

# ANALIZA MOŻLIWOŚCI ZASTOSOWANIA TECHNIKI TRANSLACJI AUTOMATYCZNEJ OPARTEJ NA WZORCACH DLA WYBRANYCH JĘZYKÓW NALEŻĄCYCH DO GRUP JĘZYKOWYCH GERMAŃSKIEJ I ROMAŃSKIEJ

## STRESZCZENIE

*Translacja automatyczna jest działem zastosowań informatyki, którego zadaniem jest wypracowanie metod algorytmizacji przekładu tekstów napisanych w pewnym języku naturalnym na inny język naturalny. Pomimo znacznego wysiłku badawczego włożonego w budowę systemów automatycznej translacji, automatyzacja przekładu dla tekstów należących do języka ogólnego wciąż nie doczekała się w pełni satysfakcjonującego rozwiązania. Jakość przekładów automatycznych wprawdzie z roku na rok wzrasta, ale systemom translacji automatycznej wciąż jest daleko do pożądanego ideału, który być może jest w ogóle nieosiągalny. Artykuł niniejszy stanowi drugą część cyklu publikacji, poświęconych opracowanej przez autora metodzie translacji automatycznej opartej na wzorcach PBMT (Pattern-Based Machine Translation). Metoda PBMT stanowi kombinację znanych wcześniej metod automatycznego przekładu opartych na przykładach EBMT (Example-Based Machine Translation) oraz na regułach gramatycznych (Rule-Based Machine Translation). W artykule zamieszczono wyniki rozważań dotyczących analizy możliwości i ograniczeń zastosowania metody PBMT w celu automatyzacji przekładu pomiędzy wybranymi językami germańskimi i romańskimi a językiem polskim.*

**Słowa kluczowe:** *lingwistyka informatyczna, translacja automatyczna, translacja oparta na wzorcach*

## THE ANALYSIS OF THE POSSIBILITIES OF IMPLEMENTATION OF THE PATTERN-BASED MACHINE TRANSLATION TECHNIQUE FOR THE GROUPS OF GERMANIC AND ROMANCE LANGUAGES

*Machine translation is a discipline of computer science the purpose of which is to deliver the methods of automatization of translation of texts written in some natural language into another. Despite the great scientific effort connected with the development of machine translation systems for the general language, the automatization of translation is still far away from its fully satisfactory solution. Besides that the quality of automatic translation is growing systematically, the fully-automatic high-quality machine translation system for unrestricted text is still something ideal that maybe will never be attained. The paper is the second one of the longer series of publications of this author devoted to the method of machine translation that was developed by himself. The method was called Pattern-Based Machine Translation (PBMT) and is a hybride of Example-Based Machine Translation and Rule-Based Machine Translation. The paper is a summary of the analysis of possibilities of implementation of PBMT for machine translation systems that translate from certain Germanic and Romance languages into the Polish language.*

**Keywords:** *computational linguistics, machine translation, pattern-based translation*

## 1. WPROWADZENIE

Artykuł niniejszy jest drugim artykułem z większego cyklu publikacji autora, poświęconego opracowanej przez niego metodzie translacji automatycznej opartej na wzorcach PBMT (Pattern-Based Machine Translation). W artykule dokonano analizy możliwości i ograniczeń zastosowania techniki PBMT dla wybranych języków germańskich i romańskich. W sumie pod uwagę wzięto dziesięć języków, w tym sześć germańskich (angielski, niemiecki, niderlandzki, szwedzki, duński i norweski) oraz cztery języki romańskie (francuski, włoski, hiszpański i portugalski) [22].

Łatwość zastosowania metody PBMT dla danego języka zależy w głównej mierze od stopnia jego fleksyjności. Im

więcej jest w danym języku form fleksyjnych, tym większą liczbę wzorców translacyjnych należy zgromadzić w bazie danych. Ponieważ każdy system PBMT jest tak dobry, jak dobra jest jego baza danych zawierająca wzorce translacyjne, wstępne oszacowanie rozmiarów tej bazy danych jest sprawą kluczową. Bowiern dopiero wtedy, gdy wstępnie znana będzie wymagana pojemność bazy danych, można przystąpić do prac nad jej tworzeniem na podstawie odpowiedniego bilingwicznego korpusu tekstów. Wtedy możliwe jest również oszacowanie czasu potrzebnego do opracowania rozważanej bazy danych oraz liczebności zespołu lingwistów, którzy będą z bilingwicznymi korpusami tekstów pracować [21].

W związku z powyższym, w kolejnych punktach artykułu zamieszczono wyniki prób oszacowania rozmiarów baz

\* Katedra Automatyki AGH

danych zawierających wzorce translacyjne dla wymienionych języków należących do germańskiej i romańskiej grupy językowej.

## 2. SZACOWANIE ROZMIARÓW BAZY DANYCH DLA JĘZYKA ANGIELSKIEGO

Język angielski należy do zachodniej gałęzi germańskiej grupy językowej. Jego cechą charakterystyczną jest niewielka liczba form fleksyjnych wyrazów [1].

W języku angielskim formy fleksyjne wyrazów można utworzyć dla rzeczowników, czasowników oraz przymiotników.

Angielski rzeczownik oprócz swej formy podstawowej posiada również formę liczby mnogiej oraz formy dzierżawcze dla liczby pojedynczej i mnogiej. Fleksja angielskiego rzeczownika została zilustrowana na przykładach zamieszczonych w tabeli 1.

**Tabela 1.** Zestawienie form fleksyjnych dla wybranych rzeczowników język angielskiego

Forma podstawowa	Forma liczby mnogiej	Forma dzierżawcza dla liczby pojedynczej	Forma dzierżawcza dla liczby mnogiej	Znaczenie
knife	knives	knife's	knives'	nóż
life	lives	life's	lives'	życie
thief	thieves	thief's	thieves'	złodziej
wolf	wolves	wolf's	wolves'	wilk
wife	wives	wife's	wives'	żona
woman	women	woman's	women's	kobieta
mouse	mice	mouse's	mice's	mysz
child	children	child's	children's	dziecko
worker	workers	worker's	workers'	robotnik
teacher	teachers	teacher's	teachers'	nauczyciel
ox	oxen	ox'es	oxen's	wół

Zatem współczynnik fleksyjności angielskiego rzeczownika wynosi  $f_{GB}(R) = 4$ .

Z kolei fleksja angielskiego czasownika została przedstawiona na przykładzie odmiany regularnego czasownika „to walk” (spacerować). W tabeli 2 zamieszczono wzór odmiany czasownika „to walk” w czasie teraźniejszym (Present Simple).

**Tabela 2.** Wzór odmiany czasownika „to walk” w czasie teraźniejszym (Present Simple)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	walk	walk
Druga	walk	walk
Trzecia	walks	walk

Z kolei w tabeli 3 zamieszczono wzór odmiany czasownika „to walk” w czasie przeszłym (Past Simple).

**Tabela 3.** Wzór odmiany czasownika „to walk” w czasie przeszłym (Past Simple)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	walked	walked
Druga	walked	walked
Trzecia	walked	walked

W związku z powyższym współczynnik fleksyjności angielskiego czasownika wynosi  $f_{GB}(C) = 3$ .

Z kolei dla angielskiego przymiotnika można utworzyć dodatkowe formy fleksyjne dla stopnia wyższego i najwyższego, co zostało zilustrowane na wybranych przykładach w tabeli 4.

**Tabela 4.** Zastawienie form stopnia wyższego i najwyższego wybranych przymiotników języka angielskiego

Forma podstawowa	Forma stopnia wyższego	Forma stopnia najwyższego	Znaczenie
good	better	best	dobry
bad	worse	worst	zły
far	farther	farthest	daleki
old	older	oldest	stary
easy	easier	easiest	łatwy
poor	poorer	poorest	biedny
narrow	narrower	narrowest	wąski
polite	politer	politest	uprzejmy
clear	clearer	clearest	czysty
tall	taller	tallest	wysoki
big	bigger	biggest	duży

Zatem współczynnik fleksyjności angielskiego przymiotnika wynosi  $f_{GB}(P) = 3$ .

Wartość współczynnika fleksyjności języka angielskiego może zostać wyznaczona ze wzoru

$$f_{GB} = w_{GB}(R) \cdot f_{GB}(R) + w_{GB}(C) \cdot f_{GB}(C) + w_{GB}(P) \cdot f_{GB}(P) \quad (1)$$

Występujące we wzorze (1) współczynniki wagowe  $w_{GB}(R)$ ,  $w_{GB}(C)$  oraz  $w_{GB}(P)$  określają częstość występowania rzeczowników, czasowników i przymiotników w słownictwie języka angielskiego.

Częstotliwość występowania rzeczowników, czasowników i przymiotników w słownictwie języka angielskiego została wyznaczona na podstawie analizy następującego

zbioru zawierającego sto haseł zaczerpniętych kolejno ze słownika [2].

- |                        |                       |
|------------------------|-----------------------|
| 1) bend-leather (R),   | 51) beryl (R),        |
| 2) beneaped (P),       | 52) beseech (C),      |
| 3) beneath (X),        | 53) beseeching (P),   |
| 4) benedick (R),       | 54) beseem (C),       |
| 5) benedictine (R),    | 55) beseeming (P),    |
| 6) benediction (R),    | 56) beset (C),        |
| 7) benefaction (R),    | 57) besetment (R),    |
| 8) benefactor (R),     | 58) besetting (P),    |
| 9) benefactress (R),   | 59) beshrew (X),      |
| 10) benefice (R),      | 60) beside (X),       |
| 11) beneficence (R),   | 61) besides (X),      |
| 12) beneficent (P),    | 62) besiege (C),      |
| 13) beneficial (X),    | 63) besieger (R),     |
| 14) beneficiary (R),   | 64) beslaver (C),     |
| 15) benefit (R),       | 65) beslobber (C),    |
| 16) benefit (C),       | 66) beslubber (C),    |
| 17) benevolence (R),   | 67) besmear (C),      |
| 18) benevolent (P),    | 68) besmirch (C),     |
| 19) benighted (P),     | 69) besom (R),        |
| 20) benign (P),        | 70) besot (C),        |
| 21) benignant (P),     | 71) bespangle (C),    |
| 22) benignity (R),     | 72) bespatter (C),    |
| 23) benjamin (R),      | 73) bespeak (C),      |
| 24) benjamin-tree (R), | 74) bespectacled (P), |
| 25) bennet (R),        | 75) bespoke (P),      |
| 26) bent (R),          | 76) besprent (P),     |
| 27) bent (C),          | 77) besprinkle (C),   |
| 28) bent-grass (R),    | 78) best (P),         |
| 29) benthos (R),       | 79) best (C),         |
| 30) benumb (C),        | 80) bestead (C),      |
| 31) benumbed (P),      | 81) bested (P),       |
| 32) benzene (R),       | 82) bestial (P),      |
| 33) benzine (R),       | 83) bestiality (R),   |
| 34) benzoin (R),       | 84) bestir (C),       |
| 35) benzol (R),        | 85) bestow (C),       |
| 36) bequeath (C),      | 86) bestowal (R),     |
| 37) bequest (R),       | 87) bestrew (C),      |
| 38) berate (C),        | 88) bestride (C),     |
| 39) bereave (C),       | 89) best-seller (R),  |
| 40) bereaved (R),      | 90) bestudded (P),    |
| 41) bereavement (R),   | 91) bet (R),          |
| 42) beret (R),         | 92) bet (C),          |
| 43) bergamot (R),      | 93) beta (R),         |
| 44) beriberi (R),      | 94) betake (C),       |
| 45) berlin (R),        | 95) beta-rays (R),    |
| 46) berm (R),          | 96) betel (R),        |
| 47) berry (R),         | 97) betel-nuts (R),   |
| 48) berserker (R),     | 98) bethink (C),      |
| 49) berth (R),         | 99) betide (C),       |
| 50) berththa (R),      | 100) betimes (X)      |

W zestawieniu zastosowano następujące oznaczenia części mowy:

- R – rzeczownik,
- C – czasownik,
- P – przymiotnik,
- X – pozostałe części mowy języka angielskiego nie podlegające fleksji.

Zestawienie to zawiera 100 wyrazów, wśród których można wyróżnić:

- 47 – rzeczowników,
- 30 – czasowników,
- 17 – przymiotników.

Stąd współczynnikiem wagowym przypisane zostały następujące wartości:

$$w_{GB}(R) = 0,47,$$

$$w_{GB}(C) = 0,30,$$

$$w_{GB}(P) = 0,17.$$

Zatem wartość współczynnika fleksyjności języka angielskiego może zostać oszacowana, zgodnie ze wzorem (1), jako  $f_{GB} = 3,29$ .

Natomiast w tabeli 5 zamieszczono zestawienie dostarczające informacji o tym, ile podczas tłumaczenia (metodą PBMT) pewnego próbnego tekstu wystąpiło wzorców translacyjnych o długości jednego wyrazu ( $l_1$ ), dwóch wyrazów ( $l_2$ ), trzech wyrazów ( $l_3$ ), czterech wyrazów ( $l_4$ ), pięciu wyrazów ( $l_5$ ) oraz dłuższych niż pięć wyrazów ( $l_{\geq 6}$ ). Podczas przeprowadzonych obliczeń, jako oddzielnych wyrazów nie liczono angielskich przedimków „a” i „the”.

**Tabela 5.** Statystyka wzorców translacyjnych

$l_1$	$l_2$	$l_3$	$l_4$	$l_5$	$l_{\geq 6}$
64	43	16	11	3	2

Minimalna liczba wzorców translacyjnych  $L_{GB}$ , jakie powinna zawierać baza danych, potrzebnych do sprawnego funkcjonowania systemu automatycznej translacji z języka angielskiego na polski, może zostać oszacowana za pomocą następującego wzoru

$$L_{GB} = S_{GB} \cdot f_{GB} \cdot \left( 1 + \frac{l_2 + l_3 + l_4 + l_5 + l_{\geq 6}}{l_1} \right) \quad (2)$$

We wzorze (2) parametr  $S_{GB}$  oznacza liczebność zbioru haseł słownika języka angielskiego. Wartość parametru  $S_{GB}$  szacowana może być w granicach od 10 000 (w przypadku uwzględnienia jedynie podstawowego słownictwa języka angielskiego) do 100 000 (w przypadku dodatkowego uwzględnienia słownictwa specjalistycznego). Na przykład słownik [3] zawiera około 50 000 haseł.

Zatem przyjmując  $S_{GB} = 50 000$  otrzymuje się oszacowanie minimalnej liczby wzorców translacyjnych  $L_{GB} = 375 300$ .

### 3. SZACOWANIE ROZMIARÓW BAZY DANYCH DLA JĘZYKA NIEMIECKIEGO

Język niemiecki należy do zachodniej podgrupy języków germańskich. Język ten zachował znaczną liczbę form fleksyjnych wyrazów.

W języku niemieckim częściami mowy, dla których można utworzyć formy fleksyjne są rzeczowniki, czasowniki oraz przymiotniki.

W języku niemieckim rzeczownik odmienia się przez przypadki. W tabeli 6 zamieszczono wzór odmiany rzeczownika „der Mensch” (człowiek).

**Tabela 6.** Wzór odmiany rzeczownika „der Mensch”

Przypadek	Liczba pojedyncza	Liczba mnoga
Nominative	der Mensch	die Menschen
Genitive	des Menschen	der Menschen
Dative	dem Menschen	den Menschen
Akkusative	den Menschen	die Menschen

Z kolei w tabeli 7 zamieszczono wzór deklinacji rzeczownika „der Wald” (las).

**Tabela 7.** Wzór odmiany rzeczownika „der Wald”

Przypadek	Liczba pojedyncza	Liczba mnoga
Nominative	der Wald	die Wälder
Genitive	des Waldes	der Wälder
Dative	dem Wald	den Wäldern
Akkusative	den Wald	die Wälder

Natomiast w tabeli 8 zamieszczono wzór odmiany rzeczownika „die Wand” (ściana).

**Tabela 8.** Wzór odmiany rzeczownika „die Wand”

Przypadek	Liczba pojedyncza	Liczba mnoga
Nominative	die Wand	die Wände
Genitive	der Wand	der Wände
Dative	der Wand	den Wänden
Akkusative	die Wand	die Wände

Jak wynika z powyższych zestawień, w języku niemieckim rzeczownik może mieć dwie, trzy lub cztery różne formy fleksyjne. Zatem współczynnikowi fleksyjności niemieckiego rzeczownika może zostać przypisana średnia wartość  $f_D(R) = 3$ .

W języku niemieckim również czasowniki posiadają wiele form fleksyjnych. Fleksja niemieckiego czasownika zostanie zilustrowana na przykładzie odmiany czasownika „kaufen” (kupować).

W tabeli 9 zamieszczono wzór koniugacji czasownika „kaufen” w czasie teraźniejszym (Präsens).

**Tabela 9.** Odmiana czasownika „kaufen” w czasie teraźniejszym (Präsens)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	kaufe	kaufen
Druga	kaufst	kauft
Trzecia	kauft	kaufen

Z kolei w tabeli 10 zamieszczono wzór koniugacji czasownika „kaufen” w czasie przeszłym (Imperfekt).

**Tabela 10.** Odmiana czasownika „kaufen” w czasie przeszłym (Imperfekt)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	kaufte	kauften
Druga	kauftest	kauftet
Trzecia	kaufte	kauften

Ponadto czasownik „kaufen” posiada dodatkową formę fleksyjną dla imiesłowu czasu przeszłego dokonanego (Perfekt) „gekauft”.

Zatem współczynnik fleksyjności niemieckiego czasownika posiada wartość  $f_D(C) = 10$ .

W języku niemieckim przymiotniki podlegają deklinacji. Zjawisko to zostało zilustrowane na następujących przykładach.

W tabeli 11 zamieszczono wzór odmiany niemieckiego przymiotnika „breit” (szeroki) dla rodzaju męskiego.

**Tabela 11.** Wzór odmiany przymiotnika „breit” dla rodzaju męskiego

Przypadek	Liczba pojedyncza	Liczba mnoga
Nominative	breiter	breite
Genitive	breiten	breiter
Dative	breitem	breiten
Akkusative	breiten	breite

Z kolei w tabeli 12 zamieszczono wzór odmiany niemieckiego przymiotnika „breit” (szeroki) dla rodzaju żeńskiego.

**Tabela 12.** Wzór odmiany przymiotnika „breit” dla rodzaju żeńskiego

Przypadek	Liczba pojedyncza	Liczba mnoga
Nominative	breite	breite
Genitive	breiter	breiter
Dative	breiter	breiten
Akkusative	breite	breite

Analogicznie w tabeli 13 zamieszczono wzór odmiany niemieckiego przymiotnika „breit” (szeroki) dla rodzaju nijakiego.

**Tabela 13.** Wzór odmiany przymiotnika „breit” dla rodzaju nijakiego

Przypadek	Liczba pojedyncza	Liczba mnoga
Nominative	breites	breite
Genitive	breiten	breiter
Dative	breitem	breiten
Akkusative	breites	breite

Ponadto niemiecki przymiotnik podlega jeszcze stopniowaniu, z czym wiąże się występowanie dodatkowych form fleksyjnych. Na przykład niemiecki przymiotnik „breit” w stopniu wyższym posiada formy „breiter”, „der, die, das breitere” oraz „die breiteren”. Podobnie w stopniu najwyższym występują formy „am breitsten”, „der, die, das breitste” oraz „die breitsten”.

Uwzględniając powyższe formy odmiany, można stwierdzić, że współczynnik fleksyjności niemieckiego przymiotnika wynosi  $f_D(P) = 10$ .

Wartość współczynnika fleksyjności języka niemieckiego może zostać wyznaczona ze wzoru

$$f_D = w_D(R) \cdot f_D(R) + w_D(C) \cdot f_D(C) + w_D(P) \cdot f_D(P) \quad (3)$$

Występujące we wzorze (3) współczynniki wagowe  $w_D(R)$ ,  $w_D(C)$  oraz  $w_D(P)$  określają częstość występowania rzeczowników, czasowników i przymiotników w słownictwie języka niemieckiego.

Częstotliwość występowania rzeczowników, czasowników i przymiotników w słownictwie języka niemieckiego została wyznaczona na podstawie analizy następującego zbioru zawierającego sto haseł pobranych kolejno ze słownika [4].

- |                      |                     |
|----------------------|---------------------|
| 1) Fest (R),         | 16) feucht (P),     |
| 2) festfressen (C),  | 17) Feuer (R),      |
| 3) festhalten (C),   | 18) feuern (C),     |
| 4) festigen (C),     | 19) Feuilleton (R), |
| 5) Festigkeit (R),   | 20) feurig (P),     |
| 6) festlegen (C),    | 21) Fiasko (R),     |
| 7) festlich (P),     | 22) Fichte (R),     |
| 8) festmachen (C),   | 23) Fiber (R),      |
| 9) festnehmen (C),   | 24) Fieber (R),     |
| 10) festsetzen (C),  | 25) fieberhaft (P), |
| 11) feststehend (P), | 26) Figur (R),      |
| 12) feststellen (C), | 27) fiktiv (P),     |
| 13) fett (P),        | 28) Filiale (R),    |
| 14) Fett (R),        | 29) Film (R),       |
| 15) Fetzen (R),      | 30) Filter (R),     |

- |                     |                      |
|---------------------|----------------------|
| 31) Filz (R),       | 66) Flaum (R),       |
| 32) Finale (R),     | 67) Flechte (R),     |
| 33) Finanzen (R),   | 68) flechten (C),    |
| 34) finanziell (P), | 69) Fleck (R),       |
| 35) finden (C),     | 70) Flecke (R),      |
| 36) Finder (R),     | 71) Flecken (R),     |
| 37) findig (P),     | 72) Fledermaus (R),  |
| 38) Finger (R),     | 73) flegelhaft (P),  |
| 39) finster (P),    | 74) flehen (C),      |
| 40) Finsternis (R), | 75) Fleisch (R),     |
| 41) firm (P),       | 76) flei-schig (P),  |
| 42) Firma (R),      | 77) Fleiß (R),       |
| 43) Firnis (R),     | 78) fleißig (P),     |
| 44) First (R),      | 79) flexibel (P),    |
| 45) Fisch (R),      | 80) Flexowriter (R), |
| 46) fischen (C),    | 81) flicken (C),     |
| 47) Fiskus (R),     | 82) Flieder (R),     |
| 48) Fistel (R),     | 83) Fliege (R),      |
| 49) fit (P),        | 84) fliegen (C),     |
| 50) fix (P),        | 85) Flieger (R),     |
| 51) fixieren (C),   | 86) fliehen (C),     |
| 52) flach (P),      | 87) Fliehkraft (R),  |
| 53) Flachfeile (R), | 88) Fliese (R),      |
| 54) Fläche (R),     | 89) Fließband (R),   |
| 55) Flachs (R),     | 90) fließen (C),     |
| 56) flackern (C),   | 91) flimmern (C),    |
| 57) Flader (R),     | 92) flink (P),       |
| 58) Flagge (R),     | 93) Flint (R),       |
| 59) Flak (R),       | 94) Flitter (R),     |
| 60) Flamme (R),     | 95) Flocke (R),      |
| 61) Flanke (R),     | 96) Floh (R),        |
| 62) Flansch (R),    | 97) Floß (R),        |
| 63) Flasche (R),    | 98) Flosse (R),      |
| 64) flattern (C),   | 99) flößen (C),      |
| 65) flau (P),       | 100) Flöte (R)       |

W powyższym zestawieniu zastosowano następujące oznaczenia części mowy:

- R – rzeczownik,
- C – czasownik,
- P – przymiotnik,
- X – pozostałe części mowy języka niemieckiego nie podlegające fleksji.

Zestawienie to zawiera 100 wyrazów, wśród których można wyróżnić:

- 58 – rzeczowników,
- 22 – czasowników,
- 20 – przymiotników.

W związku z powyższym, współczynnikiem wagowym przypisane zostały następujące wartości:

- $w_D(R) = 0,58$ ,
- $w_D(C) = 0,22$ ,
- $w_D(P) = 0,20$ .

Zatem wartość współczynnika fleksyjności języka niemieckiego może zostać oszacowana jako  $f_D = 5,72$ .

W tabeli 14 zamieszczono zestawienie dostarczające informacji o tym, ile podczas tłumaczenia pewnego tekstu (metodą PBMT) wystąpiło wzorców translacyjnych o długości jednego wyrazu ( $l_1$ ), dwóch wyrazów ( $l_2$ ), trzech wyrazów ( $l_3$ ), czterech wyrazów ( $l_4$ ), pięciu wyrazów ( $l_5$ ) oraz dłuższych niż pięć wyrazów ( $l_{\geq 6}$ ). Podczas przeprowadzonych wyliczeń, jako oddzielnych wyrazów nie liczono niemieckich rodzajników określonych „der”, „die” i „das”, rodzajników nieokreślonych „ein”, „eine” oraz ich wszelkich form fleksyjnych.

Tabela 14. Statystyka wzorców translacyjnych

$l_1$	$l_2$	$l_3$	$l_4$	$l_5$	$l_{\geq 6}$
69	30	12	10	5	3

Minimalna liczba wzorców translacyjnych  $L_D$ , jakie powinna zawierać baza danych, potrzebnych do sprawnego funkcjonowania systemu automatycznej translacji z języka niemieckiego na polski może zostać oszacowana za pomocą następującego wzoru

$$L_D = S_D \cdot f_D \cdot \left( 1 + \frac{l_2 + l_3 + l_4 + l_5 + l_{\geq 6}}{l_1} \right) \quad (4)$$

We wzorze (4) parametr  $S_D$  oznacza licznosc zbioru haseł słownika języka niemieckiego. Wartość parametru  $S_D$  szacowana może być w granicach od 10 000 (w przypadku uwzględnienia jedynie podstawowego słownictwa języka niemieckiego) do 100 000 (w przypadku dodatkowego uwzględnienia słownictwa specjalistycznego). Na przykład słownik [5] zawiera około 60 000 haseł ze zbioru słownictwa języka niemieckiego. Zatem przyjmując  $S_D = 60 000$  otrzymuje się oszacowanie minimalnej liczby wzorców translacyjnych  $L_D = 641 600$ .

#### 4. SZACOWANIE ROZMIARÓW BAZY DANYCH DLA JĘZYKA NIDERLANDZKIEGO

Język niderlandzki należy do zachodniej gałęzi germańskiej grupy językowej. Jego cechą charakterystyczną jest stosunkowo niewielka liczba form fleksyjnych wyrazów. W języku niderlandzkim fleksji podlegają takie części mowy, jak: rzeczowniki, czasowniki i przymiotniki.

Niderlandzki rzeczownik oprócz swej formy podstawowej posiada również odrębną formę liczby mnogiej, którą tworzy się w sposób regularny przez dodanie końcówki „-en” lub „-s”, bądź w sposób nieregularny. W tabeli 15 dokonano zestawienia form liczby mnogiej wybranych rzeczowników niderlandzkich.

Tabela 15. Zestawienie form liczby mnogiej wybranych rzeczowników języka niderlandzkiego

Forma podstawowa rzeczownika	Forma liczby mnogiej	Polski odpowiednik znaczeniowy formy podstawowej
buurt	buurten	dzielnica
paard	paarden	koń
moeite	moeiten	trud
zee	zeeën	morze
hamer	hamers	młot
molen	molens	młyn
radio	radio's	radio
kind	kinderen	dziecko
been	beenderen	noga
stad	steden	miasto
glas	glazen	szklanka
waarheid	waarheden	prawda
timmerman	timmerlieden	cieśla
koe	koeien	krowa

Ponadto w języku niderlandzkim od formy podstawowej rzeczownika można utworzyć jeszcze formę zdrobniąłą, od której z kolei można utworzyć formę liczby mnogiej. W tabeli 16 dokonano zestawienia form zdrobniąłych liczby pojedynczej i mnogiej dla wybranych rzeczowników języka niderlandzkiego.

Tabela 16. Zestawienie form zdrobniąłych liczby pojedynczej i mnogiej wybranych rzeczowników języka niderlandzkiego

Forma podstawowa rzeczownika	Forma zdrobniąłą w liczbie pojedynczej	Forma zdrobniąłą w liczbie mnogiej	Polski odpowiednik znaczeniowy formy podstawowej
bloem	bloemetje	bloemetjes	kwiatek
pop	poppetje	poppetjes	lalka
brug	bruggetje	bruggetjes	most
gat	gaatje	gaatjes	dziura
glas	glaasje	glaasjes	szklanka
woning	woninkje	woninkjes	mieszkanie
beslissing	beslissinkje	beslissinkjes	decyzja
film	filmpje	filmpjes	film
boek	boekje	boekjes	książka
schilderij	schilderijtje	schilderijtjes	obraz
kamer	kamertje	kamertjes	pokój

Zatem niderlandzki rzeczownik posiada oprócz swej formy podstawowej jeszcze formę liczby mnogiej oraz formy zdrobniale dla liczby pojedynczej i mnogiej, wobec tego współczynnik fleksyjności niderlandzkiego rzeczownika wynosi  $f_{NL}(P) = 4$ .

Z kolei niderlandzki czasownik oprócz formy bezokolicznika posiada formy fleksyjne występujące podczas odmiany przez osoby w czasie teraźniejszym. Na przykład czasownik „maaken” (robić) posiada wzór odmiany w czasie teraźniejszym, który zamieszczono w tabeli 17.

**Tabela 17.** Wzór odmiany niderlandzkiego czasownika „maaken” w czasie teraźniejszym

Liczba pojedyncza	Liczba mnoga
ik maak	wij maaken
jjj maakt	jullie maaken
hij maakt	zij maaken

Niderlandzki czasownik podlega również odmianie w czasie przeszłym niedokonanym. Wzorzec odmiany zamieszczono w tabeli 18.

**Tabela 18.** Wzór odmiany niderlandzkiego czasownika „maaken” w czasie przeszłym niedokonanym

Liczba pojedyncza	Liczba mnoga
ik maakte	wij maakten
jjj maakte	jullie maakten
hij maakte	zij maakten

Ponadto niderlandzkie czasowniki posiadają również formy imiesłowu czasu przeszłego dokonanego. Przykłady takich form dla wybranych czasowników zamieszczono w tabeli 19.

**Tabela 19.** Zestawienie form imiesłowów czasu przeszłego dokonanego dla wybranych czasowników języka niderlandzkiego

Forma bezokolicznika	Imiesłów czasu przeszłego dokonanego	Polski ekwiwalent znaczeniowy
hebben	gehad	mieć
zijn	gewest	być
komen	gekomen	przyjść
zien	gezien	widzieć
eten	gegeten	jeść
doen	gedaan	robić
nemen	genomen	nazywać
slapen	geslapen	spać
drinken	gedronken	pić
lezen	gelezen	czytać
spreken	gesproken	mówić

Analizując tabele 17–19 widać, że czasownik niderlandzki oprócz formy podstawowej bezokolicznika posiada jeszcze pięć innych form fleksyjnych. Zatem współczyn-

nik fleksyjności dla niderlandzkiego czasownika wynosi  $f_{NL}(P) = 6$ .

Z kolei niderlandzki przymiotnik oprócz formy podstawowej posiada jeszcze formę z końcówką „-e”, która stosowana jest przed rzeczownikiem, z wyjątkiem sytuacji, gdy jest to rzeczownik rodzaju nijakiego występujący bez rodzajnika lub z rodzajnikiem nieokreślonym. Ponadto niderlandzki przymiotnik posiada odrębne formy fleksyjne odpowiadające stopniowi wyższemu i najwyższemu. Zestawienia form fleksyjnych dla wybranych przymiotników języka niderlandzkiego dokonano w tabeli 20.

**Tabela 20.** Zestawienie form fleksyjnych wybranych przymiotników języka niderlandzkiego

Forma podstawowa	Forma stopnia wyższego	Forma stopnia najwyższego	Polski odpowiednik formy podstawowej
klein	kleiner	kleinst	mały
groot	groter	grotst	wielki
jong	jonger	jongst	młody
oud	ouder	oudst	stary
belangrijk	belangrijker	belangrijkst	znany
wijs	wijzer	wijst	mądry
duur	duurder	duurst	drogi
goed	beter	best	dobry
veel	meer	meest	duży
weinig	minder	minst	mały
graag	liever	liefst	chętny

Z powyższych rozważań wynika, że współczynnik fleksyjności dla niderlandzkiego przymiotnika wynosi  $f_{NL}(P) = 4$ .

Wartość współczynnika fleksyjności języka niderlandzkiego może zostać wyznaczona ze wzoru

$$f_{NL} = w_{NL}(R) \cdot f_{NL}(R) + w_{NL}(C) \cdot f_{NL}(C) + w_{NL}(P) \cdot f_{NL}(P) \quad (5)$$

Występujące we wzorze (5) współczynniki wagowe  $w_{NL}(R)$ ,  $w_{NL}(C)$  oraz  $w_{NL}(P)$  określają częstość występowania rzeczowników, czasowników i przymiotników w słownictwie języka niderlandzkiego.

Częstotliwość występowania rzeczowników, czasowników i przymiotników w słownictwie języka niderlandzkiego została wyznaczona na podstawie analizy następującego zbioru zawierającego sto haseł pobranych kolejno ze słownika [6].

- 1) toeter (R),  
 2) toeteren (C),  
 3) toetreden (C),  
 4) toets (R),  
 5) toetsen (C),  
 6) toetsenbord (R),  
 7) toeval (R),  
 8) toevallig (P),  
 9) toeven (C),  
 10) toeverlaat (R),  
 11) toilet (R),  
 12) toilettafel (R),  
 13) tol (R),  
 14) tolerant (P),  
 15) tolerantie (R),  
 16) tolereren (C),  
 17) tolk (R),  
 18) tolunie (R),  
 19) tomaat (R),  
 20) tombe (R),  
 21) tomeloos (P),  
 22) tomen (C),  
 23) ton (R),  
 24) tondeuse (R),  
 25) toneel (R),  
 26) tonen (C),  
 27) tong (R),  
 28) tonijn (R),  
 29) tonnage (R),  
 30) tooi (R),  
 31) tooien (C),  
 32) toom (R),  
 33) toon (R),  
 34) toonaangevend (R),  
 35) toorn (R),  
 36) toornen (C),  
 37) toornig (P),  
 38) toorts (R),  
 39) toast (R),  
 40) toosten (C),  
 41) top (R),  
 42) topografie (R),  
 43) topprestatie (R),  
 44) tor (R),  
 45) toren (R),  
 46) tornen (C),  
 47) torpederen (C),  
 48) torpedojager (R),  
 49) torsen (C),  
 50) toss (R),  
 51) tot (X),  
 52) totaal (P),  
 53) totalitair (P),  
 54) totdat (X),  
 55) totstandbrenging (R),  
 56) touperen (C),  
 57) touringcar (R),  
 58) tournee (R),  
 59) touw (R),  
 60) touw ladder (R),  
 61) tovenaar (R),  
 62) toverachtig (P),  
 63) toveren (C),  
 64) toverkol (R),  
 65) traag (P),  
 66) traan (R),  
 67) tracé (R),  
 68) trachten (C),  
 69) tractor (R),  
 70) traditie (R),  
 71) traditioneel (P),  
 72) tragedie (R),  
 73) tragisch (P),  
 74) trainen (C),  
 75) trainingspak (R),  
 76) traject (R),  
 77) traktaat (R),  
 78) traktement (R),  
 79) trakteren (C),  
 80) tralie (R),  
 81) tram (R),  
 82) trammen (C),  
 83) tranen (C),  
 84) tranig (C),  
 85) tranquillizer (R),  
 86) transactie (R),  
 87) transistor (R),  
 88) transitief (P),  
 89) transito (R),  
 90) transplantatie (R),  
 91) transport (R),  
 92) transporteren (C),  
 93) trant (R),  
 94) trap (R),  
 95) trapezium (R),  
 96) trapeuning (R),  
 97) trappelen (C),  
 98) trappen (C),  
 99) trapper (R),  
 100) trauma (R)

W powyższym zestawieniu zastosowano następujące oznaczenia części mowy:

- R – rzeczownik,  
 C – czasownik,

P – przymiotnik,

X – pozostałe części mowy języka niderlandzkiego nie podlegające fleksji.

Zestawienie to zawiera 100 wyrazów, wśród których można wyróżnić:

- 63 – rzeczowniki,  
 24 – czasowniki,  
 11 – przymiotników.

Stąd współczynnikiem wagowym przypisane zostały następujące wartości:

$$w_{NL}(R) = 0,63,$$

$$w_{NL}(C) = 0,24,$$

$$w_{NL}(P) = 0,11.$$

Zatem wartość współczynnika fleksyjności języka niderlandzkiego może zostać oszacowana jako  $f_{NL} = 4,4$ .

W tabeli 21 zamieszczono zestawienie dostarczające informacji o tym, ile podczas tłumaczenia (metodą PBMT) pewnego próbnego tekstu wystąpiło wzorców translacyjnych o długości jednego wyrazu ( $l_1$ ), dwóch wyrazów ( $l_2$ ), trzech wyrazów ( $l_3$ ), czterech wyrazów ( $l_4$ ), pięciu wyrazów ( $l_5$ ) oraz dłuższych niż pięć wyrazów ( $l_{\geq 6}$ ). Podczas przeprowadzonych wyliczeń, jako oddzielnych wyrazów nie liczono niderlandzkich rodzajników określonych „de” i „het” oraz rodzajnika nieokreślonego „een”.

**Tabela 21.** Statystyka wzorców translacyjnych

$l_1$	$l_2$	$l_3$	$l_4$	$l_5$	$l_{\geq 6}$
48	38	20	7	3	1

Minimalna liczba wzorców translacyjnych  $L_{NL}$ , jakie powinna zawierać baza danych, potrzebnych do sprawnego funkcjonowania systemu automatycznej translacji z języka niderlandzkiego na polski może zostać oszacowana za pomocą następującego wzoru

$$L_{NL} = S_{NL} \cdot f_{NL} \cdot \left( 1 + \frac{l_2 + l_3 + l_4 + l_5 + l_{\geq 6}}{l_1} \right) \quad (6)$$

We wzorze (6) parametr oznacza liczebność zbioru haseł słownika języka niderlandzkiego. Wartość parametru szacowana może być w granicach od 10 000 (w przypadku uwzględnienia jedynie podstawowego słownictwa języka niderlandzkiego) do 100 000 (w przypadku dodatkowego uwzględnienia słownictwa specjalistycznego). Na przykład słownik [6] zawiera około 24 000 haseł ze zbioru słownictwa języka niderlandzkiego. Zatem przyjmując  $S_{NL} = 24 000$ , otrzymuje się oszacowanie minimalnej liczby przykładów translacyjnych  $L_{NL} = 257 400$ .



## 5. SZACOWANIE ROZMIARÓW BAZY DANYCH DLA JĘZYKA SZWEDZKIEGO

Język szwedzki jest reprezentantem północnej gałęzi germańskiej grupy językowej. Jego cechą charakterystyczną jest stosunkowo umiarkowana liczba form fleksyjnych wyrazów. Częściami mowy podlegającymi w języku szwedzkim fleksji są rzeczowniki, czasowniki i przymiotniki [7].

Szwedzki rzeczownik oprócz swej formy podstawowej posiada jeszcze formę charakterystyczną dla liczby mnogiej oraz formy określone dla liczby pojedynczej i mnogiej. Zestawienia form fleksyjnych szwedzkiego rzeczownika dokonano w tabeli 22.

Zatem współczynnik fleksyjności szwedzkiego rzeczownika wynosi  $f_S(R) = 4$ .

**Tabela 22.** Zestawienie form fleksyjnych wybranych szwedzkich rzeczowników

Forma podstawowa	Forma określona dla liczby pojedynczej	Forma nieokreślona dla liczby mnogiej	Forma określona dla liczby mnogiej	Znaczenie formy podstawowej
huggorm	huggormen	huggormar	huggormarna	żmija
hög	högen	högar	högarna	stos
väska	väsken	väschor	väschorna	walizka
flicka	flickan	flickor	flickorna	dziewczyna
blomma	blomman	blommor	blommorna	kwiat
bil	bilen	bilar	bilarna	samochód
pojke	pojken	pojkar	pojkarne	chłopiec
tomat	tomaten	tomater	tomaterna	pomidor
apelsin	apelsinen	apelsiner	apelsinerna	pomarańcza
äpple	äpplet	äpplen	äpplena	jabłko
begravning	begravningen	begravningar	begravningarna	pogrzeb

Z kolei szwedzki czasownik oprócz formy bezokolicznika posiada również odrębne formy fleksyjne dla czasu teraźniejszego, czasu przeszłego, czasu przeszło-teraźniejszego oraz formę imiesłowu czasu przeszłego. Zestawienie form fleksyjnych wybranych szwedzkich czasowników zostało zamieszczone w tabeli 23.

Ponadto większość czasowników szwedzkich tworzy dodatkowe formy fleksyjne strony biernej dla bezokolicznika, czasu teraźniejszego, czasu przeszłego oraz czasu teraźniejszo-przeszłego. Zestawienia form fleksyjnych wybranych szwedzkich czasowników w stronie biernej dokonano w tabeli 24.

**Tabela 23.** Zestawienie form fleksyjnych dla wybranych czasowników języka szwedzkiego

Forma bezokolicznika	Forma czasu teraźniejszego	Forma czasu przeszłego	Forma czasu przeszło-teraźniejszego	Imiesłów czasu przeszłego	Znaczenie formy podstawowej
måla	målar	målade	målats	målads	malować
beställa	beställer	beställde	beställt	beställd	zamawiać
sy	sy	sydde	sytt	sydd	szyć
göra	gör	gjorde	gjort	gjord	robić
skriva	skriver	skrev	skrivit	skrivnen	pisać
dricka	dricker	drack	druckit	drucken	pić
giva	ger	gav	givit	given	dać
vinna	vinner	vann	vunnit	vunnen	wygrać
äta	äter	ät	ätit	äten	jeść
ljuga	ljuger	ljög	ljugit	ljugen	kłamać
hinna	hinner	hann	hunnit	hunnen	zdążyć

**Tabela 24.** Zestawienie form fleksyjnych strony biernej dla wybranych szwedzkich czasowników

Forma bezokolicznika	Forma czasu teraźniejszego	Forma czasu przeszłego	Forma czasu teraźniejszo-przeszłego	Polski odpowiednik znaczeniowy
tvättas	tvättas	tvättades	tvättats	myć
köpas	köpes	köptes	köpts	kupować
sys	sys	syddes	sytt	szyć
läggas	lägges	lades	lagts	kłaść
ätas	ätes	äts	ätits	jeść
beställas	beställes	beställdes	beställts	zamówić
skrivs	skrivs	skrevs	skrivits	pisać
målas	målas	målades	målats	malować

Jak widać z zestawień zamieszczonych w tabelach 23 i 24, typowy szwedzki czasownik posiada dziewięć form fleksyjnych, w związku z czym jego współczynnik fleksyjności wynosi  $f_S(C) = 9$ .

Z kolei szwedzki przymiotnik posiada odrębne formy fleksyjne dla rodzaju wspólnego i nijakiego oraz formę liczby mnogiej. Zestawienia form fleksyjnych przykładowych szwedzkich przymiotników dokonano w tabeli 25.

**Tabela 25.** Zestawienie form fleksyjnych dla wybranych szwedzkich przymiotników

Forma dla rodzaju wspólnego	Forma dla rodzaju nijakiego	Forma liczby mnogiej	Polski odpowiednik znaczeniowy
stor	stort	stora	duży
dyr	dyrt	dyra	drogi
ny	nytt	nya	nowy
liten	litet	lilla	mały
blå	blått	blåa	niebieski
billig	billigt	billiga	tani
gammal	gammalt	gammla	stary

Ponadto szwedzki przymiotnik podlega stopniowaniu, w związku z czym istnieją dodatkowe formy fleksyjne dla stopnia wyższego i najwyższego przymiotników. Zestawienie form fleksyjnych stopnia wyższego i najwyższego dla wybranych szwedzkich przymiotników zamieszczono w tabeli 26.

W ogólnym przypadku szwedzki przymiotnik posiada oprócz swej formy podstawowej jeszcze cztery dodatkowe formy fleksyjne, zatem współczynnik fleksyjności szwedzkiego przymiotnika wynosi  $f_S(P) = 5$ .

**Tabela 26.** Zestawienie form fleksyjnych stopnia wyższego i najwyższego dla wybranych szwedzkich przymiotników

Forma podstawowa	Forma stopnia wyższego	Forma stopnia najwyższego	Znaczenie formy podstawowej
dyr	dyrare	dyrast	drogi
vacker	vackrare	vackras	ładny
ung	yngrare	yngst	młody
lång	långre	långst	długi
dålig	sämre	sämst	zły
bra	bättre	bäst	dobry
gammal	äldre	äldst	stary
liten	mindre	minst	mały
många	flera	flest	dużo
hög	högre	högst	wysoki
tung	tyngre	tyngst	ciężki

Wartość współczynnika fleksyjności języka szwedzkiego może zostać wyznaczona ze wzoru

$$f_S = w_S(R) \cdot f_S(R) + w_S(C) \cdot f_S(C) + w_S(P) \cdot f_S(P) \quad (7)$$

Występujące we wzorze (7) współczynniki wagowe  $w_S(R)$ ,  $w_S(C)$  oraz  $w_S(P)$  określają częstość występowania rzeczowników, czasowników i przymiotników w słownictwie języka szwedzkiego.

Częstotliwość występowania rzeczowników, czasowników i przymiotników w słownictwie języka szwedzkiego została wyznaczona na podstawie analizy następującego zbioru zawierającego sto haseł pobranych kolejno ze słownika [8].

- 1) rabatt (R),
- 2) racerbil (R),
- 3) rad (R),
- 4) radikal (P),
- 5) radio (R),
- 6) rak (P),
- 7) raka (C),
- 8) rakapparat (R),
- 9) rakblad (R),
- 10) raket (R),
- 11) rally (R),
- 12) ram (R),
- 13) ramlä (C),
- 14) randig (P),
- 15) raps (R),
- 16) rar (P),
- 17) ras (R),
- 18) rast (R),
- 19) ratt (R),
- 20) rea (R),
- 21) reagera (P),
- 22) reaktion (R),
- 23) realisation (R),
- 24) realistisk (P),
- 25) recension (R),
- 26) recept (R),
- 27) reception (R),
- 28) reda (C),
- 29) redaktion (R),
- 30) redaktör (R),
- 31) redan (X),
- 32) rederi (R),
- 33) redogöra (C),
- 34) redovisning (R),
- 35) redskap (R),
- 36) reform (R),
- 37) regel (R),
- 38) regelbunden (P),
- 39) regera (C),
- 40) regering (R),
- 41) regi (R),
- 42) regissör (R),
- 43) registrera (C),
- 44) regn (R),
- 45) regna (C),
- 46) regning (R),
- 47) regnkappa (R),
- 48) reklam (R),
- 49) rekommendera (C),
- 50) rekord (R),
- 51) rekreation (R),
- 52) rektangel (R),
- 53) rektor (R),
- 54) relativ (P),
- 55) religion (R),
- 56) religiös (P),
- 57) ren (R),
- 58) ren (P),
- 59) renovera (C),
- 60) rep (R),
- 61) reparation (R),
- 62) reparera (C),
- 63) repetera (C),
- 64) reportage (R),
- 65) representant (R),
- 66) repris (R),
- 67) republik (R),
- 68) resa (C),
- 69) resa (R),
- 70) resebyrå (R),
- 71) resecheck (R),
- 72) reseledare (R),
- 73) resenär (R),
- 74) reserv (R),
- 75) reservdel (R),
- 76) reservera (C),
- 77) resgodsintämning (R),
- 78) resonera (C),
- 79) respekt (R),
- 80) rest (R),
- 81) restaurang (R),
- 82) restaurangvagn (R),
- 83) resultat (R),
- 84) resurs (R),
- 85) resväska (R),
- 86) reta (C),
- 87) returbiljett (R),
- 88) revolution (R),
- 89) revy (R),
- 90) rida (C),
- 91) rik (P),
- 92) rikedom (R),
- 93) riksdag (R),
- 94) riksdagsman (R),
- 95) rikssamtal (R),
- 96) riktig (P),
- 97) riktigt (X),
- 98) riktning (R),
- 99) riktnummer (R),
- 100) rimlig (P)

W zestawieniu zastosowano następujące oznaczenia części mowy:

- R – rzeczownik,
- C – czasownik,
- P – przymiotnik,
- X – pozostałe części mowy języka szwedzkiego nie podlegające fleksji.

Z zestawienie to zawiera 100 wyrazów, wśród których można wyróżnić:

- 69 – rzeczowników,
- 16 – czasowników,
- 13 – przymiotników.

Stąd współczynnikiem wagowym przypisane zostały następujące wartości:

$$w_S(R) = 0,69,$$

$$w_S(C) = 0,16,$$

$$w_S(P) = 0,13.$$

Zatem wartość współczynnika fleksyjności języka szwedzkiego może zostać oszacowana jako  $f_S = 4,85$ .

W tabeli 27 zamieszczono zestawienie dostarczające informacji o tym, ile wystąpiło podczas tłumaczenia wzorców translacyjnych o długości jednego wyrazu ( $l_1$ ), dwóch wyrazów ( $l_2$ ), trzech wyrazów ( $l_3$ ), czterech wyrazów ( $l_4$ ), pięciu wyrazów ( $l_5$ ) oraz dłuższych niż pięć wyrazów ( $l_{\geq 6}$ ). Podczas przeprowadzonych wyliczeń, jako oddzielnych wyrazów nie liczono szwedzkich rodzajników nieokreślonych „en” i „ett”.

Tabela 27. Statystyka wzorców translacyjnych

$l_1$	$l_2$	$l_3$	$l_4$	$l_5$	$l_{\geq 6}$
59	32	20	10	5	0

Minimalna liczba wzorców translacyjnych  $L_S$ , jakie powinna zawierać baza danych, potrzebnych do sprawnego funkcjonowania systemu automatycznej translacji z języka szwedzkiego na polski, może zostać oszacowana za pomocą następującego wzoru

$$L_S = S_S \cdot f_S \cdot \left( 1 + \frac{l_2 + l_3 + l_4 + l_5 + l_{\geq 6}}{l_1} \right) \quad (8)$$

We wzorze (8) parametr  $S_S$  oznacza liczebność zbioru haseł słownika języka szwedzkiego. Wartość parametru  $S_S$  szacowana może być w granicach od 10 000 (w przypadku uwzględnienia jedynie podstawowego słownictwa języka szwedzkiego) do 100 000 (w przypadku dodatkowego uwzględnienia słownictwa specjalistycznego). Na przykład słownik [9] zawiera około 11 000 haseł ze zbioru słownic-

twą języka szwedzkiego. Zatem przyjmując  $S_S = 11\ 000$  otrzymuje się następujące oszacowanie minimalnej liczby przykładów translacyjnych  $L_S = 113\ 900$ .

## 6. SZACOWANIE ROZMIARÓW BAZY DANYCH DLA JĘZYKA NORWESKIEGO

Język norweski należy do północnej gałęzi germańskiej grupy językowej. Język norweski charakteryzuje się występowaniem umiarkowanej liczby form fleksyjnych wyrazów.

Rzeczownik w języku norweskim oprócz swej formy podstawowej posiada również formy charakterystyczne dla liczby mnogiej oraz formy określone dla liczby pojedynczej i mnogiej. Zestawienie form fleksyjnych dla wybranych rzeczowników języka norweskiego zamieszczono w tabeli 28.

Tabela 28. Zestawienie form fleksyjnych wybranych rzeczowników języka norweskiego

Forma podstawowa	Forma określona dla liczby pojedynczej	Forma nieokreślona dla liczby mnogiej	Forma określona dla liczby mnogiej	Znaczenie formy podstawowej
gutt	gutten	gutter	guttene	chłopiec
hund	hunden	hunder	hundene	pies
stol	stolen	stoler	stolene	krzesło
jente	jenten	jenter	jentene	dziewczyna
katt	katten	katter	kattene	kot
bok	boken	boker	bokene	książka
bil	bilen	biler	bilene	samochód
ingeniør	ingeniøren	ingeniører	ingeniørene	inżynier
barn	barnet	barn	barna	dziecko
dyr	dyret	dyr	dyrene	zwierzę
dør	døren	dører	dørene	drzwi

Jak wynika z powyższego zestawienia, współczynnik fleksyjności norweskiego rzeczownika wynosi  $f_N(R) = 4$ .

W języku norweskim czasownik oprócz formy bezokolicznika posiada dodatkowe formy fleksyjne dla czasu teraźniejszego, czasu przeszłego oraz formę imiesłowu czasu przeszłego. Formy fleksyjne wybranych norweskich czasowników zostały zestawione w tabeli 29.

Zatem współczynnik fleksyjności dla norweskiego czasownika wynosi  $f_N(C) = 4$ .

W języku norweskim przymiotnik występować może w trzech różnych formach fleksyjnych, które zostały dla wybranych przymiotników zestawione w tabeli 30.

**Tabela 29.** Zestawienie form fleksyjnych dla wybranych czasowników języka norweskiego

Forma bezokolicznika	Forma czasu teraźniejszego	Forma czasu przeszłego	Imiesłów czasu przeszłego	Znaczenie formy podstawowej
brekke	brekker	brakk	brukket	złamać
dra	drar	drog	dratt	ciągnąć
finne	finner	fant	funnet	znaleźć
si	sir	sa	sagt	mówić
legge	legger	la	lagt	kłaść
syng	synger	sang	sunget	śpiewać
tore	torer	torde	tort	odważyć się
vite	viter	visste	visst	wiedzieć
sitte	sitter	satt	sittet	siedzieć
gråte	gråter	gråt	grått	plakać
slå	slår	slo	slått	uderzać

**Tabela 30.** Zestawienie form fleksyjnych dla wybranych przymiotników języka norweskiego

Forma dla rodzaju męskiego i żeńskiego	Forma dla rodzaju nijakiego	Forma określona	Znaczenie
rød	rødt	røde	czerwony
gul	gult	gule	żółty
brun	brunt	brune	brązowy
blå	blått	blå	niebieski
grå	grått	grå	szary
hvit	hvitt	hvite	biały
sort	sort	sorte	czarny
pen	pent	pene	ładny
stor	stort	store	duży
ny	nytt	nye	nowy
liten	lite	lille	mały

Ponadto w języku norweskim przymiotnik podlega stopniowaniu. Formy stopnia wyższego najwyższego dla wybranych przymiotników języka norweskiego zostały zebrane w tabeli 31.

Z powyższego wynika, że współczynnik fleksyjności dla norweskiego przymiotnika wynosi  $f_N(P) = 5$ .

Wartość współczynnika fleksyjności języka norweskiego może zostać wyznaczona ze wzoru

$$f_N = w_N(R) \cdot f_N(R) + w_N(C) \cdot f_N(C) + w_N(P) \cdot f_N(P) \quad (9)$$

Występujące we wzorze (9) współczynniki wagowe  $w_N(R)$ ,  $w_N(C)$  oraz  $w_N(P)$  określają częstość występowania rzeczowników, czasowników i przymiotników w słownictwie języka norweskiego.

**Tabela 31.** Formy stopnia wyższego i najwyższego dla wybranych przymiotników języka norweskiego

Przymiotnik w stopniu równym	Stopień wyższy	Stopień najwyższy	Znaczenie
lang	lengre	lengst	długi
stor	større	størst	duży
ung	yngre	yngst	młody
tung	tyngre	tyngst	ciężki
god	bedre	best	dobry
vond	verre	verst	zły
gammel	eldre	eldst	stary
mye	mer	mest	dużo
ny	nyere	nyest	nowy
varm	varmere	varmest	ciepły
kald	kaldere	kaldest	zimny

Częstotliwość występowania rzeczowników, czasowników i przymiotników w słownictwie języka norweskiego została wyznaczona na podstawie analizy następującego zbioru zawierającego sto haseł pobranych kolejno ze słownika [10].

- |                  |                      |
|------------------|----------------------|
| 1) madrass (R),  | 24) mangle (C),      |
| 2) magasin (R),  | 25) mann (R),        |
| 3) mage (R),     | 26) manndom (R),     |
| 4) mager (P),    | 27) mannfolk (R),    |
| 5) magnet (R),   | 28) mannskap (R),    |
| 6) mai (R),      | 29) manntall (R),    |
| 7) mais (R),     | 30) manuell (P),     |
| 8) majones (R),  | 31) manufaktur (R),  |
| 9) make (R),     | 32) mappe (R),       |
| 10) make (P),    | 33) marg (R),        |
| 11) makrell (R), | 34) marine (R),      |
| 12) makt (R),    | 35) mark (R),        |
| 13) male (C),    | 36) markant (P),     |
| 14) maler (R),   | 37) markblomst (R),  |
| 15) maleri (R),  | 38) marked (P),      |
| 16) malm (R),    | 39) markere (C),     |
| 17) mamma (R),   | 40) markjordbar (R), |
| 18) man (X),     | 41) marknad (R),     |
| 19) mandag (R),  | 42) marmelade (R),   |
| 20) mandel (R),  | 43) mars (R),        |
| 21) mandig (P),  | 44) marsj (R),       |
| 22) mange (X),   | 45) marsjal (R),     |
| 23) mangel (R),  | 46) mas (R),         |

47) maskin (R),	74) medgang (R),
48) maskinist (R),	75) medhjelper (R),
49) maskinskrive (C),	76) medikament (R),
50) maskulin (P),	77) medlem (R),
51) massasje (R),	78) medlidenhet (R),
52) masse (R),	79) medmenneske (R),
53) mast (R),	80) medvirke (R),
54) mat (R),	81) meg (X),
55) mate (C),	82) meget (X),
56) matematikk (R),	83) meie (C),
57) materiale (R),	84) meieri (R),
58) materialisme (R),	85) meieriprodukter (R),
59) materialistisk (P),	86) meierivarer (R),
60) materie (R),	87) meisel (R),
61) materiell (R),	88) mekaniker (R),
62) materiell (P),	89) mekanikk (R),
63) matlagning (R),	90) mekanisk (P),
64) matt (R),	91) mekanisme (R),
65) matvarer (R),	92) mekle (C),
66) med (X),	93) mekler (R),
67) medalje (R),	94) mektig (P),
68) medarbeider (R),	95) mel (R),
69) meddele (C),	96) melde (C),
70) meddelelse (R),	97) melding (C),
71) medeier (R),	98) melk (R),
72) medfølelse (R),	99) melke (C),
73) medfore (C),	100) mellom (X)

W powyższym zestawieniu zastosowano następujące oznaczenia części mowy:

- R – rzeczownik,
- C – czasownik,
- P – przymiotnik,
- X – pozostałe części mowy języka norweskiego nie podlegające fleksji.

Powyższe zestawienie zawiera 100 wyrazów, wśród których można wyróżnić:

- 71 – rzeczowników,
- 12 – czasowników,
- 11 – przymiotników.

Stąd współczynnikom wagowym przypisane zostały następujące wartości:

$$w_N(R) = 0,71,$$

$$w_N(C) = 0,12,$$

$$w_N(P) = 0,11.$$

Zatem wartość współczynnika fleksyjności języka norweskiego może zostać oszacowana jako  $f_N = 3,87$ .

W tabeli 32 zamieszczono zestawienie dostarczające informacji o tym, ile podczas tłumaczenia wystąpiło wzorców translacyjnych o długości jednego wyrazu ( $l_1$ ), dwóch wyrazów ( $l_2$ ), trzech wyrazów ( $l_3$ ), czterech wyrazów ( $l_4$ ), pięciu wyrazów ( $l_5$ ) oraz dłuższych niż pięć wyrazów ( $l_{\geq 6}$ ). Podczas przeprowadzonych wyliczeń, jako oddzielnych wyrazów nie liczono norweskich rodzajników nieokreślonych „en”, „ei” i „et”.

**Tabela 32.** Statystyka wzorców translacyjnych

$l_1$	$l_2$	$l_3$	$l_4$	$l_5$	$l_{\geq 6}$
64	61	19	9	2	1

Minimalna liczba wzorców translacyjnych  $L_N$ , jakie powinna zawierać baza danych, potrzebnych do sprawnego funkcjonowania systemu automatycznej translacji z języka norweskiego na polski, może zostać oszacowana za pomocą następującego wzoru

$$L_N = S_N \cdot f_N \cdot \left( 1 + \frac{l_2 + l_3 + l_4 + l_5 + l_{\geq 6}}{l_1} \right) \quad (10)$$

Występujący we wzorze (10) parametr  $S_N$  oznacza liczebność zbioru haseł słownika języka norweskiego. Wartość parametru  $S_N$  szacowana może być w granicach od 10 000 (w przypadku uwzględnienia jedynie podstawowego słownictwa języka norweskiego) do 100 000 (w przypadku dodatkowego uwzględnienia słownictwa specjalistycznego). Na przykład słownik [10] zawiera około 7000 haseł ze zbioru słownictwa języka norweskiego. Zatem przyjmując  $S_N = 7000$  otrzymuje się oszacowanie minimalnej liczby potrzebnych przykładów translacyjnych  $L_N = 38\,900$ .

## 7. SZACOWANIE ROZMIARÓW BAZY DANYCH DLA JĘZYKA FRANCUSKIEGO

Język francuski należy do romańskiej grupy językowej. Jego cechą charakterystyczną jest zanik form deklinacyjnych rzeczowników przy jednocześnie bardzo rozbudowanej fleksji czasownikowej [11].

Częściami mowy, dla których można utworzyć formy fleksyjne, są w języku francuskim rzeczowniki, czasowniki i przymiotniki.

Francuski rzeczownik oprócz swej formy podstawowej posiada jeszcze odrębną formę fleksyjną dla liczby mnogiej. Celem zilustrowania fleksji francuskiego rzeczownika, w tabeli 33 zamieszczono pewną liczbę wybranych rzeczowników francuskich wraz z ich formami liczby mnogiej.

**Tabela 33.** Zestawienie form fleksyjnych wybranych rzeczowników języka francuskiego

Forma podstawowa	Forma liczby mnogiej	Znaczenie
homme	hommes	mężczyzna
femme	femmes	kobieta
oiseau	oiseaux	ptak
tuyau	tuyaux	rurka
cheveu	cheveux	włos
bijou	bijoux	klejnot
cheval	chevaux	koń
travail	travaux	praca
chien	chiens	pies
chat	chats	kot
lion	lions	lew

Jak wynika z powyższego zestawienia, współczynnik fleksyjności francuskiego rzeczownika wynosi  $f_R(R) = 2$ .

W języku francuskim czasowniki posiadają wiele form fleksyjnych, ponieważ podlegają odmianie przez osoby w kilku czasach i trybach. Fleksja francuskiego czasownika została zilustrowana na przykładzie czasownika „avoir”, któremu w języku polskim odpowiada czasownik „mieć”.

Odmiana francuskiego czasownika „avoir” w czasie teraźniejszym (présent) została zamieszczona w tabeli 34.

**Tabela 34.** Formy koniugacyjne czasownika „avoir” w czasie teraźniejszym (présent)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	ai	avons
Druga	as	avez
Trzecia	a	ont

Z kolei w tabeli 35 zamieszczono wzór odmiany czasownika „avoir” w czasie przeszłym (imparfait).

**Tabela 35.** Formy koniugacyjne czasownika „avoir” w czasie przeszłym (imparfait)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	avais	avions
Druga	avais	aviez
Trzecia	avait	avaient

W tabeli 36 zamieszczono formy odmiany czasownika „avoir” w czasie przeszłym (passé simple).

**Tabela 36.** Formy koniugacyjne czasownika „avoir” w czasie przeszłym (passé simple)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	eus	eûmes
Druga	eus	eûtes
Trzecia	eut	eurent

Z kolei w tabeli 37 zamieszczono formy koniugacyjne czasownika „avoir” w czasie przyszłym (futur).

**Tabela 37.** Formy koniugacyjne czasownika „avoir” w czasie przyszłym (futur)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	aurai	aurons
Druga	auras	avez
Trzecia	aura	auront

Natomiast w tabeli 38 zamieszczono wzór odmiany czasownika „avoir” w trybie warunkowym (conditionnel présent).

**Tabela 38.** Formy koniugacyjne czasownika „avoir” w trybie warunkowym (conditionnel présent)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	aurais	aurions
Druga	aurais	auriez
Trzecia	aurait	auraient

Z kolei w tabeli 39 zamieszczono wykaz form fleksyjnych występujących podczas odmiany czasownika „avoir” w trybie łączącym (subjonctif) w czasie teraźniejszym (présent).

**Tabela 39.** Formy koniugacyjne czasownika „avoir” w trybie łączącym (subjonctif) w czasie teraźniejszym (présent)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	aie	ayons
Druga	aies	ayez
Trzecia	ait	aient

W tabeli 40 zamieszczono wzór odmiany czasownika „avoir” w trybie łączącym (subjonctif) w czasie przeszłym (imparfait).

Ponadto czasownik „avoir” posiada dodatkowe formy fleksyjne dla imiesłówów: „ayant” i „eu” oraz dla trybu rozkazującego „aie”, „ayons” i „ayez”.

**Tabela 40.** Formy koniugacyjne czasownika „avoir” w trybie łączącym (subjonctif) w czasie przeszłym (imparfait)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	eusse	eussions
Druga	eusse	eussiez
Trzecia	eût	eussent

Zatem współczynnik fleksyjności dla francuskiego czasownika wynosi  $f_F(C) = 36$ .

W języku francuskim przymiotnik oprócz formy podstawowej posiada dodatkowe formy fleksyjne dla liczby mnogiej oraz dla rodzaju żeńskiego liczby pojedynczej i mnogiej. Zestawienie form fleksyjnych wybranych przymiotników języka francuskiego zamieszczono w tabeli 41.

**Tabela 41.** Zestawienie form fleksyjnych dla wybranych przymiotników francuskich

Forma podstawowa	Forma liczby mnogiej	Forma rodzaju żeńskiego – liczba pojedyncza	Forma rodzaju żeńskiego – liczba mnoga	Znaczenie
léger	légers	légere	légeres	lekki
ancien	anciens	ancienne	anciennes	starożytny
bon	bons	bonne	bonnes	dobry
cruel	cruels	cruelle	cruelle	okrutny
beau	beaux	belle	belles	piękny
nouveau	nouveaux	nouvelle	nouvelle	nowy
blanc	blancs	blanche	blanches	biały
grand	grands	grande	grandes	duży
petit	petits	petite	petites	mały
clair	clairs	claire	claires	jasny
joli	jolis	jolie	jolies	ładny

Zatem współczynnik fleksyjności francuskiego przymiotnika wynosi  $f_F(P) = 4$ .

Wartość współczynnika fleksyjności języka francuskiego może zostać wyznaczona ze wzoru

$$f_F = w_F(R) \cdot f_F(R) + w_F(C) \cdot f_F(C) + w_F(P) \cdot f_F(P) \quad (11)$$

Występujące we wzorze (11) współczynniki wagowe  $w_F(R)$ ,  $w_F(C)$  oraz  $w_F(P)$  określają częstość występowania rzeczowników, czasowników i przymiotników w słownictwie języka francuskiego.

Częstotliwość występowania rzeczowników, czasowników i przymiotników w słownictwie języka francuskiego została wyznaczona na podstawie analizy następującego zbioru zawierającego sto haseł pobranych kolejno ze słownika [12].

- |                          |                       |
|--------------------------|-----------------------|
| 1) calendrier (R),       | 51) carafe (R),       |
| 2) calme (P),            | 52) caravane (R),     |
| 3) calmer (C),           | 53) carburant (R),    |
| 4) calonnier (C),        | 54) carburateur (R),  |
| 5) camarade (R),         | 55) cardiaque (P),    |
| 6) cambrioler (C),       | 56) cardinal (P),     |
| 7) cambrioleur (R),      | 57) careme (R),       |
| 8) caméra (R),           | 58) caresse (R),      |
| 9) camion (R),           | 59) caresser (C),     |
| 10) camionnette (R),     | 60) cargaison (R),    |
| 11) camomille (R),       | 61) cargo (R),        |
| 12) camp (R),            | 62) caricature (R),   |
| 13) campagne (R),        | 63) carnassier (P),   |
| 14) camper (C),          | 64) carnet (R),       |
| 15) campeur (R),         | 65) carotte (R),      |
| 16) Canada (R),          | 66) carpe (R),        |
| 17) canadien (R),        | 67) carré (P),        |
| 18) canal (R),           | 68) careau (R),       |
| 19) canalisation (R),    | 69) carrefour (R),    |
| 20) canapé (R),          | 70) carrément (P),    |
| 21) canard (R),          | 71) carriere (R),     |
| 22) cancer (R),          | 72) carriere (R),     |
| 23) candidat (R),        | 73) carrosserie (R),  |
| 24) candidature (R),     | 74) cartable (R),     |
| 25) canif (R),           | 75) carte (R),        |
| 26) canne (R),           | 76) carton (R),       |
| 27) cannelle (R),        | 77) cartouche (R),    |
| 28) canon (R),           | 78) cas (R),          |
| 29) canot (R),           | 79) casanier (R),     |
| 30) cantatrice (R),      | 80) case (R),         |
| 31) cantine (R),         | 81) caserne (R),      |
| 32) canton (R),          | 82) casier (R),       |
| 33) caoutchouc (R),      | 83) casque (R),       |
| 34) cap (R),             | 84) casquette (R),    |
| 35) capable (P),         | 85) casser (C),       |
| 36) capacité (R),        | 86) casserole (R),    |
| 37) capitaine (R),       | 87) cassis (R),       |
| 38) capital (P),         | 88) catalogue (R),    |
| 39) capitalisme (R),     | 89) catastrophe (R),  |
| 40) capitaliste (P),     | 90) catéchisme (R),   |
| 41) capitulation (R),    | 91) catégorie (R),    |
| 42) capituler (C),       | 92) cathédrale (R),   |
| 43) caporal (R),         | 93) catholicisme (R), |
| 44) caprice (R),         | 94) catholique (P),   |
| 45) captivité (R),       | 95) cauchemar (R),    |
| 46) capturer (C),        | 96) cause (R),        |
| 47) capuchon (R),        | 97) causer (C),       |
| 48) car (X),             | 98) cavalerie (R),    |
| 49) caractere (R),       | 99) cavalier (P),     |
| 50) caractéristique (P), | 100) cave (R)         |

W zestawieniu zastosowano następujące oznaczenia części mowy:

- R – rzeczownik,
- C – czasownik,
- P – przymiotnik,
- X – pozostałe części mowy języka francuskiego nie podlegające fleksji.

Powyższe zestawienie zawiera 100 wyrazów, wśród których można wyróżnić:

- 78 – rzeczowników,
- 9 – czasowników,
- 12 – przymiotników.

Stąd współczynnikiem wagowym przypisane zostały następujące wartości:

$$w_F(R) = 0,78,$$

$$w_F(C) = 0,09,$$

$$w_F(P) = 0,12.$$

Zatem wartość współczynnika fleksyjności języka francuskiego może zostać oszacowana jako  $f_F = 5,28$ .

W tabeli 42 zamieszczono zestawienie dostarczające informacji o tym, ile podczas tłumaczenia wystąpiło wzorców translacyjnych o długości jednego wyrazu ( $l_1$ ), dwóch wyrazów ( $l_2$ ), trzech wyrazów ( $l_3$ ), czterech wyrazów ( $l_4$ ), pięciu wyrazów ( $l_5$ ) oraz dłuższych niż pięć wyrazów ( $l_{\geq 6}$ ). Podczas przeprowadzonych wyliczeń, jako oddzielnych wyrazów nie liczono francuskich rodzajników określonych „le”, „la” i „les” oraz rodzajników nieokreślonych „un”, „une” i „des”.

**Tabela 42.** Statystyka wzorców translacyjnych

$l_1$	$l_2$	$l_3$	$l_4$	$l_5$	$l_{\geq 6}$
51	41	18	9	12	1

Minimalna liczba wzorców translacyjnych  $L_F$ , jakie powinna zawierać baza danych, potrzebnych do sprawnego funkcjonowania systemu automatycznej translacji z języka francuskiego na polski może zostać oszacowana za pomocą następującego wzoru

$$L_F = S_F \cdot f_F \cdot \left( 1 + \frac{l_2 + l_3 + l_4 + l_5 + l_{\geq 6}}{l_1} \right) \quad (12)$$

Występujący we wzorze (12) parametr  $S_F$  oznacza liczebność zbioru haseł słownika języka francuskiego. Wartość parametru  $S_F$  szacowana może być w granicach od 10 000 (w przypadku uwzględnienia jedynie podstawowego słownictwa języka francuskiego) do 100 000 (w przypadku dodatkowego uwzględnienia słownictwa specjalistycznego). Na przykład słownik [14] zawiera około 60 000 haseł ze zbioru słownictwa języka francuskiego. Zatem przyjmując  $S_F = 60 000$  otrzymuje się oszacowanie liczby potrzebnych wzorców translacyjnych  $L_F = 511 200$ .

## 8. SZACOWANIE ROZMIARÓW BAZY DANYCH DLA JĘZYKA WŁOSKIEGO

Język włoski należy do romańskiej grupy językowej. Jego cechą charakterystyczną jest zanik form deklinacyjnych rzeczowników przy jednoczesnej bardzo rozbudowanej fleksji czasownikowej.

Częściami mowy, dla których można utworzyć formy fleksyjne, są w języku włoskim rzeczowniki, czasowniki i przymiotniki.

W języku włoskim zanikły całkowicie formy deklinacyjne rzeczowników, zatem włoski rzeczownik oprócz swej formy podstawowej posiada jeszcze tylko formę liczby mnogiej. W tabeli 43 zamieszczono wybrane włoskie rzeczowniki wraz z ich formami liczby mnogiej.

**Tabela 43.** Zestawienie form fleksyjnych wybranych rzeczowników języka włoskiego

Forma podstawowa	Forma liczby mnogiej	Znaczenie
uomo	uomini	człowiek
templo	templi	świątynia
dito	dita	palec
uovo	uova	jajko
ala	ali	skrzydło
porta	porte	drzwi
torre	torri	wieża
mano	mani	ręka
albero	alberi	drzewo
padre	padri	ojciec
pacco	pacchi	paczka

Jak wynika z powyższego zestawienia, współczynnik fleksyjności włoskiego rzeczownika wynosi  $f_I(R) = 2$ .

Z kolei włoskie czasowniki posiadają wiele form fleksyjnych, ponieważ podlegają koniugacji w kilku czasach i trybach. Zjawiska związane z fleksją włoskich czasowników zostały zilustrowane na przykładzie odmiany czasownika „essere”, któremu w języku polskim odpowiada czasownik „być”.

Odmiana czasownika „essere” w trybie orzekającym w czasie teraźniejszym (presente) została zamieszczona w tabeli 44.

**Tabela 44.** Odmiana czasownika „essere” w czasie teraźniejszym (presente)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	sono	siamo
Druga	sei	siete
Trzecia	è	sono



Z kolei w tabeli 45 zamieszczono wzór odmiany czasownika „essere” w czasie przeszłym (imperfetto).

**Tabela 45.** Wzór odmiany czasownika „essere” w czasie przeszłym (imperfetto)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	ero	eravamo
Druga	eri	eravate
Trzecia	era	erano

W tabeli 46 zamieszczono wzór odmiany czasownika „essere” w czasie przeszłym (passato remoto).

**Tabela 46.** Wzór odmiany czasownika „essere” w czasie przeszłym (passato remoto)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	fui	fummo
Druga	fosti	foste
Trzecia	fu	furono

Natomiast w tabeli 47 zamieszczono wzór odmiany czasownika „essere” w czasie przyszłym (futuro).

**Tabela 47.** Wzór odmiany czasownika „essere” w czasie przyszłym (futuro)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	sarò	saramo
Druga	sarai	sarete
Trzecia	sarà	saranno

Ponadto czasownik „essere” posiada formy trybu rozkazującego (imperativo), które zestawiono w tabeli 48.

**Tabela 48.** Zestawienie form fleksyjnych czasownika „essere” w trybie rozkazującym (imperativo)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	–	siamo
Druga	sii	siate
Trzecia	sia	siano

Włoski czasownik „essere” odmienia się również w trybie łączącym (congiuntivo). Wzór odmiany w czasie teraźniejszym (presente) zamieszczono w tabeli 49.

**Tabela 49.** Zestawianie form fleksyjnych czasownika „essere” w trybie łączącym (congiuntivo) w czasie teraźniejszym (presente)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	sia	siamo
Druga	sia	siate
Trzecia	sia	siano

Z kolei w tabeli 50 dokonano zestawienia form odmiany czasownika „essere” w trybie łączącym (congiuntivo) w czasie przeszłym (imperfetto).

**Tabela 50.** Zestawianie form fleksyjnych czasownika „essere” w trybie łączącym (congiuntivo) w czasie teraźniejszym (presente)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	fossi	fossimo
Druga	fossi	foste
Trzecia	fosse	fossero

Z kolei w tabeli 51 zamieszczono wzór odmiany czasownika „essere” w trybie warunkowym (condizionale).

**Tabela 51.** Wzór odmiany czasownika „essere” w trybie warunkowym (condizionale)

Osoba pierwsza	sarei	saremmo
Osoba druga	saresti	sareste
Osoba trzecia	sarebbe	sarebbero

Ponadto czasownik „essere” posiada dodatkowe formy imiesłowowe: „stato” (participio) oraz „essendo” (gerundio).

Zatem współczynnik fleksyjności włoskiego czasownika wynosi  $f_I(C) = 41$ .

W języku włoskim przymiotnik oprócz swej formy podstawowej posiada jeszcze dodatkowe formy fleksyjne dla liczby mnogiej, a także dla rodzaju żeńskiego liczby pojedynczej i mnogiej. W tabeli 52 dokonano zestawienia form fleksyjnych wybranych włoskich przymiotników.

**Tabela 52.** Zestawienie form fleksyjnych wybranych włoskich przymiotników

Forma podstawowa	Forma liczby mnogiej	Forma rodzaju żeńskiego – liczba pojedyncza	Forma rodzaju żeńskiego – liczba mnoga	Znaczenie
bello	belli	bella	belle	piękny
cattivo	cattivi	cattiva	cative	zły
piccolo	piccoli	piccola	piccole	mały
alto	alti	alta	alte	wysoki
basso	bassi	bassa	basse	niski
moderno	moderni	moderna	moderne	nowoczesny
spazioso	spaziosi	spaziosa	spaziose	przestronny
stanco	stanci	stanca	stance	zmęczony
ultimo	ultimi	ultima	ultime	ostatni
nuovo	nuovi	nuova	nuove	nowy
vicino	vicini	vicina	vicine	bliski

Zatem współczynnik fleksyjności włoskiego przymiotnika wynosi  $f_I(P) = 4$ .

Wartość współczynnika fleksyjności języka włoskiego może zostać wyznaczona ze wzoru

$$f_I = w_I(R) \cdot f_I(R) + w_I(C) \cdot f_I(C) + w_I(P) \cdot f_I(P) \quad (13)$$

Występujące we wzorze (13) współczynniki wagowe  $w_I(R)$ ,  $w_I(C)$  oraz  $w_I(P)$  określają częstość występowania rzeczowników, czasowników i przymiotników w słownictwie języka włoskiego.

Częstotliwość występowania rzeczowników, czasowników i przymiotników w słownictwie języka włoskiego została wyznaczona na podstawie analizy następującego zbioru zawierającego sto haseł pobranych kolejno ze słownika [15].

- |                      |                         |
|----------------------|-------------------------|
| 1) manuale (R),      | 40) marmellata (R),     |
| 2) manubrio (R),     | 41) marmitta (R),       |
| 3) manutenzione (R), | 42) marmo (R),          |
| 4) manzo (R),        | 43) marmoreo (P),       |
| 5) mappa (R),        | 44) marmotta (R),       |
| 6) mappamondo (R),   | 45) marna (R),          |
| 7) maraschino (R),   | 46) marocchino (R),     |
| 8) marasmo (R),      | 47) marrone (P),        |
| 9) maratona (R),     | 48) martedì (R),        |
| 10) maraviglia (R),  | 49) martellare (C),     |
| 11) marca (R),       | 50) martellata (R),     |
| 12) marcare (C),     | 51) martello (R),       |
| 13) marchesa (R),    | 52) martire (R),        |
| 14) marchese (R),    | 53) martirio (R),       |
| 15) marchio (R),     | 54) martirizzare (C),   |
| 16) marcia (R),      | 55) martirologio (R),   |
| 17) marciapiede (R), | 56) martoriare (C),     |
| 18) marciare (C),    | 57) marxismo (R),       |
| 19) marcio (P),      | 58) marxista (R),       |
| 20) marcire (C),     | 59) marzapane (R),      |
| 21) marciume (R),    | 60) marziale (P),       |
| 22) marco (R),       | 61) marziano (R),       |
| 23) mare (R),        | 62) marzo (R),          |
| 24) marea (R),       | 63) mascalzone (R),     |
| 25) maremma (R),     | 64) mascella (R),       |
| 26) maresciallo (R), | 65) mascellare (P),     |
| 27) margherita (R),  | 66) maschera (R),       |
| 28) marginale (P),   | 67) mascheramento (R),  |
| 29) margine (R),     | 68) mascherare (C),     |
| 30) marina (R),      | 69) mascherina (R),     |
| 31) marinaio (R),    | 70) maschile (P),       |
| 32) marinare (C),    | 71) maschio (R),        |
| 33) marinaresco (P), | 72) massa (R),          |
| 34) marino (P),      | 73) massacrare (C),     |
| 35) maritare (C),    | 74) massacro (R),       |
| 36) maritata (P),    | 75) massaggiatore (R),  |
| 37) marito (R),      | 76) massaggiatrice (R), |
| 38) marittimo (P),   | 77) massaggio (R),      |
| 39) mariuolo (R),    | 78) massaia (R),        |

- |                     |                       |
|---------------------|-----------------------|
| 79) masseria (R),   | 90) materassa (R),    |
| 80) massiccio (R),  | 91) materia (R),      |
| 81) massima (R),    | 92) materiale (R),    |
| 82) massimale (P),  | 93) materialismo (R), |
| 83) massimo (R),    | 94) maternita (R),    |
| 84) masso (R),      | 95) materno (P),      |
| 85) massone (R),    | 96) matita (R),       |
| 86) masticare (C),  | 97) matriarcato (R),  |
| 87) mastro (R),     | 98) matrice (R),      |
| 88) matassa (R),    | 99) matricida (R),    |
| 89) matematica (R), | 100) matricola (R)    |

W powyższym zestawieniu zastosowano następujące oznaczenia części mowy:

- R – rzeczownik,
- C – czasownik,
- P – przymiotnik,
- X – pozostałe części mowy języka włoskiego nie podlegające fleksji.

Zestawienie to zawiera 100 wyrazów, wśród których można wyróżnić:

- 77 – rzeczowników,
- 11 – czasowników,
- 12 – przymiotników.

Stąd współczynnikiem wagowym przypisane zostały następujące wartości:

$$w_I(R) = 0,77,$$

$$w_I(C) = 0,11,$$

$$w_I(P) = 0,12.$$

Zatem wartość współczynnika fleksyjności języka włoskiego może zostać oszacowana jako  $f_I = 6,53$ .

W tabeli 53 zamieszczono zestawienie dostarczające informacji o tym, ile podczas tłumaczenia wystąpiło wzorców translacyjnych o długości jednego wyrazu ( $l_1$ ), dwóch wyrazów ( $l_2$ ), trzech wyrazów ( $l_3$ ), czterech wyrazów ( $l_4$ ), pięciu wyrazów ( $l_5$ ) oraz dłuższych niż pięć wyrazów ( $l_{\geq 6}$ ). Podczas przeprowadzonych wyliczeń, jako oddzielnych wyrazów nie liczono włoskich rodzajników określonych „il”, „lo”, „la”, „le”, „i” i „gli” oraz rodzajników nieokreślonych „un”, „una” i „uno”.

**Tabela 53.** Statystyka wzorców translacyjnych

$l_1$	$l_2$	$l_3$	$l_4$	$l_5$	$l_{\geq 6}$
72	31	23	12	2	1

Minimalna liczba wzorców translacyjnych  $L_I$ , jakie powinna zawierać baza danych, potrzebnych do sprawnego funkcjonowania systemu automatycznej translacji z języka włoskiego na polski, może zostać oszacowana za pomocą następującego wzoru

$$L_I = S_I \cdot f_I \cdot \left( 1 + \frac{l_2 + l_3 + l_4 + l_5 + l_{\geq 6}}{l_1} \right) \quad (14)$$

We wzorze (14) parametr  $S_I$  oznacza liczebność zbioru haseł słownika języka włoskiego. Wartość parametru  $S_I$  szacowana może być w granicach od 10 000 (w przypadku uwzględnienia jedynie podstawowego słownictwa języka włoskiego) do 100 000 (w przypadku dodatkowego uwzględnienia słownictwa specjalistycznego). Na przykład słownik [16] zawiera około 60 000 haseł ze zbioru słownictwa języka włoskiego. Zatem przyjmując  $S_I = 60\,000$  otrzymuje się oszacowanie liczby potrzebnych wzorców translacyjnych  $L_I = 767\,300$ .

## 9. SZACOWANIE ROZMIARÓW BAZY DANYCH DLA JĘZYKA HISZPAŃSKIEGO

Język hiszpański należy do romańskiej grupy językowej. Jego cechą charakterystyczną jest zanik form deklinacyjnych rzeczowników przy jednoczesnej bardzo rozbudowanej fleksji czasownikowej [17].

Częściami mowy, dla których można utworzyć formy fleksyjne, są w języku hiszpańskim rzeczowniki, czasowniki i przymiotniki.

W języku hiszpańskim rzeczownik oprócz swojej formy podstawowej posiada również formę liczby mnogiej. Dla przykładu w tabeli 54 zamieszczono kilka wybranych rzeczowników hiszpańskich wraz z ich formami liczby mnogiej.

**Tabela 54.** Zestawienie form fleksyjnych wybranych rzeczowników języka hiszpańskiego

Forma podstawowa	Forma liczby mnogiej	Znaczenie
ojo	ojos	oko
planta	plantas	roślina
acuerdo	acuerdos	uchwała
pared	paredes	ściana
pez	peces	smoła
reloj	relojes	zegar
manzana	manzanas	jabłko
artista	artistas	artysta

Jak widać, współczynnik fleksyjności dla hiszpańskiego rzeczownika wynosi  $f_E(R) = 2$ .

Z kolei hiszpański czasownik posiada bardzo wiele form fleksyjnych, ponieważ odmienia się przez osoby w różnych czasach gramatycznych i trybach. Fleksja hiszpańskiego czasownika została zilustrowana na przykładzie odmiany czasownika „estar”, któremu w języku polskim odpowiada ją czasowniki „być” lub „znajdować się”.

Odmiana hiszpańskiego czasownika „estar” w czasie teraźniejszym (presente) została zamieszczona w tabeli 55.

**Tabela 55.** Formy koniugacyjne czasownika „estar” w czasie teraźniejszym (presente)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	estoy	estamos
Druga	estás	estáis
Trzecia	está	están

Z kolei w tabeli 56 zamieszczono wzór odmiany czasownika „estar” w czasie przeszłym (pretérito indefinido).

**Tabela 56.** Formy koniugacyjne czasownika „estar” w czasie przeszłym (pretérito indefinido)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	estuve	estuvimos
Druga	estuviste	estuvisteis
Trzecia	estuvo	estuvieron

W tabeli 57 zamieszczono wzór odmiany czasownika „estar” w czasie przeszłym (pretérito imperfecto).

**Tabela 57.** Formy koniugacyjne czasownika „estar” w czasie przeszłym (pretérito imperfecto)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	estaba	estábamos
Druga	estabas	estabais
Trzecia	estaba	estaban

W tabeli 58 zamieszczono wzór odmiany czasownika „estar” w czasie przyszłym (futuro imperfecto).

**Tabela 58.** Formy koniugacyjne czasownika „estar” w czasie przyszłym (futuro imperfecto)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	estaré	estaremos
Druga	estarás	estaréis
Trzecia	estará	estarán

W tabeli 59 zamieszczono wzór odmiany czasownika „estar” w trybie warunkowym prostym (potencial simple).

**Tabela 59.** Formy koniugacyjne czasownika „estar” w trybie warunkowym prostym (potencial simple)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	estaría	estaríamos
Druga	estarías	estaríais
Trzecia	estaría	estarían

W tabeli 60 zamieszczono wzór odmiany czasownika „estar” w trybie łączącym (modo subjuntivo) w czasie teraźniejszym (presente).

**Tabela 60.** Formy koniugacyjne czasownika „estar” w trybie łączącym (modo subjuntivo) w czasie teraźniejszym (presente)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	esté	estemos
Druga	estés	estéis
Trzecia	esté	estén

Hiszpański czasownik w trybie łączącym (modo subjuntivo) posiada w czasie przeszłym (pretérito imperfecto) dwa alternatywne wzorce odmiany, które dla czasownika „estar” zamieszczono w tabelach 61 i 62.

**Tabela 61.** Formy koniugacyjne czasownika „estar” w trybie łączącym (modo subjuntivo) w czasie przeszłym (pretérito imperfecto) – pierwszy wariant odmiany

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	estuviera	estuviéramos
Druga	estuvieras	estuvierais
Trzecia	estuviera	estuvieran

**Tabela 62.** Formy koniugacyjne czasownika „estar” w trybie łączącym (modo subjuntivo) w czasie przeszłym (pretérito imperfecto) – drugi wariant odmiany

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	estuviese	estuviésemos
Druga	estuvieses	estuvieseis
Trzecia	estuviese	estuviesen

W tabeli 63 zamieszczono wzór odmiany czasownika „estar” w trybie łączącym (modo subjuntivo) w czasie przyszłym (futuro imperfecto).

**Tabela 63.** Formy koniugacyjne czasownika „estar” w trybie łączącym (modo subjuntivo) w czasie przyszłym (futuro imperfecto)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	estuviere	estuviéremos
Druga	estuvieres	estuviereis
Trzecia	estuviere	estuvieren

Ponadto czasownik „estar” posiada dodatkowe formy fleksyjne odpowiadające imiesłowowi czynnemu (gerun-

dio) – „estando” oraz biernemu (participio pasivo) – „estado”. Ponadto wyróżnia się jeszcze formy fleksyjne dla trybu rozkazującego (imperativo) dla liczby pojedynczej „está” i mnogiej „estad”.

Zatem współczynnik fleksyjności hiszpańskiego czasownika wynosi  $f_E(C) = 52$ .

Z kolei hiszpański przymiotnik oprócz formy podstawowej posiada odrębne formy fleksyjne dla liczby mnogiej oraz dla rodzaju żeńskiego w liczbie pojedynczej i mnogiej. Zestawienie form fleksyjnych dla wybranych przymiotników hiszpańskich zamieszczono w tabeli 64.

**Tabela 64.** Zestawienie form fleksyjnych dla wybranych przymiotników hiszpańskich

Forma podstawowa	Forma liczby mnogiej	Forma rodzaju żeńskiego – liczba pojedyncza	Forma rodzaju żeńskiego – liczba mnoga	Znaczenie
bueno	buenos	bueno	buenas	dobry
malo	malos	mala	malas	zły
alto	altos	alta	altas	wysoki
bajo	bajos	baja	bajas	niski
pequeño	pequeños	pequeña	pequeñas	mały
español	españoles	española	españolas	hiszpański
delgado	delgados	delgada	delgadas	szczupły
gordo	gordos	gorda	gordas	gruby
caro	caros	cara	caras	drogi
extranjero	extranjeros	extranjera	extranjeras	zagraniczny
decidido	decididos	decidida	decididas	zdecydowany

Zatem współczynnik fleksyjności hiszpańskiego przymiotnika wynosi  $f_E(P) = 4$ .

Wartość współczynnika fleksyjności języka hiszpańskiego może zostać wyznaczona ze wzoru

$$f_E = w_E(R) \cdot f_E(R) + w_E(C) \cdot f_E(C) + w_E(P) \cdot f_E(P) \quad (15)$$

Występujące we wzorze (15) współczynniki wagowe  $w_E(R)$ ,  $w_E(C)$  oraz  $w_E(P)$  określają częstość występowania rzeczowników, czasowników i przymiotników w słownictwie języka hiszpańskiego.

Częstotliwość występowania rzeczowników, czasowników i przymiotników w słownictwie języka hiszpańskiego została wyznaczona na podstawie analizy następującego zbioru zawierającego sto haseł pobranych kolejno ze słownika [17].

- |                          |                           |
|--------------------------|---------------------------|
| 1) majada (R),           | 51) malva (R),            |
| 2) majadero (P),         | 52) malvado (P),          |
| 3) majar (C),            | 53) mal-versación (R),    |
| 4) majestad (R),         | 54) malversar (C),        |
| 5) majestuoso (P),       | 55) malla (R),            |
| 6) majo (P),             | 56) mama (R),             |
| 7) mal (R),              | 57) mamá (R),             |
| 8) malacostumbrado (P),  | 58) mamar (C),            |
| 9) malandanza (R),       | 59) mamífero (R),         |
| 10) malavenido (P),      | 60) mamón (R),            |
| 11) malaventura (R),     | 61) mampostería (R),      |
| 12) malbaratar (C),      | 62) manada (R),           |
| 13) malcontento (P),     | 63) manantial (R),        |
| 14) malcriado (P),       | 64) manar (C),            |
| 15) maldad (R),          | 65) mancar (C),           |
| 16) maldecir (C),        | 66) manceba (R),          |
| 17) maldición,           | 67) mancilla (R),         |
| 18) maldispuesto (P),    | 68) manco (R),            |
| 19) maldita (R),         | 69) mancomunadamente (P), |
| 20) maldito (P),         | 70) mancha (R),           |
| 21) maleable (P),        | 71) manchar (C),          |
| 22) maleante (R),        | 72) mandado (R),          |
| 23) malecón (R),         | 73) mandamiento (R),      |
| 24) maledicencia (R),    | 74) mandar (C),           |
| 25) maleficio (R),       | 75) mandatario (R),       |
| 26) maléfico (P),        | 76) mandíbula (R),        |
| 27) malestar (R),        | 77) mandil (R),           |
| 28) maleta (R),          | 78) mando (R),            |
| 29) maletilla (R),       | 79) mandolín (R),         |
| 30) malevolencia (R),    | 80) mandolina (R),        |
| 31) malévolo (P),        | 81) manducar (C),         |
| 32) maleza (R),          | 82) manecilla (R),        |
| 33) malgastar (C),       | 83) manejable (P),        |
| 34) malhadado (P),       | 84) manejar (C),          |
| 35) malhechor (R),       | 85) manera (R),           |
| 36) malhumorado (P),     | 86) manga (R),            |
| 37) malicia (R),         | 87) manganesa (R),        |
| 38) malicioso (P),       | 88) mango (R),            |
| 39) maligno (P),         | 89) manguera (R),         |
| 40) malintencionado (P), | 90) manguito (R),         |
| 41) malo (P),            | 91) manía (R),            |
| 42) malograr (C),        | 92) maniático (R),        |
| 43) malparto (R),        | 93) manicomio (R),        |
| 44) malquerencia (R),    | 94) manicestación (R),    |
| 45) malquisto (P),       | 95) manifestar (C),       |
| 46) malrotar (C),        | 96) manifiesto (P),       |
| 47) malsano (P),         | 97) manija (R),           |
| 48) maltratamiento (R),  | 98) manilla (R),          |
| 49) maltrato (R),        | 99) maniobra (R),         |
| 50) maltratar (C),       | 100) manipulación (R)     |

W powyższym zestawieniu zastosowano następujące oznaczenia części mowy:

- R – rzeczownik,
- C – czasownik,
- P – przymiotnik,
- X – pozostałe części mowy języka hiszpańskiego nie podlegające fleksji.

Zestawienie to zawiera 100 wyrazów, wśród których można wyróżnić:

- 60 – rzeczowników,
- 16 – czasowników,
- 24 – przymiotniki.

Stąd współczynnikom wagowym przypisane zostały następujące wartości:

$$w_{GB}(R) = 0,60,$$

$$w_{GB}(C) = 0,16,$$

$$w_{GB}(P) = 0,4.$$

Zatem wartość współczynnika fleksyjności języka hiszpańskiego może zostać oszacowana jako  $f_E = 10,48$ .

W tabeli 65 zamieszczono zestawienie dostarczające informacji o tym, ile podczas tłumaczenia wystąpiło wzorców translacyjnych o długości jednego wyrazu ( $l_1$ ), dwóch wyrazów ( $l_2$ ), trzech wyrazów ( $l_3$ ), czterech wyrazów ( $l_4$ ), pięciu wyrazów ( $l_5$ ) oraz dłuższych niż pięć wyrazów ( $l_{\geq 6}$ ). Podczas przeprowadzonych wyliczeń, jako oddzielnych wyrazów nie liczone hiszpańskich rodzajników określonych „le”, „la”, „los” i „las” oraz rodzajników nieokreślonych „un”, „una”, „unos” i „unas”.

**Tabela 65.** Statystyka wzorców translacyjnych

$l_1$	$l_2$	$l_3$	$l_4$	$l_5$	$l_{\geq 6}$
33	31	26	11	2	2

Minimalna liczba wzorców translacyjnych  $L_E$ , jakie powinna zawierać baza danych, potrzebnych do sprawnego funkcjonowania systemu automatycznej translacji z języka hiszpańskiego na polski, może zostać oszacowana za pomocą następującego wzoru

$$L_E = S_E \cdot f_E \cdot \left( 1 + \frac{l_2 + l_3 + l_4 + l_5 + l_{\geq 6}}{l_1} \right) \quad (16)$$

Występujący we wzorze (16) parametr  $S_E$  oznacza liczebność zbioru haseł słownika języka hiszpańskiego. Wartość parametru  $S_E$  szacowana może być w granicach od 10 000 (w przypadku uwzględnienia jedynie podstawowego słownictwa języka hiszpańskiego) do 100 000 (w przypadku dodatkowego uwzględnienia słownictwa specjalistycznego). Na przykład słownik [13] zawiera około 40 000 haseł ze zbioru słownictwa języka hiszpańskiego. Zatem przyjmując  $S_E = 40 000$  otrzymuje się oszacowanie liczby potrzebnych przykładów translacyjnych  $L_E = 1 333 800$ .

## 10. SZACOWANIE ROZMIARÓW BAZY DANYCH DLA JĘZYKA PORTUGALSKIEGO

Język portugalski należy do romańskiej rodziny językowej. Jego cechą charakterystyczną jest niewielka liczba form fleksyjnych wyrazów [18].

W języku portugalskim formy fleksyjne można utworzyć od takich części mowy, jak rzeczowniki, czasowniki oraz przymiotniki.

W języku portugalskim rzeczownik oprócz formy liczby pojedynczej posiada jeszcze formę liczby mnogiej. Nie występuje natomiast deklinacja rzeczowników. Zestawienie przykładowych portugalskich rzeczowników zostało zamieszczone w tabeli 66.

**Tabela 66.** Zestawienie form liczby pojedynczej i mnogiej przykładowych portugalskich rzeczowników

Forma liczby pojedynczej	Forma liczby mnogiej	Znaczenie
caderno	cadernos	zeszyt
mesa	mesas	stół
coração	corações	serce
cidadão	cidadãos	obywatel
irmão	irmãos	brat
botão	botões	guzik
lição	lições	lekcja
pão	pães	chleb
cão	cães	pies
fim	fins	koniec
juiz	juizes	sędzia

Zatem współczynnik fleksyjności portugalskiego rzeczownika wynosi  $f_{PO}(R) = 2$ .

Natomiast portugalski czasownik charakteryzuje się bardzo dużą liczbą możliwych do utworzenia form fleksyjnych, odpowiadających różnym czasom i trybom. Fleksja portugalskiego czasownika została omówiona na przykładzie czasownika „vender” (sprzedawać).

W tabeli 67 zamieszczono wzór odmiany czasownika „vender” w czasie teraźniejszym (presente).

**Tabela 67.** Formy koniugacyjne czasownika „vender” w czasie teraźniejszym (presente)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	vendo	vendemos
Druga	vendes	vendeis
Trzecia	vende	vendem

Z kolei w tabeli 68 zamieszczono wzór odmiany czasownika „vender” w czasie przeszłym niedokonanym (pretérito imperfeito).

W tabeli 69 zamieszczono wzór odmiany czasownika „vender” w czasie przeszłym dokonanym (pretérito perfeito).

**Tabela 68.** Formy koniugacyjne czasownika „vender” w czasie przeszłym niedokonanym (pretérito imperfeito)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	vendia	vendíamos
Druga	vendias	vendíeis
Trzecia	vendia	vendíam

**Tabela 69.** Formy koniugacyjne czasownika „vender” w czasie przeszłym dokonanym (pretérito perfeito)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	vendi	vendemos
Druga	vendeste	vendestes
Trzecia	vendeu	venderam

Z kolei w tabeli 70 zamieszczono wzór odmiany czasownika „vender” w czasie zaprzyszłym (mais-que-perfeito simples).

**Tabela 70.** Formy koniugacyjne czasownika „vender” w czasie zaprzyszłym (mais-que-perfeito simples)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	vendera	vendêramos
Druga	venderas	vendêreis
Trzecia	vendera	venderam

W tabeli 71 zamieszczono wzór odmiany czasownika „vender” w czasie przyszłym niedokonanym (futuro imperfeito).

**Tabela 71.** Formy koniugacyjne czasownika „vender” w czasie przyszłym niedokonanym (futuro imperfeito)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	venderei	venderemos
Druga	venderás	vendereis
Trzecia	venderá	venderão

Z kolei w tabeli 72 zamieszczono wzór odmiany czasownika „vender” w trybie warunkowym (modo condicional) w czasie (imperfeito).

**Tabela 72.** Formy koniugacyjne czasownika „vender” w trybie warunkowym (modo condicional) w czasie (imperfeito)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	venderia	venderíamos
Druga	venderias	venderíeis
Trzecia	venderia	venderiam

W tabeli 73 zamieszczono natomiast wzór odmiany czasownika „vender” w trybie łączącym (modo subjuntivo) w czasie (presente).

**Tabela 73.** Formy koniugacyjne czasownika „vender” w trybie łączącym (modo subjuntivo) w czasie (presente)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	venda	vendamos
Druga	vendas	vendais
Trzecia	venda	vendam

Z kolei w tabeli 74 zamieszczono wzór odmiany czasownika „vender” w trybie łączącym (modo subjuntivo) w czasie (pretérito imperfeito).

**Tabela 74.** Formy koniugacyjne czasownika „vender” w trybie łączącym (modo subjuntivo) w czasie (pretérito imperfeito)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	vendessee	vendêssemos
Druga	vendessee	vendêsseis
Trzecia	vendessee	vendessem

W tabeli 75 zamieszczono wzór odmiany czasownika „vender” w trybie łączącym (modo subjuntivo) w czasie (futuro imperfeito).

**Tabela 75.** Formy koniugacyjne czasownika „vender” w trybie łączącym (modo subjuntivo) w czasie (futuro imperfeito)

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	vender	vendermos
Druga	venderes	venderdes
Trzecia	vender	venderem

Ponadto w języku portugalskim wyróżnia się również osobowe formy bezokolicznika w czasie teraźniejszym (modo infinitivo presente pessoal), które dla czasownika „vender” zostały zestawione w tabeli 76.

**Tabela 76.** Zestawienie osobowych form bezokolicznika w czasie teraźniejszym (modo infinitivo presente pessoal) dla czasownika „vender”

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	vender	vendermos
Druga	venderes	venderdes
Trzecia	vender	venderem

Ponadto dla portugalskiego czasownika można utworzyć jeszcze formy imiesłowowe (gerúndio) „vendendo”, (participio passado) „vendido” oraz formy trybu rozkazującego „vende” i „vendei”.

Z przedstawionego zestawienia wynika, że współczynnik fleksyjności portugalskiego czasownika wynosi  $f_{PO}(C) = 50$ .

W języku portugalskim przymiotnik oprócz swej formy podstawowej odpowiadającej rodzajowi męskiemu liczby pojedynczej posiada także formy odpowiadające rodzajowi żeńskiemu liczby pojedynczej oraz formy liczby mnogiej rodzaju męskiego i żeńskiego. Zestawienia form fleksyjnych portugalskiego przymiotnika dokonano w tabeli 77.

**Tabela 77.** Zestawienie form fleksyjnych wybranych portugalskich przymiotników

Forma liczby pojedynczej rodzaju męskiego	Forma liczby pojedynczej rodzaju żeńskiego	Forma liczby mnogiej rodzaju męskiego	Forma liczby mnogiej rodzaju żeńskiego	Znaczenie
celso	celsa	celsos	celsas	wysoki
dileto	dileta	diletos	diletas	ulubiony
digno	digna	dignos	dignas	godny
diminuto	diminuta	diminutos	diminutas	drobny
lento	lenta	lentos	lentas	powolny
lerdo	lerda	lerdos	lerdas	ociężały
ríspido	ríspida	ríspidos	ríspidas	chropowaty
rochoso	rochosa	rochosos	rochosas	skalisty
rôto	rôta	rôtos	rôtas	podarty
roxo	roxa	roxos	roxas	siny
sólido	sólida	sólidos	sólidas	trwały

Zatem współczynnik fleksyjności portugalskiego przymiotnika wynosi  $f_{PO}(P) = 4$ .

Wartość współczynnika fleksyjności języka portugalskiego może zostać wyznaczona ze wzoru

$$f_{PO} = w_{PO}(R) \cdot f_{PO}(R) + w_{PO}(C) \cdot f_{PO}(C) + w_{PO}(P) \cdot f_{PO}(P) \quad (17)$$

Występujące we wzorze (17) współczynniki wagowe  $w_{PO}(R)$ ,  $w_{PO}(C)$  oraz  $w_{PO}(P)$  określają częstość występowania rzeczowników, czasowników i przymiotników w słownictwie języka portugalskiego.

Częstotliwość występowania rzeczowników, czasowników i przymiotników w słownictwie języka portugalskiego została wyznaczona na podstawie analizy następującego zbioru zawierającego sto haseł zaczerpniętych kolejno ze słownika [19].

- |                      |                        |
|----------------------|------------------------|
| 1) fa (R),           | 51) faia (R),          |
| 2) fábrica (R),      | 52) faiança (R),       |
| 3) fabricação (R),   | 53) faina (R),         |
| 4) fabricante (R),   | 54) faisao (R),        |
| 5) fabricante (R),   | 55) faísca (R),        |
| 6) fabricar (C),     | 56) faiscar (C),       |
| 7) fabrico (R),      | 57) faixa (R),         |
| 8) fabril (P),       | 58) fala (R),          |
| 9) fábula (R),       | 59) falasao (R),       |
| 10) fabular (C),     | 60) falácia (R),       |
| 11) fabulista (R),   | 61) falacioso (P),     |
| 12) fabuloso (P),    | 62) falado (P),        |
| 13) faca (R),        | 63) falador (P),       |
| 14) fachada (R),     | 64) falar (C),         |
| 15) fachadista (R),  | 65) falatório (R),     |
| 16) façanha (R),     | 66) falaz (P),         |
| 17) facao (R),       | 67) falcao (R),        |
| 18) fação (R),       | 68) falecer (C),       |
| 19) faccionário (P), | 69) falecido (P),      |
| 20) facciosismo (R), | 70) falecimento (R),   |
| 21) faccioso (P),    | 71) falencia (R),      |
| 22) face (R),        | 72) falésia (R),       |
| 23) facécia (R),     | 73) falha (R),         |
| 24) faceiro (P),     | 74) falhar (C),        |
| 25) faceta (R),      | 75) falho (P),         |
| 26) faceto (P),      | 76) falido (R),        |
| 27) fachada (R),     | 77) falir (C),         |
| 28) facho (R),       | 78) falível (P),       |
| 29) facial (P),      | 79) falsar (C),        |
| 30) facies (R),      | 80) falsário (R),      |
| 31) fácil (P),       | 81) falsear (C),       |
| 32) facilidade (R),  | 82) falsete (R),       |
| 33) facilitar (C),   | 83) falsidade (R),     |
| 34) facilitário (R), | 84) falsificação (R),  |
| 35) facinora (R),    | 85) falsificador (R),  |
| 36) factício (P),    | 86) falsificar (C),    |
| 37) façudo (P),      | 87) falso (P),         |
| 38) faculdade (R),   | 88) falta (R),         |
| 39) facultar (C),    | 89) faltar (C),        |
| 40) facultativo (P), | 90) fama (R),          |
| 41) facundo (P),     | 91) famélico (P),      |
| 42) fada (R),        | 92) famigerado (P),    |
| 43) fadado (P),      | 93) família (R),       |
| 44) fadar (C),       | 94) familiar (P),      |
| 45) fadário (R),     | 95) familiaridade (R), |
| 46) fadiga (R),      | 96) familiarizar (C),  |
| 47) fadista (R),     | 97) faminto (P),       |
| 48) fado (R),        | 98) famoso (P),        |
| 49) fagote (R),      | 99) fâmullo (R),       |
| 50) fagulha (R),     | 100) fanal (R)         |

W powyższym zestawieniu zastosowano następujące oznaczenia części mowy:

- R – rzeczownik,
- C – czasownik,
- P – przymiotnik,
- X – pozostałe części mowy języka portugalskiego nie podlegające fleksji.

Zestawienie to zawiera 100 wyrazów, wśród których można wyróżnić:

- 59 – rzeczowników,
- 15 – czasowników,
- 26 – przymiotników.

Stąd współczynnikiem wagowym przypisane zostały następujące wartości:

$$w_{GB}(R) = 0,59,$$

$$w_{GB}(C) = 0,15,$$

$$w_{GB}(P) = 0,26.$$

Zatem wartość współczynnika fleksyjności języka portugalskiego może zostać oszacowana jako  $f_{PO} = 9,72$ .

W tabeli 78 zamieszczono zestawienie dostarczające informacji o tym, ile podczas tłumaczenia wystąpiło wzorców translacyjnych o długości jednego wyrazu ( $l_1$ ), dwóch wyrazów ( $l_2$ ), trzech wyrazów ( $l_3$ ), czterech wyrazów ( $l_4$ ), pięciu wyrazów ( $l_5$ ) oraz dłuższych niż pięć wyrazów ( $l_{\geq 6}$ ). Podczas przeprowadzonych wyliczeń, jako oddzielnych wyrazów nie liczono rodzajników określonych „o”, „a”, „os” i „as” oraz nieokreślonych „um”, „umos”, „uma” i „umas”.

**Tabela 78.** Statystyka wzorców translacyjnych dla języka portugalskiego

$l_1$	$l_2$	$l_3$	$l_4$	$l_5$	$l_{\geq 6}$
55	20	16	8	1	1

Minimalna liczba wzorców translacyjnych  $L_{PO}$ , jakie powinna zawierać baza danych, potrzebnych do sprawnego funkcjonowania systemu automatycznej translacji z języka portugalskiego na polski, może zostać oszacowana za pomocą następującego wzoru

$$L_{PO} = S_{PO} \cdot f_{PO} \cdot \left( 1 + \frac{l_2 + l_3 + l_4 + l_5 + l_{\geq 6}}{l_1} \right) \quad (18)$$

We wzorze (18) parametr  $S_{PO}$  oznacza liczebność zbioru haseł słownika języka portugalskiego. Wartość parametru  $S_{PO}$  szacowana może być w granicach od 10 000 (w przypadku uwzględnienia jedynie podstawowego słownictwa języka portugalskiego) do 100 000 (w przypadku dodatkowego uwzględnienia słownictwa specjalistycznego). Na przykład słownik [19] zawiera około 20 000 haseł ze zbioru słownictwa języka portugalskiego. Zatem przyjmując  $S_{PO} = 20 000$  otrzymuje się oszacowanie liczby potrzebnych wzorców translacyjnych  $L_{PO} = 357 000$ .

## 11. SZACOWANIE ROZMIARÓW BAZY DANYCH DLA JĘZYKA DUŃSKIEGO

Język duński należy do północnej gałęzi germańskiej grupy językowej. Jego cechą charakterystyczną jest umiarkowana liczba form fleksyjnych wyrazów. W języku duńskim częściami mowy, od których można utworzyć formy fleksyjne, są: rzeczowniki, czasowniki i przymiotniki.



Duński rzeczownik oprócz swej formy podstawowej posiada jeszcze formę liczby mnogiej oraz formy określone dla liczby pojedynczej i mnogiej. Zestawienia form fleksyjnych duńskiego rzeczownika dokonano w tabeli 79.

**Tabela 79.** Zestawienie form fleksyjnych wybranych duńskich rzeczowników

Forma podstawowa	Forma określona dla liczby pojedynczej	Forma nieokreślona dla liczby mnogiej	Forma określona dla liczby mnogiej	Znaczenie formy podstawowej
hest	hesten	heste	hestene	koń
by	byen	byer	byerne	miasto
stol	stolen	stole	stolene	krzesło
sild	silden	sild	sildene	śledź
sko	skoen	sko	skoene	but
bord	bordet	borde	bordene	stół
hus	huset	huse	husene	dom
æble	æblet	æbler	æblerne	jabłko
år	året	år	årene	rok
skib	skibet	skibe	skibene	statek
ord	ordet	ord	ordene	słowo

Zatem współczynnik fleksyjności duńskiego rzeczownika wynosi  $f_{DK}(R) = 4$ .

Z kolei duński czasownik oprócz formy bezokolicznika posiada również odrębne formy dla trybu rozkazującego, czasu teraźniejszego, czasu przeszłego, formę imiesłowu czasu teraźniejszego oraz formę imiesłowu czasu przeszłego. Zestawienie form fleksyjnych wybranych duńskich czasowników zostało zamieszczone w tabeli 80.

**Tabela 80.** Zestawienie form fleksyjnych dla wybranych czasowników języka duńskiego

Forma bezokolicznika	Forma trybu rozkazującego	Forma czasu teraźniejszego	Forma czasu przeszłego	Imie-słów czasu teraźniejszego	Imie-słów czasu przeszłego	Znaczenie formy podstawowej
have	hav	har	havde	havende	haft	mieć
være	vær	er	var	værende	været	być
finde	find	finder	fandt	findende	fundet	znaleźć
drikke	drikk	drikker	drak	drikkende	drukket	pić
binde	bind	binder	bandt	bindende	bundet	wiązać
give	giv	giver	gav	givende	givet	dać
gribe	grib	griber	greb	gribende	grebet	chwycić
skrive	skriv	skriver	skrev	skrivende	skrevet	pisać
sove	sov	sover	sov	sovende	sovet	spać
tage	tag	tager	tog	tagende	taget	brać
ligge	ligg	ligger	lå	liggende	ligget	leżeć

Zatem, jak wynika z tabeli 80, współczynnik fleksyjności duńskiego czasownika wynosi  $f_{DK}(C) = 6$ .

Z kolei dla duńskiego przymiotnika występują odrębne formy fleksyjne, które można utworzyć dla rodzaju nijakiego, liczby mnogiej oraz dla stopnia wyższego i najwyższego. Ich zestawienia dokonano w tabeli 81.

**Tabela 81.** Zestawienie form fleksyjnych wybranych duńskich przymiotników

Rodzaj wspólny	Rodzaj nijaki	Liczba mnoga	Stopień wyższy	Stopień najwyższy	Znaczenie
fin	fint	fine	finere	finest	ładny
stor	stort	store	større	størst	duży
rolig	roligt	rolige	roligere	roligst	spokojny
høj	højt	høje	højere	højest	wysoki
bred	bredt	brede	bredere	bredest	szeroki
smuk	smukt	smuke	smukkere	smukkest	piękny
kedelig	kedeligt	kedelige	kedeligere	kedeligst	nudny
ivrig	ivrigt	ivrige	ivrigere	ivrigst	pilny
morsom	morsomt	morsomme	morsommere	morsommst	zabawny
god	godt	gode	bedre	bedst	dobry
gammel	gammelt	gamle	ældre	ældst	stary

Jak wynika z zestawienia zamieszczonego w tabeli 81, współczynnik fleksyjności duńskiego przymiotnika wynosi  $f_{DK}(P) = 5$ .

Wartość współczynnika fleksyjności języka duńskiego może zostać wyznaczona ze wzoru

$$f_{DK} = w_{DK}(R) \cdot f_{DK}(R) + w_{DK}(C) \cdot f_{DK}(C) + w_{DK}(P) \cdot f_{DK}(P) \quad (19)$$

Występujące we wzorze (19) współczynniki wagowe  $w_{DK}(R)$ ,  $w_{DK}(C)$  oraz  $w_{DK}(P)$  określają częstość występowania rzeczowników, czasowników i przymiotników w słownictwie języka duńskiego.

Częstotliwość występowania rzeczowników, czasowników i przymiotników w słownictwie języka duńskiego została wyznaczona na podstawie analizy następującego zbioru zawierającego sto haseł zaczerpniętych kolejno ze słownika [20]:

- |                 |                    |
|-----------------|--------------------|
| 1) da (X),      | 9) dal (R),        |
| 2) dadel (R),   | 10) dam (R),       |
| 3) dadle (C),   | 11) dame (R),      |
| 4) dag (R),     | 12) damp (R),      |
| 5) dagblad (R), | 13) damper (R),    |
| 6) dagbog (R),  | 14) Danmark (R),   |
| 7) daggry (R),  | 15) danne (C),     |
| 8) daglig (P),  | 16) dannelsen (R), |

17) danse (C),	59) din (X),
18) dansk (R),	60) dirk (R),
19) dansk (P),	61) dirre (C),
20) dansker (R),	62) dis (R),
21) datid (R),	63) diset (P),
22) dato (R),	64) disk (R),
23) datter (R),	65) dit (X),
24) davarende (P),	66) divan (R),
25) de (X),	67) diat (R),
26) De (X),	68) djavel (R),
27) december (R),	69) dobelt (P),
28) dejlig (P),	70) dog (X),
29) del (R),	71) dom (R),
30) dele (C),	72) domfalde (C),
31) deltage (C),	73) domhus (R),
32) delvis (X),	74) dommer (R),
33) dem (X),	75) domstol (R),
34) Dem (X),	76) dorsk (P),
35) den (X),	77) douche (R),
36) dengang (X),	78) dovenskab (R),
37) denne (X),	79) drab (R),
38) der (X),	80) drage (R),
39) deraf (X),	81) drage (C),
40) derefter (X),	82) drager (R),
41) deres (X),	83) dragt (R),
42) Deres (X),	84) drama (R),
43) derfor (X),	85) dranker (R),
44) derfra (X),	86) dreje (C),
45) deriblandt (X),	87) drejebog (R),
46) derigennem (X),	88) drejer (R),
47) dersom (X),	89) dreng (R),
48) dervarende (P),	90) dreven (P),
49) des (X),	91) drevenhed (R),
50) deslige (P),	92) drift (R),
51) desuden (X),	93) driftskapital (R),
52) det (X),	94) driftsomkostninger (R),
53) did (X),	95) drik (R),
54) digt (R),	96) drikke (C),
55) digte (C),	97) dristig (P),
56) digter (R),	98) drive (C),
57) digtning (R),	99) drivhus (R),
58) dild (R),	100) dronning (R)

W powyższym zestawieniu zastosowano następujące oznaczenia części mowy:

- R – rzeczownik,
- C – czasownik,
- P – przymiotnik,
- X – pozostałe części mowy języka duńskiego nie podlegające fleksji.

Zestawienie to zawiera 100 wyrazów, wśród których można wyróżnić:

- 51 – rzeczowników,
- 12 – czasowników,
- 11 – przymiotników.

Stąd współczynnikom wagowym przypisane zostały następujące wartości:

$$w_{DK}(R) = 0,51,$$

$$w_{DK}(C) = 0,12,$$

$$w_{DK}(P) = 0,11.$$

Zatem wartość współczynnika fleksyjności języka duńskiego może zostać oszacowana jako  $f_{DK} = 3,31$ .

W tabeli 82 zamieszczono zestawienie dostarczające informacji o tym, ile podczas tłumaczenia wystąpiło wzorców translacyjnych o długości jednego wyrazu ( $l_1$ ), dwóch wyrazów ( $l_2$ ), trzech wyrazów ( $l_3$ ), czterech wyrazów ( $l_4$ ), pięciu wyrazów ( $l_5$ ) oraz dłuższych niż pięć wyrazów ( $l_{\geq 6}$ ). Podczas przeprowadzonych wyliczeń, jako oddzielnych wyrazów nie liczone duńskich rodzajników nieokreślonych „en” i „et”.

**Tabela 82.** Statystyka wzorców translacyjnych

$l_1$	$l_2$	$l_3$	$l_4$	$l_5$	$l_{\geq 6}$
62	39	17	11	1	1

Minimalna liczba wzorców translacyjnych  $L_{DK}$ , jakie powinna zawierać baza danych, potrzebnych do sprawnego funkcjonowania systemu automatycznej translacji z języka duńskiego na polski, może zostać oszacowana za pomocą następującego wzoru

$$L_{DK} = S_{DK} \cdot f_{DK} \cdot \left( 1 + \frac{l_2 + l_3 + l_4 + l_5 + l_{\geq 6}}{l_1} \right) \quad (20)$$

Występujący we wzorze (20) parametr  $S_{DK}$  oznacza liczebność zbioru haseł słownika języka duńskiego. Wartość parametru  $S_{DK}$  szacowana może być w granicach od 10 000 (w przypadku uwzględnienia jedynie podstawowego słownictwa języka duńskiego) do 100 000 (w przypadku dodatkowego uwzględnienia słownictwa specjalistycznego). Na przykład słownik [20] zawiera około 6000 haseł ze zbioru słownictwa języka duńskiego. Zatem  $S_{DK} = 6000$  przyjmując otrzymuje się oszacowanie liczby potrzebnych przykładów translacyjnych  $L_{DK} = 42\,280$ .

## 12. ZAKOŃCZENIE

W artykule podjęto próbę oszacowania rozmiarów baz danych zawierających wzorce translacyjne dla dziesięciu wybranych języków należących do romańskiej i germańskiej grup języków indoeuropejskich. Przeprowadzone oszacowania wykazały, że takie bazy danych powinny posiadać co najmniej około kilkuset tysięcy wzorców translacyjnych, aby oparte na nich systemy automatycznego przekładu były w stanie dostarczyć jakichkolwiek praktycznie użytecznych rezultatów. Aby osiągnąć wysoką jakość przekładu dla tek-

stów należących do języka ogólnego, liczba wzorców translacyjnych zawartych w bazie danych powinna być prawdopodobnie co najmniej o rząd wielkości większa, a zatem musiałaby wynosić kilka milionów. Bazy danych o tak potężnych rozmiarach nie można zgromadzić w krótkim czasie, ponieważ każdy z wzorców translacyjnych wymaga indywidualnego przetłumaczenia przez specjalistę. Zatem opracowanie sprawnego systemu translacji opartego na metodzie wzorców translacyjnych wymaga wielu lat pracy zespołu lingwistów, którzy bazując na bilingwicznych korpusach tekstów, byłoby w stanie zapełnić bazę danych przynajmniej najczęściej występującymi w danym języku związkami frazeologicznymi. Trud włożony w opracowanie takiego systemu powinien się jednak opłacić, ponieważ zgodnie ze wstępnymi wynikami badań przeprowadzonymi przez autora, jakość przekładu uzyskanego metodą wzorców translacyjnych powinna być znacznie wyższa niż w przypadku systemów opartych tylko na regułach gramatycznych i bardziej przypominać rezultat pracy człowieka tłumacza.

Przeprowadzone w artykule próby oszacowania rozmiarów baz danych dla wybranych języków pokazały, że różne języki są w różnym stopniu podatne na implementację dla nich metody translacji automatycznej opartej na wzorcach. Języki germańskie okazały się przy tym znacznie lepsze od języków romańskich, ponieważ w przypadku tych pierwszych występuje znacznie mniej możliwych do utworzenia form fleksyjnych wyrazów. Uwaga to dotyczy zwłaszcza języków należących do północnej gałęzi grupy germańskiej, gdzie w ogóle nie występuje deklinacja rzeczowników i koniugacja czasowników. Na drugim biegunie znajdują się współczesne języki romańskie, w przypadku których również nie występuje deklinacja rzeczowników, ale niestety mamy za to do czynienia z prawdziwą mnogością form fleksyjnych czasowników, które mogą występować w różnych czasach i trybach gramatycznych.

## Literatura

- [1] Smólska J., Zawadzka A.: *Gramatyka języka angielskiego*. Warszawa, Wydawnictwa Szkolne i Pedagogiczne 1987, ISBN 83-02-00258-5
- [2] Stanisławski J.: *Wielki słownik angielsko-polski z suplementem*. Warszawa, Wiedza Powszechna 1983
- [3] Stanisławski J., Billip K., Chociłowska Z.: *Podręczny słownik angielsko-polski*. Warszawa, Wiedza Powszechna 1983
- [4] Bender A.: *Mały słownik niemiecko-polski*. Warszawa, Wydawnictwa Naukowo-Techniczne 1991
- [5] Chodera J., Kubica S.: *Podręczny słownik niemiecko-polski*. Warszawa, Wiedza Powszechna 1989
- [6] Martens N., Miec E.: *Mały słownik niderlandzko-polski i polsko-niderlandzki*. Warszawa, Wiedza Powszechna 1991
- [7] Lindquist M., Stypińska M.: *På svenska – Podręcznik do nauki języka szwedzkiego dla początkujących*. Warszawa, PWN 1991
- [8] Leonard P.: *Słownik minimum szwedzko-polski*. Warszawa, Wiedza Powszechna, 1994
- [9] Sikorski L.: *Mały słownik szwedzko-polski i polsko-szwedzki*. Warszawa, Wiedza Powszechna 1990
- [10] Frank-Oborzynska E.: *Słownik minimum norwesko-polski i polsko-norweski*. Warszawa, Wiedza Powszechna 1993
- [11] Lebek H.: *Zarys gramatyki francuskiej*. Warszawa, PWN 1967
- [12] Kwiatkowski W., Sobczyński K.: *Nowy słownik francusko-polski i polsko-francuski*. Warszawa, Wydawnictwo „Spotkania” 1989
- [13] Wawrzykiewicz S., Hiszpański K.: *Podręczny słownik hiszpańsko-polski*. Warszawa, Wiedza Powszechna 1983
- [14] Kupisz K., Kielski B.: *Podręczny słownik francusko-polski z suplementem*. Warszawa, Wiedza Powszechna 1993
- [15] Soja S., Zawadzka C., Zawadzki Z.: *Mały słownik włosko-polski i polsko-włoski*. Warszawa, Wiedza Powszechna 1993
- [16] Meisels W.: *Podręczny słownik włosko-polski*. Warszawa, Wiedza Powszechna 1989
- [17] Perlin J.: *Mała gramatyka języka hiszpańskiego*. Warszawa, Wiedza Powszechna 1995
- [18] Łukaszczuk E.: *Język portugalski od A do Z – repetytorium*. Warszawa, Wydawnictwo „Kram” 1998
- [19] Sachs V., Sachs I.: *Mały słownik portugalsko-polski*. Warszawa, Wiedza Powszechna 1996
- [20] Frank-Obodrzycka E.: *Słownik minimum duńsko-polski i polsko-duński*. Warszawa, Wiedza Powszechna 1996
- [21] Gajer M.: *Towards a Fully-Automatic High-Quality Machine Translation System for Unrestricted Text*. Machine Translation Review, No. 12, 2001, 10–20
- [22] Gajer M.: *The Pattern-Based French-to-Polish machine Translation System*. Machine Translation Review, No. 13, 2002, 7–41

Wpłynęło: 10.02.2005

Miroslaw GAJER



Urodził się 25 kwietnia 1971 roku. Ukończył kierunek elektronika ze specjalnością aparatura elektroniczna na Wydziale Elektrotechniki, Automatyki i Elektroniki Akademii Górniczo-Hutniczej im. Stanisława Staszica w Krakowie. Dyplom magistra inżyniera otrzymał w 1996 roku. Następnie kontynuował naukę na studiach doktoranckich na Wydziale Elektrotechniki, Automatyki i Elektroniki Akademii Górniczo-Hutniczej, które zakończył złożeniem rozprawy doktorskiej. W dniu 26 października 2000 roku Rada Wydziału Elektrotechniki, Automatyki, Informatyki i Elektroniki Akademii Górniczo-Hutniczej nadała mu tytuł doktora nauk technicznych w dyscyplinie informatyka oraz przyznane mu zostało wyróżnienie. Obecnie jest pracownikiem naukowo-dydaktycznym w Katedrze Automatyki AGH, zatrudnionym na stanowisku adiunkta. Swoje obecne zainteresowania wiąże z dziedziną sztucznej inteligencji, a zwłaszcza z technikami przetwarzania języka naturalnego i automatycznej translacji.

e-mail: mgajer@ia.agh.edu.pl