

Katarzyna Marszałek–Kowalewska

Irańska polityka językowa wobec zapożyczeń:
korpusowe badanie angielskich zapożyczeń
technicznych w języku perskim.

Praca magisterska napisana
w Instytucie Językoznawstwa
Uniwersytetu im. Adama Mickiewicza
pod kierunkiem prof. dr hab. Piotra Wierzchonia

Poznań, 2011

ROZDZIAŁ 4: METODOLOGIA BADAŃ

4.1. Cel badawczy

Podstawowym celem badawczym niniejszej pracy jest porównanie wyników badań nad angielskimi zapożyczeniami w perskim. W roku 2009 autorka przeprowadziła leksykograficzno-korpusowe badanie wybranej grupy zapożyczeń oraz ich perskich odpowiedników. Analizowaną grupą jest słownictwo techniczne. Korpusem użytym w badaniu w 2009 roku był Persian Linguistic Database (PLDB) –korpus języka perskiego. Ciekawość naukowa skłoniła autorkę do ponownego porównania danych par jednostek leksykalnych, tym razem jednak na korpusie przygotowanym przez zainteresowaną. Tematem tego rozdziału jest przedstawienie zastosowanej metodologii w omawianym badaniu.

4.2. Dane

Analizowane dane stanowią 10 par wyrazów. Każda para składa się z zapożyczenia angielskiego oraz jego perskiego odpowiednika. Lista ta została stworzona za pomocą słownika *Dictionnaire des Mots Européens en Persian*, autorstwa M. Moshiri oraz listy słów zaproponowanych przez Perską Akademię Językową - *A collection of Terms Approved by The Academy of Persian Language and Literature*. Najpierw przeanalizowano słownik w celu przygotowania listy wszystkich zapożyczeń angielskich. Następnie sprawdzono, które zapożyczenia mają swoje odpowiedniki zaproponowane przez Akademię Językową. W trakcie przygotowywania listy zapożyczeń angielskich, okazało się, iż tworzą one pewne grupy semantyczne: technologia, edukacja, urządzenia kuchenne, sport, żywność, pojazdy i urządzenia samochodowe, medycyna oraz tabu. Podjęto decyzję o przeanalizowaniu tylko jednej grupy, a mianowicie jednostek leksykalnych należących do technicznej grupy

semantycznej. Wyrazy, które zostały poddane analizie, to: *computer* „computer”, *data* „dane”, *digital* „cyfrowy”, *file* „plik”, *user* „użytkownik”, *fax* „faks”, *freezer* „zamrażarka”, *printer* „drukarka”, *site* „strona (internetowa)”, *lens* „soczewki” oraz ich perskie odpowiedniki.

Zarówno pełną listę zapożyczeń angielskich w języku perskim stworzoną na podstawie słownika *Dictionnaire des Mots Européens en Persian*, jak i listę angielskich zapożyczeń oraz ich perskich ekwiwalentów można znaleźć w załącznikach.

4.3. Materiały

4.3.1. PLDB

Persian Linguistic Database był korpusem wykorzystanym w badaniu w roku 2009. Jest to korpus przygotowany przez Sayyda Mostafa Assiego – dyrektora wydziału językoznawstwa w Instytucie Nauk Humanistycznych w Teheranie. Poniższy cytat pochodzi z oficjalnej strony korpusu:

To pierwsza baza danych współczesnego języka perskiego dostępna online, zaprojektowana i rozwinięta przez dr S.M. Assi z Instytutu Nauk Humanistycznych (IHCS) w Iranie. Baza zawiera ogromne ilości wyselekcjonowanego korpusu współczesnego języka perskiego w formie bieżących tekstów. Niektóre z tekstów zostały opatrzone adnotacjami dotyczącymi gramatyki, wymowy oraz haseł słownikowych. Specjalne i skuteczne oprogramowanie umożliwia różnego rodzaju analizy oraz pozwala na przeprowadzenie obliczeń statystycznych w całej bazie danych. Baza danych jest stale ulepszana i rozbudowywana.

[This is the first on-line database for the contemporary (Modern) Persian designed and developed by Dr. S. M. Assi at the Institute for Humanities and Cultural Studies (IHCS), Iran. The database contains huge selected corpora of all varieties of the Modern Persian language in the form of running texts. Some of the texts are annotated with grammatical, pronunciation and lemmatization tags. A special and powerful software provides different types of search and statistical listing facilities through the whole database or any selective corpus made up of a group of texts. The database is constantly improved and expanded (www.pldb.ihcs.ac.ir/).]

Korpus w chwili obecnej zawiera 50 milionów słowoform. W jego skład wchodzi:

- proza
- poezja
- artykuły z magazynów naukowych
- sztuki teatralne

- scenariusze filmowe
- literatura dziecięca
- prasa codzienna
- podręczniki
- przepisy prawne
- teksty administracyjne
- listy urzędowe
- napisy na murach
- nagrane a później przetranskrybowane konwersacje
- transkrybowane programy telewizyjne oraz radiowe (Assi 2000: 69).

4.3.2. Korpus autorki

Decydując się na skompilowanie własnego korpusu, autorka świadoma swoich ludzkich ograniczeń, zmuszona była do określenia granic korpusu. Korpus zawiera materiały prasowe z agencji informacyjnych, magazynów oraz gazet powstałych w okresie pomiędzy pierwszym stycznia a trzydziestym pierwszym grudnia 2010 roku. Wybór dziennikarskiego rejestru językowego spowodowany był faktem, iż wszystkie wyniki w badaniu z 2009 roku pochodziły właśnie z takiego rejestru. Rok 2010 został wybrany jako ograniczenie czasowe. Korpus prasowy autorki zawiera 8 mln słowoform. W trakcie kompilowania korpusu zapadła decyzja, aby stworzyć jeszcze jeden korpus – korpus techniczny. Autorka chciała sprawdzić, czy proporcje w użyciu zapożyczeń i ich perskich odpowiedników różnią się w korpusie bardziej ogólnym – prasowym – i korpusie specjalistycznym – technicznym. Materiały zawarte w drugim korpusie stanowią artykuły technologiczne, magazyny oraz naukowe czasopisma techniczne, jak również artykuły pojawiające się na blogach dotyczących technologii. Korpus posiada 360,000 słowoform i jest ograniczony czasowo: wszystkie zawarte w nim materiały pochodzą z 2010 roku.

4.4. Badanie porównawcze

Zasadniczym celem tego badania jest porównanie wyników:

1. PLDB oraz prasowego korpusu autorki, oraz
2. Prasowego oraz technicznego korpusu autorki

Ze względu na różnice w wielkości porównywanych korpusów, Autorka zdecydowała się nie porównywać tylko częstości wystąpień poszczególnych par lecz skorzystał z dostępnych narzędzi statystycznych w celu określenia, które różnice w danej proporcji można uznać za statystycznie istotne. Zastosowanym testem jest test-Z Karla Pearsona – prekursora statystyki matematycznej.

4.5. Test-Z

Test-Z jest testem dla dwóch proporcji. Znany jest również jako test wskaźników struktury. Wartość statystyki testowej obliczmy z wzoru:

$$Z = \frac{\left| \frac{O_{11}}{O_1} - \frac{O_{21}}{O_2} \right|}{\sqrt{\frac{O_1 O_0}{n} \frac{O_2 O_0}{n}}} \sqrt{\frac{O_1 O_2}{n}}$$

Rozkładem tej statystyki, przy prawdziwości hipotezy zerowej (przy braku istotnych różnic) jest (granicznie) rozkład normalny standaryzowany.

Na potrzeby badania został stworzony specjalny arkusz w programie Excel. Arkusz ten oblicza czy dla danej pary różnica proporcji jej użycia jest statystycznie istotna. Pod uwagę brane były tylko te wyniki, które wykazywały statystyczną istotność na poziomie co najmniej 95%. W statystyce powszechnie obowiązuje założenie poziomu istotności dla 95%. Oznacza to, że maksymalne prawdopodobieństwo popełnienia błędu wynosi 5%.

1														
2		n1	sample size 1	80	0,0125	0,025					p1	0,404		
3		n2	sample size 2	80	0,0125						p2	0,56		
4		p1%	percentage 1	40,4							x1	6400		
5		p2%	percentage 2	56,0							x2	160		
6											x3	0,156		
7			z-value	1,97454							x4	0,499675895		
8											p	0,482		
9			sig. at 90% level	yes							q	0,518		
10			sig. at 95% level	yes							t	1,974541177		
11			sig. at 99% level	no								0,249676		
12												0,079005696		
13												1,974541177		
14	sample	n1	n2											
15	sizes	80	80											
16														
17	percentages	p1%	p2%	z-value	sig. at 90% level	sig. at 95% level	sig. at 99% level							
18	1	40,4	56	-1,974541177	yes	yes				p1	40,4	56	48,2	51,8
19	2	70	60	1,325987088						p	70	60	65	35
20	3									q	0	0	0	100
21	4										0	0	0	100

4.6. Concordancer

Narzędziem użytym do analizy korpusów była najnowsza wersja programu dr Mike Scotta - WordSmith 5.0. Najważniejszymi elementami tego konkordancera są:

- lista słów (*WordList*) – lista najczęściej występujących jednostek leksykalnych,
- concord (*Concord*) – program generujący konkordancje,
- słowa kluczowe (*KeyWords*) – program służący do tworzenia listy słów kluczowych danego tekstu,
- statystyka (*Statistics*) – *WordSmith* umożliwia przeprowadzenie prostych kalkulacji statystycznych.

Przewaga programu *WordSmith* nad innymi konkordancerami wynika z faktu, że *WordSmith* jako jedyny program (ze znanych autorce concordancerów) umożliwił poprawne analizy dla języka perskiego w kodowaniu *Unicode* (UTF-8).

ROZDZIAŁ 5: BADANIE EMPIRYCZNE

Celem tego rozdziału jest przedstawienie wyników przeprowadzonego badania. Poniżej znajduje się zbiorcza tabelka z wynikami:

Tabela 9. Zbiorcze wyniki przeprowadzonych badań.

	analizowany leksem	2010 korpus prasowy	2010 korpus techniczny	2008 PLDB
1.	کامپیوتر [kāmpyuter]	115	277	111
	رایانه [rāyāne]	277	10	304
2.	دیجیتال [dijital]	0	0	0
	رقمی [raqmi]	209	11	65
3.	فایل [fāil]	64	98	46
	پروانجا [parvanjā]	0	0	0
4.	سایت [sāyt]	1598	516	229
	ایستگاه [istgāh]	656	11	119
5.	دینا [deytā]	91	8	3
	داده [dādeh]	7484	475	2605
6.	فریزر [frizer]	11	0	25
	یخ زن [yakh zan]	0	0	0
7.	پرینتر [printer]	0	5	0
	چاپگر [čāpgar]	28	5	9

8.	يوزر [yuzer]	0	0	1
	كاربر [kārbar]	731	618	65
9.	لنز [lenz]	40	19	15
	عدسی [adasi]	16	2	7
10.	فاکس [fāks]	33	4	62
	دورنگار [durnagār]	2	0	0

Każda para wyrazów jest szczegółowo omówiona w kolejnych podrozdziałach.

5.1. COMPUTER

Zapóżyczenie: کامپیوتر [kāmpyuter]

Perski odpowiednik: رایانه [rāyāne]

Data wprowadzenia ekiwalentu: marzec 2005

5.1.1. PLDB

کامپیوتر [kāmpyuter]: 111

رایانه [rāyāne]: 304

Data wyszukiwania: 3 kwiecień 2009

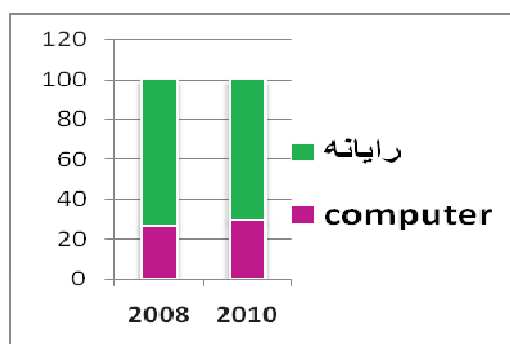
5.1.2. Korpus prasowy

کامپیوتر [kāmpyuter]: 115

رایانه [rāyāne]: 227

Data wyszukiwania: 15 luty 2011

5.1.3. PLDB vs. korpus prasowy



Z przedstawionego wykresu wynika, iż użycie angielskiego zapożyczenia nieznacznie wzrosło w przeciągu dwóch lat. Test-Z wykazał natomiast że powyższa różnica, przy wartości z równej 0,81887735, nie jest istotna statystycznie.

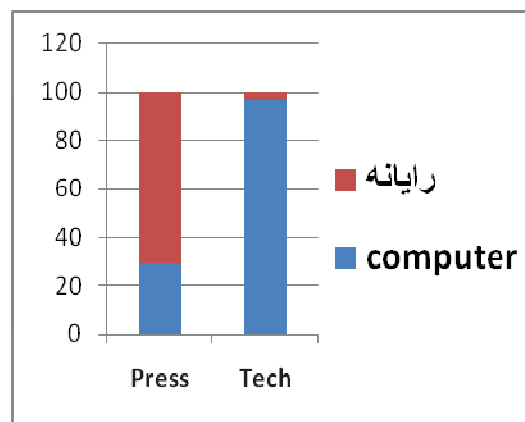
5.1.4. Korpus techniczny

کامپیوتر [kāmpyuter]: 227

رایانه [rāyāne]: 10

Data wyszukiwania: 27 luty 2011

5.1.5. Korpus prasowy vs. korpus techniczny



Porównując wyniki korpusu technicznego oraz prasowego, można zaobserwować znacznie większe użycie zapożyczenia w stosunku do perskiego odpowiednika. Obserwacje z wykresu potwierdza test-Z, wykazując, iż przy wartości z wynoszącej 17,50523981, różnica użycia danych jednostek jest statystycznie istotna na poziomie aż 99%.

5.2. DIGITAL

Zapożyczenie: دیجیتل [dijital]

Perski odpowiednik: رقمی [raqmi]

Data wprowadzenia odpowiednika: marzec 2005

5.2.1. PLDB

ديجيتل [dijital]: 0

رقمی [raqmi]: 65

Data wyszukiwania: 20 kwiecień 2009

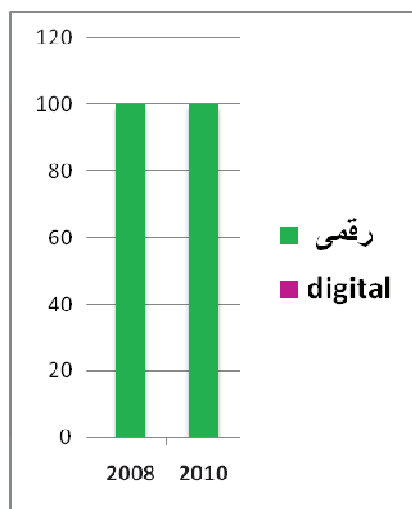
5.2.2. Korpus prasowy

ديجيتل [dijital]: 0

رقمی [raqmi]: 209

Data wyszukiwania: 15 luty 2011

5.2.3. PLDB vs korpus prasowy



Powyższy wykres pokazuje przewagę perskiego ekwiwalentu nad zapożyczeniem angielskim. Z uwagi na fakt, iż nie odnotowano występowania wyrazu angielskiego w obu korpusach, przeprowadzenie testu nie było konieczne: nie mamy tutaj do czynienia z proporcją a więc nie ma podstaw do przeprowadzenia analizy metodą testu-Z.

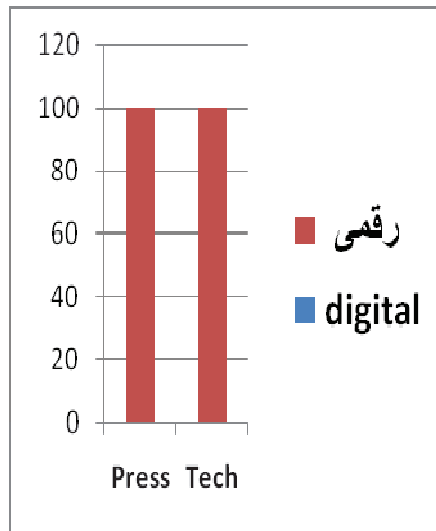
5.2.4. Korpus techniczny

ديجيتل [dijital]: 0

رقمی [raqmi]: 11

Data wyszukiwania: 27 luty 2011

5.2.5. Korpus prasowy vs korpus techniczny



Podobnie jak w poprzednim przypadku, również porównanie korpusu prasowego z technicznym wykazało jednoznaczną przewagę perskiego ekwiwalentu. Z uwagi na fakt, iż nie odnotowano występowania wyrazu angielskiego w obu korpusach przeprowadzenie testu nie było konieczne: nie mamy tutaj do czynienia z proporcją, a więc nie ma podstaw do przeprowadzenia analizy metodą testu-Z.

5.3. FILE

Zapożyczenie: فایل [fāil]

Perski ekwiwalent: پرونجا [parvanjā]

Data wprowadzenia ekwiwalentu: marzec 2005

5.3.1. PLDB

فایل [fāil]: 46

پرونجا [parvanjā]: 0

Data wyszukiwania: 11 kwiecień 2009

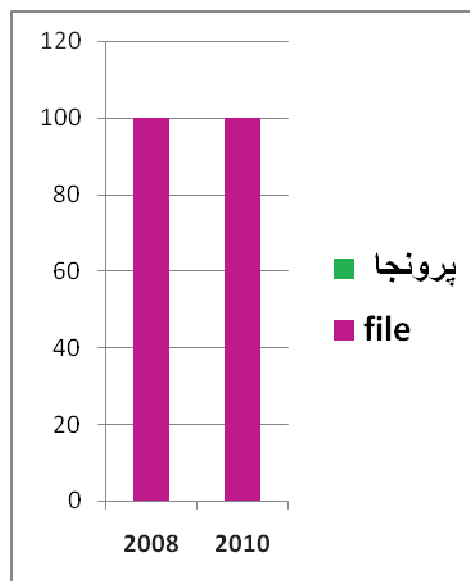
5.3.2. Korpus prasowy

فایل [fāil]: 64

پرونجا [parvanjā]: 0

Data wyszukiwania: 15 luty 2011

5.3.3. PLDB vs Korpus prasowy



Powyższy wykres pokazuje przewagę zapożyczenia angielskiego nad odpowiednikiem perskim. Z uwagi na fakt, iż nie odnotowano występowania wyrazu perskiego w obu korpusach przeprowadzenie testu nie było konieczne: nie mamy tutaj do czynienia z proporcją, a więc nie ma podstaw do przeprowadzenia analizy metodą testu-Z.

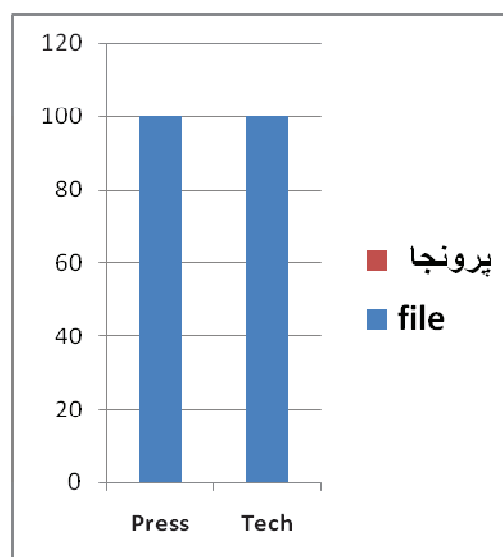
5.3.4. Korpus techniczny

فایل [fāil]: 98

پرونگا [parvanjā]: 0

Daty wyszukiwania: 27 luty 2011

5.3.5. Korpus prasowy vs korpus techniczny



Podobnie jak w poprzednim przypadku, również porównanie korpusu prasowego z technicznym wykazało jednoznaczną przewagę zapożyczenia z języka angielskiego. Z uwagi na fakt, iż nie odnotowano występowania perskiego odpowiednika w obu korpusach, przeprowadzenie testu statystycznego nie było konieczne: nie mamy tutaj do czynienia z proporcją, a więc nie ma podstaw do przeprowadzenia analizy metodą testu-Z.

5.4. SITE

Zapożyczenie: سایت [sāyt]

Perski ekwiwalent: ایستگاه [istgāh]

Data wprowadzenia ekwiwalentu: marzec 2005

5.4.1. PLDB

سایت [sāyt]: 229

ایستگاه [istgāh]: 119

Data wyszukiwania: 10 kwiecień 2009

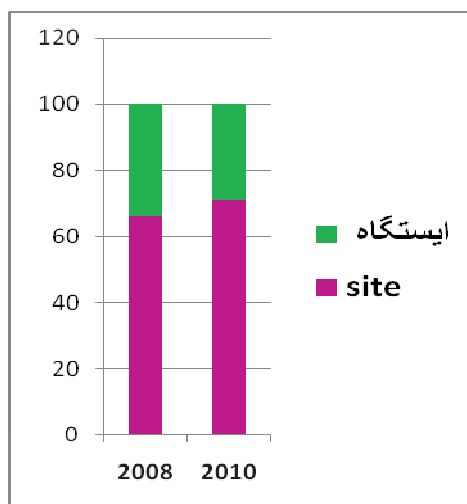
5.4.2. Korpus prasowy

سایت [sāyt]: 1598

ایستگاه [istgāh]: 656

Data wyszukiwania: 15 luty 2011

5.4.3. PLDB vs korpus prasowy



Analizując wyniki z korpusu prasowego i PLDB, test-Z wykazuje, że przy wartości z wynoszącej 1,93309526 różnica w użyciu obu jednostek wykazuje statystyczną istotność na poziomie 90%. Przy założeniu, że za statystycznie istotne uznajemy wyniki na poziomie przynajmniej 95% istotności, powyższa różnica nie może zostać uznana za statystycznie istotną.

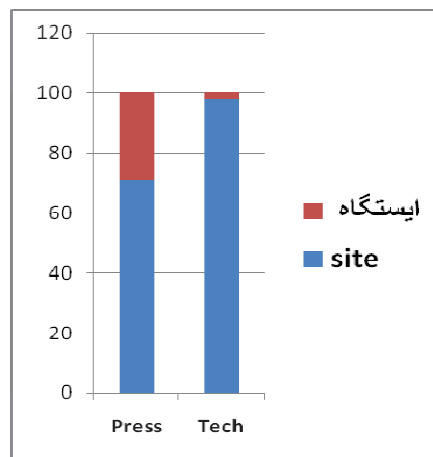
5.4.4. Korpus techniczny

سایت [sāyt]: 516

ایستگاه [istgāh]: 11

Data wyszukiwania: 27 luty 2011

5.4.5. Korpus prasowy vs korpus techniczny



Porównując wyniki korpusu technicznego oraz prasowego, można zaobserwować znacznie większe użycie zapożyczenia w stosunku do perskiego odpowiednika. Obserwacje z wykresu potwierdza test-Z, wykazując, iż przy wartości z wynoszącej 13,07666203, różnica użycia danych jednostek jest statystycznie istotna na poziomie aż 99%.

5.5. FREEZER

Zapożyczenie: فریزر [frizer]

Perski ekwiwalent: یخ زن [yakh zan]

Data wprowadzenia odpowiednika: marzec 2005

5.5.1. PLDB

فریزر [frizer]: 25

یخ زن [yakh zan]: 0

Data wyszukiwania: 11 kwiecień 2009

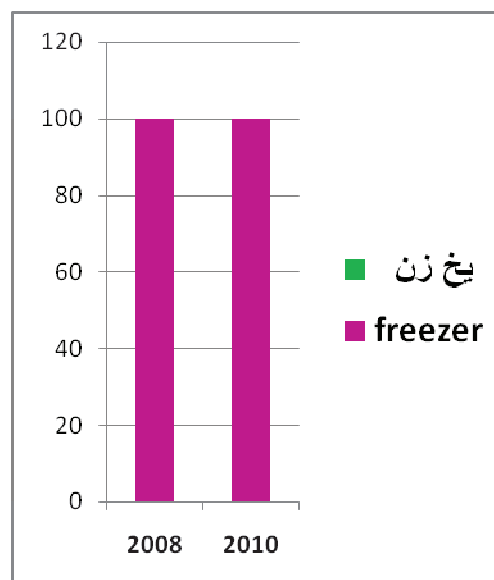
5.5.2. Korpus prasowy

فریزر [frizer]: 11

یخ زن [yakh zan]: 0

Data wyszukiwania: 15 luty 2011

5.5.3. PLDB vs korpus prasowy



Powyższy wykres pokazuje przewagę zapożyczenia angielskiego nad odpowiednikiem perskim. Z uwagi na fakt, iż nie odnotowano występowania wyrazu perskiego w obu korpusach, przeprowadzenie testu nie było konieczne: nie mamy tutaj do czynienia z proporcją a więc nie ma podstaw do przeprowadzenia analizy metodą testu-Z.

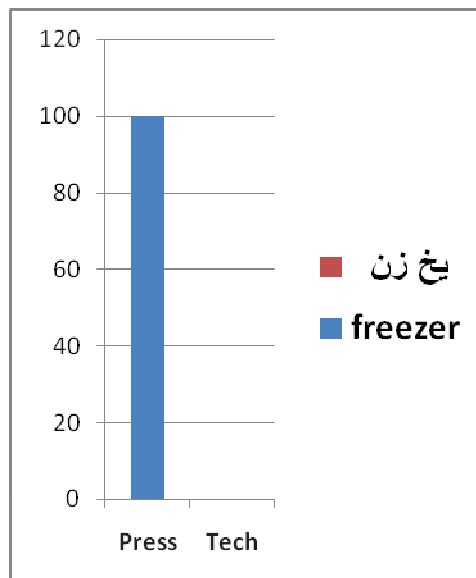
5.5.4. Korpus techniczny

فریزر [frizer]: 0

یخ زن [yakh zan]: 0

Data wyszukiwania: 27 luty 2011

5.5.5. Korpus prasowy vs korpus techniczny



W korpusie technicznym nie odnotowano występowania ani zapożyczenia ani jego perskiego odpowiednika. Z uwagi na ten fakt, przeprowadzenie testu nie było konieczne: nie mamy tutaj do czynienia z proporcją, a więc nie ma podstaw do przeprowadzenia analizy metodą testu-Z.

5.6. DATA

Zapożyczenie: دیتا [deytā]

Perski ekwiwalent: داده [dādeh]

Data wprowadzenia ekwiwalentu: kwiecień 2005

5.6.1. PLDB

دیتا [deytā]: 3

داده [dādeh]: 2605

Data wyszukiwania: 11 kwiecień 2009

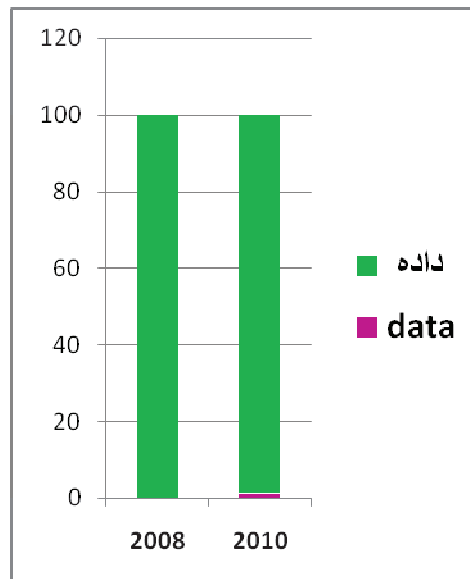
5.6.2. Korpus prasowy

دیتا [deytā]: 91

داده [dādeh]: 7484

Data wyszukiwania: 15 luty 2011

5.6.3. PLDB vs. korpus prasowy



Porównując wyniki korpusu prasowego oraz PLDB można zaobserwować znacznie większe użycie perskiego odpowiednika w stosunku do angielskiego zapożyczenia. W 2010 roku obserwujemy wzrost użycia zapożyczenia. Test Z-wykazał, że różnica w użyciu jest statystycznie istotna i dla wartości z równej 5,003115866 osiąga 99% poziom istotności.

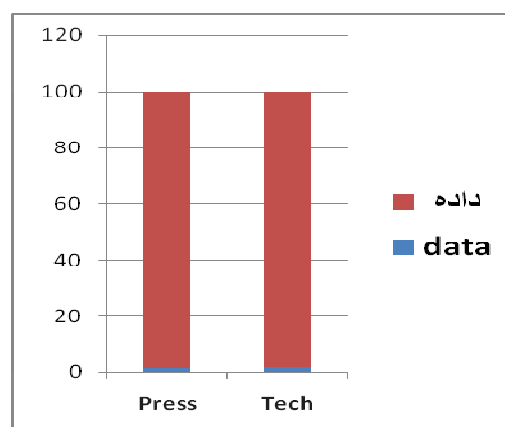
5.6.4. Korpus techniczny

ديتا [deytā]: 8

داده [dādeh]: 475

Data wyszukiwania: 27 luty 2011

5.6.5. Korpus prasowy vs. korpus techniczny



Analizując wyniki z korpusu prasowego i technicznego, obserwujemy znaczną przewagę perskiego odpowiednika nad angielskim zapożyczeniem. Test-Z wykazuje, że przy wartości z wynoszącej 0,880111868 różnica w użyciu obu jednostek nie jest statystycznie istotna.

5.7. PRINTER

Zapożyczenie: پرينتر [printer]

Perski ekwiwalent: چاپگر [čāpgar]

Data wprowadzenia ekwiwalentu: marzec 2005

5.7.1. PLDB

پرينتر [printer]: 0

چاپگر [čāpgar]: 9

Data wyszukiwania: 10 kwiecień 2009

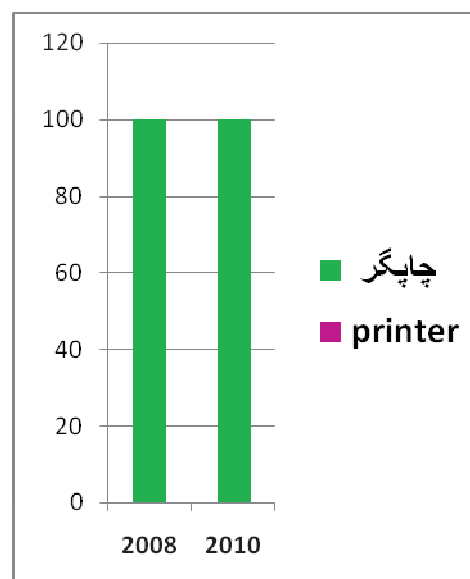
5.7.2. Korpus prasowy

پرينتر [printer]: 0

چاپگر [čāpgar]: 28

Data wyszukiwania: 15 luty 2011

5.7.3. PLDB vs korpus prasowy



Powyższy wykres pokazuje jednoznaczną przewagę perskiej jednostki. Z uwagi na fakt, iż nie odnotowano występowania wyrazu angielskiego w obu korpusach, przeprowadzenie testu nie było konieczne: nie mamy tutaj do czynienia z proporcją, a więc nie ma podstaw do przeprowadzenia analizy metodą testu-Z.

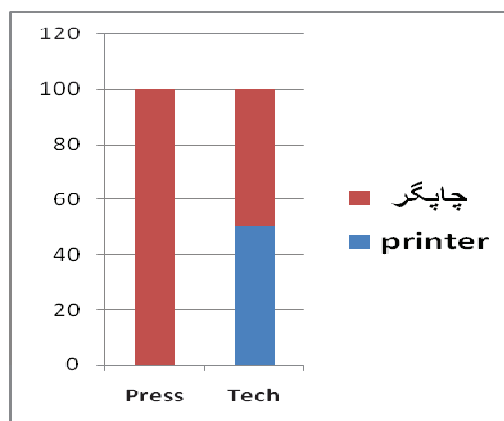
5.7.4. Korpus techniczny

پرینتر [printer]: 5

چاپگر [čāpgar]: 5

Data wyszukiwania: 27 luty 2011

5.7.5. Korpus prasowy vs korpus techniczny



Porównując wyniki korpusu technicznego oraz prasowego, można zaobserwować znaczną różnicę w użyciu zapożyczenia. Obserwację z wykresu potwierdza test-Z, wykazując, iż przy wartości z wynoszącej 4,015122927, różnica użycia danych jednostek jest statystycznie istotna na poziomie aż 99%.

5.8. USER

Zapożyczenie: یوزر [yuzer]

Perski ekwiwalent: کاربر [kārbar]

Data wprowadzenia ekwiwalentu: kwiecień 2005

5.8.1. PLDB

یوزر [yuzer]: 1

کاربر [kārbar]: 65

Data wyszukiwania: 3 kwiecień 2009

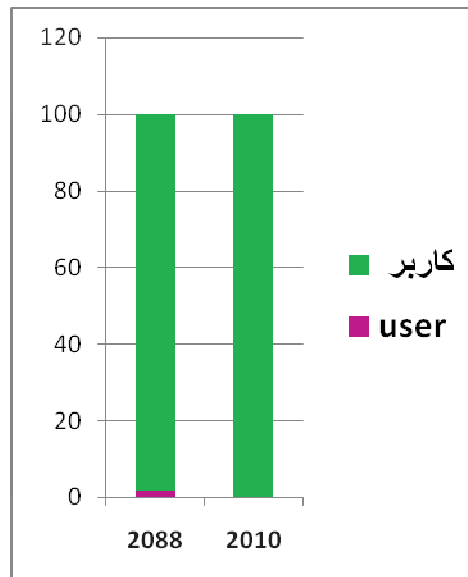
5.8.2. Korpus prasowy

يوزر [yuzer]: 0

کاربر [kārbar]: 731

Data wyszukiwania: 15 luty 2011

5.8.3. PLDB vs korpus prasowy



Porównując wyniki korpusu prasowego oraz PLDB, można zaobserwować znacznie większe użycie perskiego odpowiednika w stosunku do angielskiego zapożyczenia. Perskiego leksem wyparł zupełnie angielskie. Test-Z wykazał, że różnica w użyciu jest statystycznie istotna i dla wartości z równej 3,33011589 osiąga 99% poziom istotności.

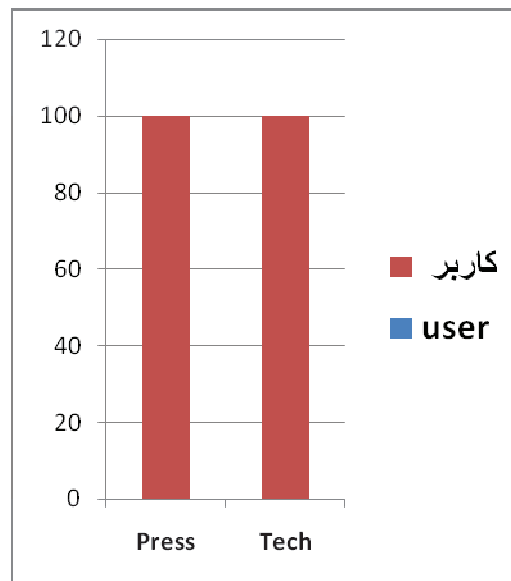
5.8.4. Korpus techniczny

يوزر [yuzer]: 0

کاربر [kārbar]: 618

Data wyszukiwania: 27 luty 2011

5.8.5. Korpus prasowy vs korpus techniczny



Powyższy wykres ilustruje jednoznaczną przewagę perskiej jednostki leksykalnej. Z uwagi na fakt, iż nie odnotowano występowania wyrazu angielskiego w obu korpusach, przeprowadzenie testu nie było konieczne: nie mamy tutaj do czynienia z proporcją, a więc nie ma podstaw do przeprowadzenia analizy metodą testu-Z.

5.9. LENS

Zapóżylenie: لنز [lenz]

Perski ekwiwalent: عدسی [adasi]

Data wprowadzenia ekwiwalentu: marzec 2005

5.9.1. PLDB

لنز [lenz]: 15

عدسی [adasi]: 7

Data wyszukiwania: 3 kwiecień 2009

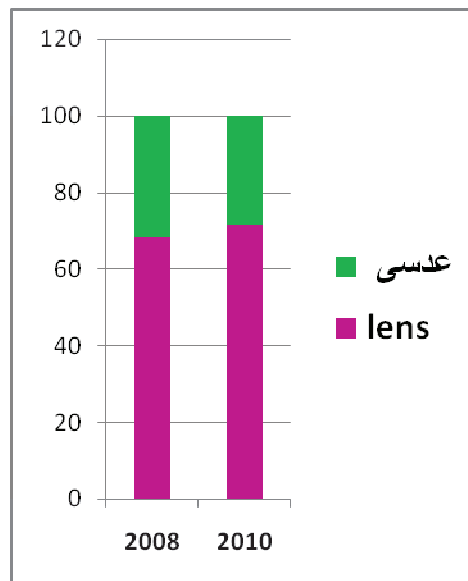
5.9.2. Korpus prasowy

لنز [lenz]: 40

عدسی [adasi]: 16

Data wyszukiwania: 15 luty 2011

5.9.3. PLDB vs korpus prasowy



Porównanie wyników z PLDB oraz korpusu prasowego obserwujemy przewagę zapożyczenia nad jego perskim odpowiednikiem. Test Z nie wykazuje, aby powyższe różnice użycia były istotne statystycznie, przy wartości z wynoszącej 0,282980608.

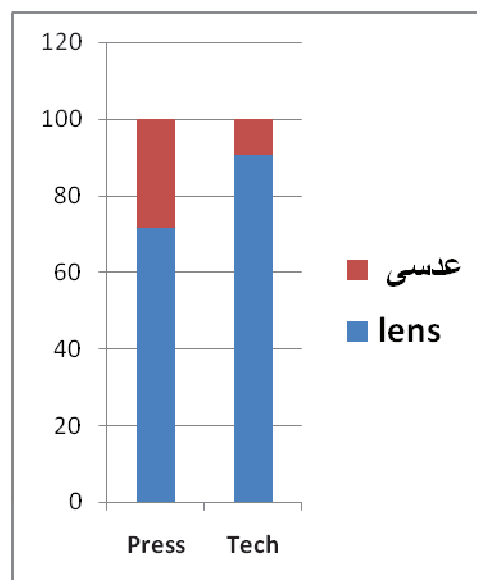
5.9.4. Korpus techniczny

لنز [lens]: 19

عدسی [adasi]: 2

Data wyszukiwania: 27 luty 2011

5.9.5. Korpus prasowy vs korpus techniczny



Powyższy wykres ilustruje, że w rejestrze technicznym częściej używane jest zapożyczenie. Różnice w użyciu pomiędzy oboma korpusami wykazują statystyczną istotność na poziomie 90%, przy wartości z równej 1,758844567. Przy założeniu, że za statystycznie istotne uznajemy wyniki na poziomie przynajmniej 95% istotności, powyższa różnica nie może zostać uznana za statystycznie istotną.

5.10. FAX

Zapożyczenie: فاكس [fāks]

Perski ekwiwalent: دورنگار [durnagār]

Data wprowadzenia ekwiwalentu: marzec 2005

5.10.1. PLDB

فاکس [fāks]: 62

دورنگار [durnagār]: 0

Data wyszukiwania: 3 kwiecień 2009

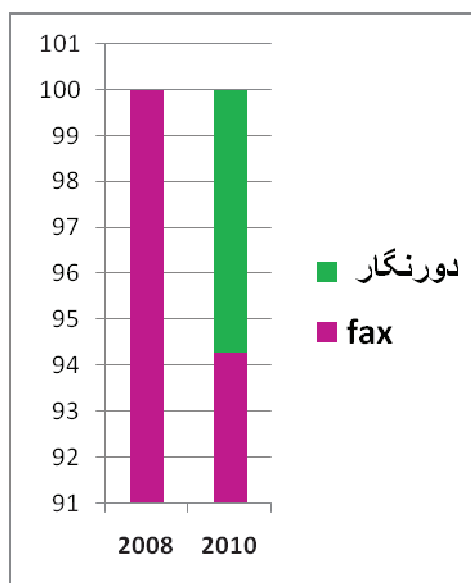
5.10.2. Korpus prasowy

فاکس [fāks]: 33

دورنگار [durnagār]: 2

Data wyszukiwania: 15 luty 2011

5.10.3. PLDB vs korpus prasowy



Na powyższym wykresie obserwujemy znaczny wzrost użycia perskiego odpowiednika. Różnice w użyciu pomiędzy oboma korpusami wykazują statystyczną istotność na poziomie 90%, przy wartości z równej 1,901957836. Przy założeniu, że za statystycznie istotne uznajemy wyniki na poziomie przynajmniej 95% istotności, powyższa różnica nie może zostać uznana za statystycznie istotną.

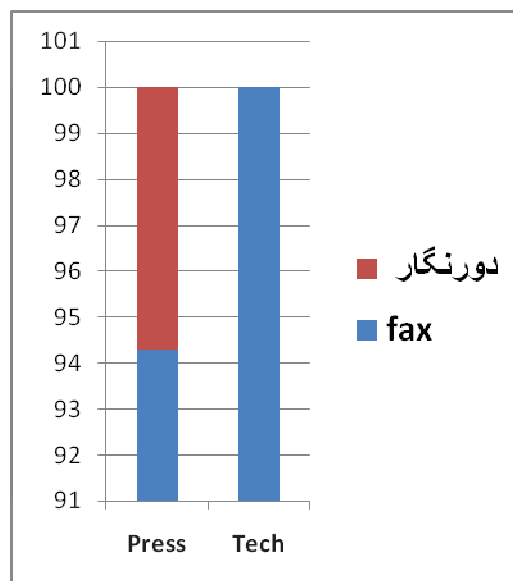
5.10.4. Korpus techniczny

فاکس [fāks]: 4

دورنگار [durnagār]: 0

Data wyszukiwania: 27 luty 2011

5.10.5. Korpus prasowy vs korpus techniczny



Powyższy wykres ilustruje, że w rejestrze technicznym częściej używane jest zapożyczenie angielskie. Różnice w użyciu pomiędzy oboma korpusami wykazują statystyczną istotność na poziomie 90%, przy wartości Z równej 0,490842786. Przy założeniu, że za statystycznie istotne uznajemy wyniki na poziomie przynajmniej 95% istotności, powyższa różnica nie może zostać uznana za statystycznie istotną.

5.11. Podsumowanie

Wyniki z korpusu prasowego z 2010 roku pokazują, że aż w 6 przypadkach perska jednostka leksykalna jest częściej używana niż angielskie zapożyczenie. Są to wyrazy: رابانه „komputer”, ىرقم „cyfrowy”, داده „dane”, چاپگر „drukarka”, كاربر „użytkownik” oraz دورنگار „fax”. Porównując to z badaniem przeprowadzonym w 2008 roku jest to o jeden wyraz więcej. Warto zauważyć, że wyrazem, który w 2008 roku nie występował w ogóle, a obecnie jest częściej używany niż angielskie zapożyczenie, jest perski wyraz دورنگار „fax”. Analiza statystyczna wykazała jednak, iż tylko w dwóch przypadkach różnice w użyciu danych par wyrazów są statystycznie istotne. Te dwa przypadki to داده „dane” i كاربر „użytkownik”. W obu mamy do czynienia z przewagą perskich odpowiedników.

Korpus techniczny pokazuje natomiast, że w 5 przypadkach większą popularnością cieszy się angielskie zapożyczenie: *computer* „komputer”, *file* „plik”, *site* „strona”, *lens* „soczewki” oraz *fax* „faks”. Wyraz *printer* „drukarka” oraz jego perski odpowiednik mają taki sam stopień użycia. Nie odnotowano natomiast wystąpień wyrazów oznaczających „zamrażarkę”. W związku z tym jednoznaczną przewagą perski ekwiwalent posiada tylko w 3 przypadkach: ىرقم „cyfrowy”, داده „dane”, كاربر „użytkownik”. Analiza statystyczna wykazała, że różnica w proporcji użyć konkretnych par wyrazów w korpusie technicznym oraz prasowym jest statystycznie istotna w 3 przypadkach, a mianowicie: *computer* „komputer”, *site* „strona” oraz *printer* „drukarka”.

Powyższe wyniki pokazują, że częstość użycia zaproponowanych przez Akademię Językową wyrazów różni się w obu analizowanych rejestrach językowych. W rejestrze prasowym obserwujemy rosnące użycie perskich jednostek. Istotny wpływ na taką tendencję ma fakt, iż propozycje Akademii są obowiązkowe dla prasy. Z drugiej jednak strony, świadczyć to może o dosyć wolnym dostosowywaniu się do wprowadzonych zmian. Wszystkie analizowane wyrazy zostały zaproponowane przez Akademię w 2005 roku a warto podkreślić, że Akademia wprowadza zazwyczaj 3 – 5 letni okres, w czasie którego zaproponowana jednostka leksykalna ma zastąpić zapożyczenie.

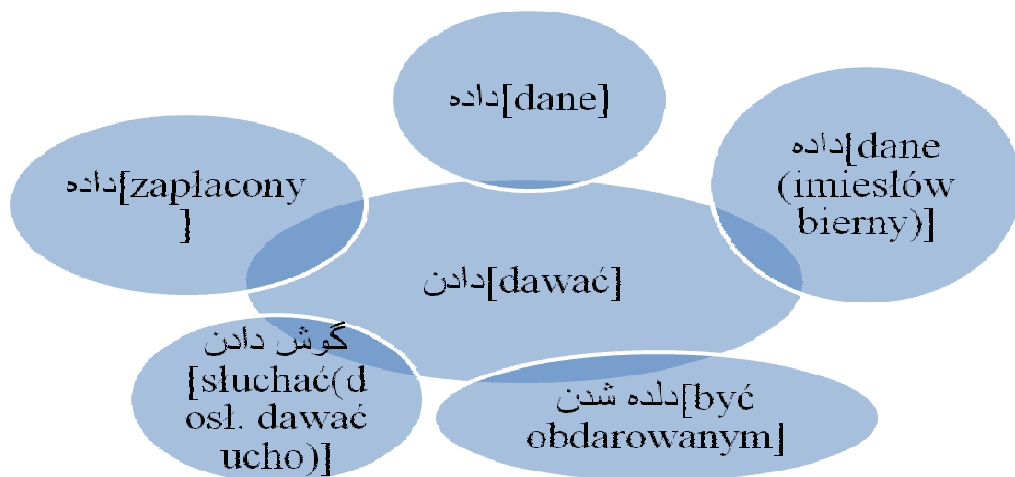
Przewaga perskich wyrazów może wynikać również z faktu, iż większość z nich w momencie pojawienia się na liście *Collection of Terms Approved* to były istniejące jednostki, które rozszerzyły swoje znaczenie. Tak jest w przypadku wyrazów: رابانه

„komputer”, عرقم „cyfrowy” oraz داده „dane”. W przypadku natomiast, kiedy jako propozycję podano nowo utworzone perskie słowo obserwujemy tendencję, w której to zapożyczenie jest częściej używane, np. *freezer* „zamrażarka” czy *file* „plik”.

Kolejną obserwacją wynikającą z przeprowadzonego badania jest stwierdzenie, że te zapożyczenia, które dopasowały się do fonetycznego i morfologicznego systemu języka perskiego wydają się być trudniejsze do zastąpienia. Widać to na przykładzie wyrazów *file* „plik” oraz *lens* „soczewki”.

Należy również zwrócić uwagę na problem z interpretacją niektórych danych związanych z występowaniem homonimii i polisemii. Fakt, że wyraz perski داده „dane” może posiadać więcej znaczeń niż to analizowane w pracy, powoduje iż wyniki dotyczące tej jednostki leksykalnej są niejednoznaczne. Wyraz ten może oznaczać „mieć coś dane” oraz „zapłacone”. Jest to również imiesłów przymiotnikowy bierny od czasownika دادن „dawać”. Warto w tym miejscu zauważyć, że w języku polskim wyraz *dane* również jest taki sam jak imiesłów bierny czasownika *dawać*.

Wreszcie analiza wybranych 10 par wyrazów prowadzi do pytania: czy w przypadku, kiedy słowo perskie poszerzyło swoje znaczenie, mamy do czynienia z rozszerzeniem znaczenia, czy raczej z procesem zapożyczania semnacyjnego? W tym drugim przypadku bowiem, we wszystkich tego rodzaju przypadkach mówilibyśmy o przewadze zapożyczenia nad jego rodzimym odpowiednikiem. Jednym ze sposobów sprawdzenia tego stanu mogła by być analiza sieci radialnych (Lakoff 1987: 116). Sieć radialna ma za zadanie pokazać połączenia pomiędzy różnymi znaczeniami tej samej formy. Składa się ona z elementu centralnego oraz elementów peryferyjnych. Dla perskiego wyrazu داده „dane” sieć radialna mogłaby przedstawiać się następująco:



Jednostką centralną jest w tym przypadku czasownik دادن „dawać”. Wykorzystując sieć radialną dla ustalenia, czy dane znaczenie jest zapożyczeniem semantycznym, należałoby ustalić czy jest ono powiązane semantycznie ze znaczeniem centralnym. W pierwszym przypadku mielibyśmy do czynienia z rozszerzeniem semantycznym. W przeciwnym natomiast z zapożyczeniem semantycznym. Wydaje się, że w przypadku rzeczownika داده „dane” mamy do czynienia z rozszerzeniem semantycznym.

ZAKOŃCZENIE

Głównymi celami niniejszej pracy była analiza angielskich zapożyczeń w języku perskim oraz sprawdzenie czy irańska polityka językowa wobec zapożyczeń z technicznego pola semantycznego jest skuteczna.

Cel pierwszy został osiągnięty: została sporządzona lista anglicyzmów we współczesnej perszczyźnie. Jednostki leksykalne zostały sklasyfikowane według pól semantycznych oraz Haugenowskiego podziału zapożyczeń językowych.

Drugi cel – analiza i ocena skuteczności działania irańskiej Akademii Językowej również został zrealizowany. Przeprowadzone badanie pokazało, iż perskie wyrazy zaproponowane przez Akademię Językową coraz częściej wypierają angielskie pożyczki. Punktem wyjścia w badaniu korpusowym były wyniki badania przeprowadzonego w 2008 roku. Porównanie tych wyników z badaniem przeprowadzonym w 2010 roku, przy zastosowaniu określonych narzędzi statystycznych, wskazało na pewne tendencje w użyciu zapożyczeń oraz ich perskich odpowiedników przygotowanych i zatwierdzonych przez irańską Akademię Językową. Tendencje te dotyczą przede wszystkim różnic w użyciu analizowanych jednostek w korpusie prasowym i technicznym.

Pomimo, iż praca ta nie wyczerpuje tematu zapożyczeń angielskich w języku perskim, zawiera ona określone wnioski, które mogą posłużyć analizie zapożyczeń z pozostałych pól semantycznych. Jednym z najważniejszych wniosków płynących z tej pracy jest zbadanie czy w przypadku kiedy mamy do czynienia z sytuacją, w której istniejąca jednostka leksykalna zapożyczyła dodatkowe znaczenie, to czy zjawisko to będzie zapożyczeniem semantycznym czy może rozszerzeniem semantycznym. Co więcej, stworzone na potrzeby tego badania korpusy mogą być rozbudowywane i posłużyć do monitorowania zmian zachodzący w proporcji użycia analizowanych jednostek leksykalnych.