish_IDO_Verb_P1.txt", z druga osobą gramatyczna powiązany jest plik „Polish_IDO_Verb_P2.txt", a z trzecia osoba gramatyczna skojarzony jest plik „Polish_IDO_Verb_P3.txt.

Budowa rekordów zawartych w wymienionych plikach zostanie omówiona na przykładzie polskojęzycznej frazy czasownikowej „jeść jabłko", którą tłumaczy się na sztuczny język ido jako „manjar pomo".

W związku z powyższym odpowiedni rekord danych lingwistycznych zapisany w pliku „Polish_IDO_Verb_P1.txt" ma postać:

***jeść|jabłko|jabłkajemjemy$będę jeść$będziemy jeść$jadłem$jadłam**
$jedliśmy$jadłyśmy$jadłbym$jadłabym$jedlibyśmy$jadłybyśmy%_manjar_pomo#

Zawarta w rozważanym rekordzie danych wiedza lingwistyczna pozwala systemowi wielojęzycznego generatora fraz czasownikowych na wygenerowanie następujących form fleksyjnych polskojęzycznej frazy czasownikowej „jeść jabłko" dla pierwszej osoby gramatycznej:

ja jem jabłko
my jemy jabłko
ja będę jeść jabłko
my będziemy jeść jabłko
ja jadłem jabłko
ja jadłam jabłko
my jedliśmy jabłko
my jadłyśmy jabłko

ja jadłbym jabłko
ja jadłabym jabłko
my jedlibyśmy jabłko
my jadłybyśmy jabłko

Ponadto za pomocą projektowanego przez autorów wielojęzycznego systemu generatora form fleksyjnych fraz czasownikowych możliwe jest także wygenerowanie form fleksyjnych polskojęzycznej frazy czasownikowej „jeść jabłko" także w formie przeczącej, czyli:

ja nie jem jabłka
my nie jemy jabłka
ja nie będę jeść jabłka
my nie będziemy jeść jabłka
ja nie jadłem jabłka
ja nie jadłam jabłka
my nie jedliśmy jabłka
my nie jadłyśmy jabłka
ja nie jadłbym jabłka
ja nie jadłabym jabłka
my nie jedlibyśmy jabłka
my nie jadłybyśmy jabłka

Jak wynika z zamieszczonego powyżej przykładu, każdy nowy rekord danych występujący w rozważanym pliku rozpoczyna się od znaku „*", po którym występuje bezokolicznik danego czasownika przechodniego, czyli w rozpatrywanym przypadku jest to czasownik „jeść". Następnie po znaku „|" umiejscowione jest dopełnienie bliższe rozważanego czasownika występujące w formie twierdzącej, czyli „jabłko". Podobnie po kolejnym znaku „|" umiejscowione jest dopełnienie bliższe rozważanego czasownika występujące w formie przeczącej, czyli „jabłka".

Następnie po kolejnych znakach „$" występują formy fleksyjne czasownika „jeść" charakterystyczne odpowiednio dla:

- pierwszej osoby liczby pojedynczej czasu teraźniejszego,
- pierwszej osoby liczby mnogiej czasu teraźniejszego,
- pierwszej osoby liczby pojedynczej czasu przyszłego,
- pierwszej osoby liczby mnogiej czasu przyszłego,
- pierwszej osoby liczby pojedynczej rodzaju męskiego czasu przeszłego,
- pierwszej osoby liczby pojedynczej rodzaju żeńskiego czasu przeszłego,
- pierwszej osoby liczby mnogiej rodzaju męskiego czasu przeszłego,
- pierwszej osoby liczby mnogiej rodzaju żeńskiego czasu przeszłego,

– pierwszej osoby liczby pojedynczej rodzaju męskiego trybu warunkowego,
– pierwszej osoby liczby pojedynczej rodzaju żeńskiego trybu warunkowego,
– pierwszej osoby liczby mnogiej rodzaju męskiego trybu warunkowego,
– pierwszej osoby liczby mnogiej rodzaju żeńskiego trybu warunkowego.

Następnie po znaku „%" występuje przekład polskiej frazy czasownikowej „jeść jabłko" na sztuczny język pośredniczący ido „_manjar_pomo", przy czym przyjęto konwencję zapisu, zgodnie z którą fraza czasownikowa w języku ido rozpoczyna się zawsze od znaku „_" oraz kolejne jej elementy są także rozdzielone znakami „_". Przyjęcie takiej formy zapisu pozwoli w przyszłości uniknąć błędów, polegających na przykład na zagubieniu pewnych elementów końcowych frazy czasownikowej występującej w języku ido.

Jak wynika z zamieszczonego przykładu, każdy nowy rekord danych lingwistycznych zakończony jest znakiem „#".

Z kolei odpowiedni rekord danych lingwistycznych zapisany w pliku „Polish_IDO_Verb_P2.txt" ma postać:

***jeść|jabłko|jabłka$jesz$jecie$będziesz jeść$będziecie jeść$jadłeś$jadłaś**
$jedliście$jadłyście$jadłbyś$jadłabyś$jedlibyście$jadłybyście
$jedz$jedzcie%_manjar_pomo#

Zawarta w rozważanym rekordzie danych wiedza lingwistyczna pozwala systemowi wielojęzycznego generatora fraz czasownikowych na wygenerowanie następujących form fleksyjnych polskojęzycznej frazy czasownikowej „jeść jabłko" dla drugiej osoby gramatycznej:

ty jesz jabłko
wy jecie jabłko
ty będziesz jeść jabłko
wy będziecie jeść jabłko
ty jadłeś jabłko
ty jadłaś jabłko
wy jedliście jabłko
wy jadłyście jabłko
ty jadłbyś jabłko
ty jadłabyś jabłko
wy jedlibyście jabłko

wy jadłybyście jabłko
jedz jabłko
jedźcie jabłko

Dodatkowo przy użyciu projektowanego przez autorów wielojęzycznego systemu generatora form fleksyjnych fraz czasownikowych możliwe jest także wygenerowanie form fleksyjnych polskojęzycznej frazy czasownikowej „jeść jabłko" także w formie przeczącej, czyli:

ty nie jesz jabłka
wy nie jecie jabłka
ty nie będziesz jeść jabłka
wy nie będziecie jeść jabłka
ty nie jadłeś jabłka
ty nie jadłaś jabłka
wy nie jedliście jabłka
wy nie jadłyście jabłka
ty nie jadłbyś jabłka
ty nie jadłabyś jabłka
wy nie jedlibyście jabłka
wy nie jadłybyście jabłka
nie jedz jabłka
nie jedźcie jabłka

W rozważanym rekordzie danych występują następujące formy fleksyjne polskiego czasownika „jeść", które są charakterystyczne dla:

- drugiej osoby liczby pojedynczej czasu teraźniejszego,
- drugiej osoby liczby mnogiej czasu teraźniejszego,
- drugiej osoby liczby pojedynczej czasu przyszłego,
- drugiej osoby liczby mnogiej czasu przyszłego,
- drugiej osoby liczby pojedynczej rodzaju męskiego czasu przeszłego,
- drugiej osoby liczby pojedynczej rodzaju żeńskiego czasu przeszłego,
- drugiej osoby liczby mnogiej rodzaju męskiego czasu przeszłego,
- drugiej osoby liczby mnogiej rodzaju żeńskiego czasu przeszłego,
- drugiej osoby liczby pojedynczej rodzaju męskiego trybu warunkowego,
- drugiej osoby liczby pojedynczej rodzaju żeńskiego trybu warunkowego,
- drugiej osoby liczby mnogiej rodzaju męskiego trybu warunkowego,
- drugiej osoby liczby mnogiej rodzaju żeńskiego trybu warunkowego,
- liczby pojedynczej trybu rozkazującego,
- liczby mnogiej trybu rozkazującego.

Analogicznie, odpowiedni rekord danych lingwistycznych zapisany w pliku „Polish_IDO_Verb_P3.txt" ma postać:

***jeść|jabłko|jabłkajejedzą$będzie jeść$będzą jeść$jadł$jadła$jadło $jedli$jadły$jadłby$jadłaby$jadłoby$jedliby$jadłyby%_manjar_pomo#**

Zawarta w rozważanym rekordzie danych wiedza lingwistyczna pozwala systemowi wielojęzycznego generatora fraz czasownikowych na wygenerowanie następujących form fleksyjnych polskojęzycznej frazy czasownikowej „jeść jabłko" dla trzeciej osoby gramatycznej:

on je jabłko
ona je jabłko
ono je jabłko
oni jedzą jabłko
one jedzą jabłko
on będzie jeść jabłko
ona będzie jeść jabłko
ono będzie jeść jabłko
oni będą jeść jabłko
one będą jeść jabłko
on jadł jabłko
ona jadła jabłko
ono jadło jabłko
oni jedli jabłko
one jadły jabłko
on jadłby jabłko
ona jadłaby jabłko
ono jadłoby jabłko
oni jedliby jabłko
one jadłyby jabłko

Dodatkowo przy użyciu projektowanego przez autorów wielojęzycznego systemu generatora form fleksyjnych fraz czasownikowych możliwe jest także wygenerowanie form fleksyjnych polskojęzycznej frazy czasownikowej „jeść jabłko" także w formie przeczącej, czyli:

on nie je jabłka
ona nie je jabłka
ono nie je jabłka
oni nie jedzą jabłka

one nie jedzą jabłka
on nie będzie jeść jabłka
ona nie będzie jeść jabłka
ono nie będzie jeść jabłka
oni nie będą jeść jabłka
one nie będą jeść jabłka
on nie jadł jabłka
ona nie jadła jabłka
ono nie jadło jabłka
oni nie jedli jabłka
one nie jadły jabłka
on nie jadłby jabłka
ona nie jadłaby jabłka
ono nie jadłoby jabłka
oni nie jedliby jabłka
one nie jadłyby jabłka

W rozważanym rekordzie danych występują następujące formy fleksyjne polskiego czasownika „jeść", które są charakterystyczne dla:

- trzeciej osoby liczby pojedynczej rodzaju męskiego czasu teraźniejszego,
- trzeciej osoby liczby pojedynczej rodzaju żeńskiego czasu teraźniejszego
- trzeciej osoby liczby mnogiej rodzaju męskoosobowego czasu teraźniejszego,
- trzeciej osoby liczby mnogiej rodzaju niemęskoosobowego czasu teraźniejszego,
- trzeciej osoby liczby pojedynczej rodzaju męskiego czasu przyszłego,
- trzeciej osoby liczby pojedynczej rodzaju żeńskiego czasu przyszłego
- trzeciej osoby liczby mnogiej rodzaju męskoosobowego czasu przyszłego,
- trzeciej osoby liczby mnogiej rodzaju niemęskoosobowego czasu przyszłego
- trzeciej osoby liczby pojedynczej rodzaju męskiego czasu przeszłego,
- trzeciej osoby liczby pojedynczej rodzaju żeńskiego czasu przeszłego,
- trzeciej osoby liczby pojedynczej rodzaju nijakiego czasu przeszłego,
- trzeciej osoby liczby mnogiej rodzaju męskoosobowego czasu przeszłego,
- trzeciej osoby liczby mnogiej rodzaju niemęskoosobowego czasu przeszłego,
- trzeciej osoby liczby pojedynczej rodzaju męskiego trybu warunkowego,

– trzeciej osoby liczby pojedynczej rodzaju żeńskiego trybu warunkowego,
– trzeciej osoby liczby pojedynczej rodzaju nijakiego trybu warunkowego,
– trzeciej osoby liczby mnogiej rodzaju męskoosobowego trybu warunkowego,
– trzeciej osoby liczby mnogiej rodzaju niemęskoosobowego trybu warunkowego.

Analogicznie, wiedza lingwistyczna potrzebna do przetłumaczenia frazy czasownikowej zapisanej w języku pośredniczącym ido na język polski zawarta jest dla poszczególnych osób gramatycznych w następujących plikach:
– IDO_Polish_Verb_P1.txt,
– IDO_Polish_Verb_P2.txt,
– IDO_Polish_Verb_P3.txt.

Zawartość pliku „IDO_Polish_Verb_P1.txt" powiązana jest z pierwszą osobą gramatyczną, a przykładowy rekord danych ma następującą postać:

***_manjar_pomojemjemy$zjem$zjemy$zjadłem$zjadłam $zjedliśmy$zjadłyśmy$zjadłbym$zjadłabym$zjedlibyśmy$zjadłybyśmy|jabłko|jabłka#**

W oparciu o zamieszczony powyżej rekord danych możliwe jest wygenerowanie następujących form fleksyjnych tłumaczenia na język polski frazy języka ido „_manjar_pomo" dla pierwszej osoby gramatycznej:

ja jem jabłko
my jemy jabłko
ja zjem jabłko
my zjemy jabłko
ja zjadłem jabłko
ja zjadłam jabłko
my zjedliśmy jabłko
my zjadłyśmy jabłko
ja zjadłbym jabłko
ja zjadłabym jabłko
my zjedlibyśmy jabłko
my zjadłybyśmy jabłko

Ponadto wymienione formy fleksyjne tłumaczenia na język polski frazy czasownikowej „_manjar_pomo" mogą również wystąpić w formie zaprzeczonej:

ja nie jem jabłka
my nie jemy jabłka

ja nie zjem jabłka
my nie zjemy jabłka
ja nie zjadłem jabłka
ja nie zjadłam jabłka
my nie zjedliśmy jabłka
my nie zjadłyśmy jabłka
ja nie zjadłbym jabłka
ja nie zjadłabym jabłka
my nie zjedlibyśmy jabłka
my nie zjadłybyśmy jabłka

Podobnie, zawartość pliku „IDO_Polish_Verb_P2.txt" powiązana jest z drugą osobą gramatyczną, a przykładowy rekord danych ma następującą postać:

***_manjar_pomo$jesz$jecie$zjesz$zjecie$zjadłeś$zjadłaś $zjedliście$zjadłyście$zjadłbyś$zjadłabyś$zjedlibyście$zjadłybyście $zjedz$zjedzcie|jabłko|jabłka#**

W oparciu o zamieszczony powyżej rekord danych możliwe jest wygenerowanie następujących form fleksyjnych tłumaczenia na język polski frazy języka ido „_manjar_pomo" dla drugiej osoby gramatycznej:

ty jesz jabłko
wy jecie jabłko
ty zjesz jabłko
wy zjecie jabłko
ty zjadłeś jabłko
ty zjadłaś jabłko
wy zjedliście jabłko
wy zjadłyście jabłko
ty zjadłbyś jabłko
ty zjadłabyś jabłko
wy zjedlibyście jabłko
wy zjadłybyście jabłko
zjedz jabłko
zjedźcie jabłko

Ponadto wymienione formy fleksyjne tłumaczenia na język polski frazy czasownikowej „_manjar_pomo" mogą również wystąpić w formie zaprzeczonej:

ty nie jesz jabłka
wy nie jecie jabłka

ty nie zjesz jabłka
wy nie zjecie jabłka
ty nie zjadłeś jabłka
ty nie zjadłaś jabłka
wy nie zjedliście jabłka
wy nie zjadłyście jabłka
ty nie zjadłbyś jabłka
ty nie zjadłabyś jabłka
wy nie zjedlibyście jabłka
wy nie zjadłybyście jabłka
nie zjedz jabłka
nie zjedźcie jabłka

Analogicznie, odpowiedni rekord danych lingwistycznych zapisany w pliku „IDO_Polish_Verb_P3.txt" ma postać:

***_manjar_pomojejedzązjezjedzą$zjadł$zjadła$zjadło $zjedli$zjadły$zjadłby$zjadłaby$zjadłoby$zjedliby$zjadłyby|jabłko|jabłka#**

W oparciu o zamieszczony powyżej rekord danych możliwe jest wygenerowanie następujących form fleksyjnych tłumaczenia na język polski frazy języka ido „_manjar_pomo" dla trzeciej osoby gramatycznej:

on je jabłko
ona je jabłko
ono je jabłko
oni jedzą jabłko
one jedzą jabłko
on zje jabłko
ona zje jabłko
ono zje jabłko
oni zjedzą jabłko
one zjedzą jabłko
on zjadł jabłko
ona zjadła jabłko
ono zjadło jabłko
oni zjedli jabłko
one zjadły jabłko
on zjadłby jabłko

ona zjadłaby jabłko
ono zjadłoby jabłko
oni zjedliby jabłko
one zjadłyby jabłko

Ponadto wymienione formy fleksyjne tłumaczenia na język polski frazy czasownikowej „_manjar_pomo" mogą również wystąpić w formie zaprzeczonej:

on nie je jabłka
ona nie je jabłka
ono nie je jabłka
oni nie jedzą jabłka
one nie jedzą jabłka
on nie zje jabłka
ona nie zje jabłka
ono nie zje jabłka
oni nie zjedzą jabłka
one nie zjedzą jabłka
on nie zjadł jabłka
ona nie zjadła jabłka
ono nie zjadło jabłka
oni nie zjedli jabłka
one nie zjadły jabłka
on nie zjadłby jabłka
ona nie zjadłaby jabłka
ono nie zjadłoby jabłka
oni nie zjedliby jabłka
one nie zjadłyby jabłka

Na zakończenie warto jeszcze zwrócić uwagę, że w sztucznym języku ido w ogóle nie istnieje taka kategoria gramatyczna jak aspekt czasownika, natomiast w języku polskim zdecydowana większość czasowników występuje zarówno w formie dokonanej, jak i niedokonanej. Z tego powodu fraza czasownikowa języka ido „manjar pomo" jest semantycznym odpowiednikiem zarówno polskiej frazy czasownikowej „jeść jabłko", jak i „zjeść jabłko". Dlatego niezależnie od tego, czy użytkownik wybierze z menu frazę „jeść jabłko" czy też „zjeść jabłko", to w obu wypadkach frazy te zostaną zastąpione w języku ido jednostką leksykalną „_manjar_pomo".

Niestety, jak widać na zamieszczonym przykładzie, wprowadzenie tłumaczenia automatycznego z wykorzystaniem języka pośredniczącego przekładu może prowadzić do utraty pewnej porcji informacji zawartej w tekście wyj-

ściowym. Jest to zarazem ceną, jaką trzeba zapłacić za uzyskanie możliwości stosunkowo łatwej budowy systemu wielojęzycznego dzięki wprowadzeniu koncepcji języka pośredniczącego przekładu (interlingua).

Z wymienionych powyżej powodów, gdyby użytkownik systemu wybrał w roli zarówno języka źródłowego przekładu, jak i języka docelowego język polski, wówczas jako tłumaczenie frazy „on będzie jeść jabłko" otrzyma nieco zmodyfikowana jej parafrazę w postaci „on zje jabłko", co spowodowane jest, jak już uprzednio wspomniano brakiem występowania w języku ido gramatycznej kategorii aspektu czasownika.

5. Podsumowanie

W niniejszym opracowaniu autorzy przedstawili koncepcję budowy wielojęzycznego generatora form fleksyjnych fraz czasownikowych. Proponowane rozwiązanie stanowi nowatorskie podejście, ponieważ do chwili obecnej według najlepszej wiedzy autorów nie są znane jeszcze tego typu rozwiązania. Dostępne za pomocą Internetu generatory form fleksyjnych czasowników, które opracowano już dla wielu różnych języków świata, są w stanie wypisać jedynie ich poszczególne formy fleksyjne bez możliwości ich dalszego automatycznego tłumaczenia na inne języki. Natomiast proponowany przez autorów generator form fleksyjnych fraz czasownikowych jest z założenia systemem wielojęzycznym, który umożliwia komputerowy przekład odmienianych fraz czasownikowych również na inne języki, które zostały wbudowane w opracowywany system.

Dużą zaletą proponowanego przez autorów rozwiązania jest wprowadzenie automatycznej odmiany nie tyle pojedynczych czasowników, co całych fraz czasownikowych, które zbudowane są najczęściej z czasownika przechodniego i jego dopełnienia bliższego, ewentualnie pomiędzy tymi elementami mogą wystąpić odpowiednie przyimki. Tego rodzaju frazy czasownikowe, w przeciwieństwie do izolowanych czasowników, posiadają zwykle już dość precyzyjnie określoną semantykę, co umożliwia tym samym ich tłumaczenie z odpowiednio wysoka jakością także na inne wybrane przez użytkownika systemu języki.

Niewątpliwy atutem proponowanego rozwiązania jest oparcie działania systemu wielojęzycznego generatora form fleksyjnych fraz czasownikowych na sztucznym języku ido. Dzięki takiemu podejściu potencjalnie łatwą staje się rozbudowa projektowanego systemu o dalsze języki. Po prostu, w takim wypadku należy dla każdego nowo dodanego języka opracować moduł tłumaczenia na język ido oraz moduł przekładający z języka ido na nowo dodany do systemu język.

W głębokim przekonaniu autorów proponowany system może stanowić niezwykle interesujące i przydatne narzędzie, które może być z powodzeniem wykorzystywane przez językoznawców (pracujących zwłaszcza w obszarze lingwistyki kontrastywnej – porównywanie ze sobą systemów gramatycznych wybranych języków), a także filologów, przekładoznawców, tłumaczy oraz osoby zainteresowane nauką danych języków obcych. Jak już uprzednio wspomniano, planowany system ma za zadanie umożliwić nie tylko generowanie odpowiednich form fleksyjnych czasowników, ale ma zarazem stanowić swego rodzaju repozytorium związków frazeologicznych, w które mogą wchodzić poszczególne czasowniki podlegające odmianie przez czasy, tryby, osoby, liczby i rodzaje gramatyczne, co jest niezmiernie istotne z punktu widzenia zagadnień translatoryki, ponieważ pozwala na prawidłowe użycie odpowiednich związków składniowych w tłumaczonych na język obcy zdaniach.

Bibliografia

Majewicz A.F., *Języki świata i ich klasyfikowanie*, Państwowe Wydawnictwo Naukowe, Warszawa 1989.

Dalewska-Greń H., *Języki słowiańskie*, Wydawnictwo Naukowe PWN, Warszawa 2002.

Danecki J., *Gramatyka języka arabskiego*, Wydawnictwo Akademickie Dialog, Warszawa 1994.

Hutchins W.J., *Machine translation – past, present, future*, Ellis Horwood Series in Computers and their Applications, London 1986.

Arnold D., Balkan L., Meijer S., Humphreys R.L., Sadler L., *Machine translation: an introductory guide*, NCC Blackwell, London 1994.

Allen J.F., *Natural language understanding*, The Benjamin/Cummings Publishing Company, New York 1995.

Whitelock P., Kilby K., *Linguistic and computational techniques in machine translation system design*, UCL Press, London 1995.

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" 2003, vol. 51, no. 1.

Gajer M., *The implementation of the example-based machine translation technique for Greek-to-Polish automatic translation system*, "Foundation of Computing and Decision Sciences" 2003, vol. 28, no. 2.

Gajer M., *The implementation of the example-based machine translation technique for Norwegian-to-Polish automatic translation system*, "Control and Cybernetics" 2004, vol. 34, no. 2.
Gajer M., *The pattern-based French-to-Polish machine translation system*, "Machine Translation Review" 2002, no. 13.
Gajer M., *Towards a fully-automatic high-quality machine translation system for unrestricted text*, "Machine Translation Review" 2001, no. 12.
Gajer M., *Specialized fully automatic machine translation system delivering high quality of translated texts*, „Task Quarterly" 2009, vol. 13, no. 4.
Gajer M., *Analiza możliwości i ograniczeń systemów translacji automatycznej wspomaganej przez człowieka na przykładzie systemu tłumaczącego z języka szwedzkiego na polski*, „Kwartalnik Elektroniki i Telekomunikacji" 2003, t. 49, z. 1.
Gajer M., *System translacji automatycznej oparty na uogólnionych przykładach translacyjnych*, „Kwartalnik Elektroniki i Telekomunikacji" 2003, t. 49, z. 3.
Gajer M., *Zastosowanie algorytmów ewolucyjnych w lingwistyce komputerowej celem oceny stopnia pokrewieństwa języków naturalnych*, „Kwartalnik Elektroniki i Telekomunikacji" 2005, t. 51, z. 2.
Gajer M., *The implementation of the example-based machine translation technique for Persian-to-Polish automatic translation system*, „Folia Societatis Scientiarum Lublinensis" 2002, vol. 11.
Gajer M., *The implementation of the example-based machine translation technique for French-to-Polish automatic translation system*, „Task Quarterly" 2002, vol. 6, no. 3.
Gajer M., *The implementation of the example-based machine translation technique for German-to-Polish automatic translation system*, „Informatica 2002", vol. 13, no. 4.
Gajer M., *The proposition of using the example-based machine translation technique for Italian-to-Polish automatic translation system*, „Studia Informatica" 2002, vol. 23, no. 4.
Gajer M., *Zastosowanie metody Example-based machine translation w specjalizowanych systemach komputerowego przekładu*, „Przegląd Elektrotechniczny" 2011, r. LXXXVII, no. 2.
Szczepaniak L., Królikowski Z., *Kontrolowane języki naturalne – przegląd rozwiązań i zastosowań*, „Pro Dialog" 2000, nr 11.
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" 2014, nr 4.

Canals R., Esteve A., Garrido A., *InterNOSTRUM: a Spanish-Catalan machine translation system*, „Machine Translation Review" 2000, no. 11.

Fukutomi O., *Report on commercial machine translation in a manufacturing industry domain*, „Machine Translation Review" 1999, no. 10.

Guidère M., *Toward corpus-based machine translation for standard Arabic*, „Translation Journal" 2002, vol. 6, no. 1.

Somers H., *EBMT seen as case-based reasoning*, MT Summit VIII, Santiago de Compostela, 2001.

Sumita E., Iida H., Experiments and prospects of example-based machine translation, Proceedings of the 29th Annulal Meeting of the Association for Computational Linguistics, 1991, 185–192.

Turcato D., Popowich F., *What is example-based machine translation?*, MT Summit VIII, Santiago de Compostela, 2001.