

UOT 004.89

Sadıqov İ.C.

AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

ismayil.sadigov@gmail.com

AZƏRBAYCAN DİLİ MƏTNLƏRİNİN MÜRƏKKƏBLİYİNİN QIYMƏTLƏNDİRİLMƏSİ ÜÇÜN MODİFİKASIYA OLUNMUŞ FLEŞ DÜSTURU

Məqalədə ingilis dilindəki mətnlər üçün Fleşin oxuma asanlıığı düsturunun Azərbaycan dilindəki mətnlərə tətbiqi probleminə toxunulur. Bu problemin həlli üçün düsturdakı orta cümlə uzunluğu və orta söz uzunluğu kimi parametrlərə uyğun əmsalların korrektəsi metodikasından danışılır. İngilis və Azərbaycan dillərində olan identik mətnlər (bədi əsər nümunələri, akademik (elmi) mətnlər, ayrı-ayrı cümlələr və onların hərfi tərcüməsi) verilmiş metodika əsasında tədqiq edilir və nəticədə ingilis dillindəki mətnlər üçün olan Fleş düsturundakı əmsallar Azərbaycan dilindəki mətnlər üçün uyğunlaşdırılır.

Açar sözlər: mətnin mürəkkəbliyi, oxunabilirlik düsturları, Fleşin oxuma asanlıığı düsturu.

Giriş

Mətnlərin mürəkkəbliyi ilə bağlı elmi tədqiqatlara XIX əsrin sonlarında ABŞ-da başlanılmışdır. Maraqlıdır ki, görkəmli Azərbaycan maarifçisi Abbasqulu ağa Bakıxanov hələ həmin əsrin birinci yarısında bu məsələyə diqqət yönəltmişdir. O, 1836-cı ildə “Kitabi-nəsihət” (“Nəsihətnamə”) adlı pedaqoji əsərinin müqəddiməsində tərbiyə və təlimə dair kitabların çətin dildə olmasından şikayətlənərək yazırdı: “*Mən elə bir dərslük görmədim ki, öz dil və məzmununa görə asan olsun. Əldə olan kitablar o qədər çətin və qəliz dildə yazılmışdır ki, uşaqlar hələ bir yana dursun, onların mənasını müəllimlərin özləri də başa düşmürlər, bəzən cümlələr elə uzundur ki, onu mənimsəmək və anlamaq çox çətin olur*” [1, 2].

Göründüyü kimi, A.Bakıxanov mətnlərin mürəkkəbliyini cümlələrin həddən artıq uzun olması və mürəkkəb mənalı (çətin və qəliz) sözlərin çoxluğu ilə əlaqələndirirdi. Maraqlıdır ki, 1920-ci illərdə ABŞ-da tədqiqatçı pedaqoqlar mətnin mürəkkəblik səviyyəsini qabaqcadan müəyyənləşdirmək üçün sözlərin çətinliyindən və cümlələrin uzunluğundan istifadə etmək yolunu aşkar etdilər [3].

1. "Oxunabilirlik" anlayışı

Mətnin mürəkkəblik səviyyəsini, onun qavranılma dərəcəsini göstərmək üçün "oxunabilirlik" (*ing.readability*) terminindən istifadə olunur. Rus dilində “*читабельность*”, türk dilində “*okunabilirlik*” kimi işlədilən bu söz O.Türksevər-Musayevin rəhbərliyi ilə hazırlanmış “İngiliscə-Azərbaycanca Lüğət”də (İngiliscə-Azərbaycanca Lüğət, Bakı: “Qismət” nəşriyyatı, 2003, 1696 səh.) “*asan oxuna bilmə*”, “*oxunuşu asan olma*” kimi göstərilib. Əlbəttə, bu tərcümələrin heç biri termin kimi işlədilə bilməz. Nəzərdə tutulan xüsusiyyətlərə malik mətnləri məhz “*oxuna bilir*” ifadəsi daha dəqiq xarakterizə etdiyindən bundan sonra biz də “*readability*” terminini dilimizdə “*oxunabilirlik*” kimi işlədəcəyik.

Oxunabilirlik nədir? Oxunabilirlik bəzi mətnləri başqalarına nisbətən daha asan oxunan edir. Bu anlayışı çox zaman xəttə, şriftə, mətnin düzəninə aid işlədilən “*aydınlıq, səlislik*” (*ing. legibility*) anlayışı ilə qarışdırırlar. Oxunabilirliyi “*oxunaqlılıq*” ilə də eyniləşdirmək olmaz, çünki dilimizdə “*oxunaqlı*” anlayışı mətnin məzmunu ilə bağlı olub, “*həvəslə oxunan*”, “*maraqlı*” anlamında işlədilir [2].

Corc Kler (*ing. George Klare*) oxunabilirliyə “*asan başa düşülən, yaxud yazı üslubu sayəsində anlaşılan*” kimi tərif verir [4]. Bu tərif diqqəti mətnin məzmununa, məntiqi ardıcılığına, quruluşuna deyil, yazı üslubuna cəmləşdirir. Qrethen Hareis (*ing. Gretchen Hargis*) və onun IBM şirkətindəki həmkarları da bənzər fikirdədirlər: “*sözlərin və cümlələrin asan oxunması*”, yəni oxunabilirlik, mətnin aydınlığının əlamətidir [5].

SMOG oxunabilirlik düsturunun yaradıcısı Q.Harri MakLaflin (*ing.G.Harry McLaughlin*) oxunabilirliyə “*verilmiş sinfə aid insanların müəyyən oxu materialını nə dərəcədə inandırıcı və anlaşıqlı*” hesab etməsi kimi tərif verir [6]. Bu tərif mətn ilə oxu bacarığı, ilkin biliklər və motivasiya kimi məlum xarakteristikalara malik oxucular sinfi arasındakı qarşılıqlı əlaqəni xüsusi qeyd edir.

Ola bilsin, oxunabilirliyin ən dolğun tərifini Edqar Deyl (*ing.Edgar Dale*) və Janna Çell (*ing.Jeanne Chall*) vermişlər: oxunabilirlik “*çap materialının verilməmiş hissəsində oxucular qrupunun uğuruna təsir edən elementlərin hamısının (bütün qarşılıqlı təsirlər daxil olmaqla) ümumi cəmidir. Uğur dedikdə, oxucuların mətni nə dərəcədə başa düşməsi, optimal sürətlə oxuması və maraqlı hesab etməsi nəzərdə tutulur*” [7].

2. Oxunabilirlik düsturları

Oxunabilirliyin qiymətləndirilməsi üçün bir çox riyazi modellər qurulmuşdur ki, onların da əksəriyyəti riyazi düsturlar şəklindədir. *Oxunabilirlik düsturları* (*ing.readability formulas*) adlandırılan bu düsturlar elmi və kütləvi ədəbiyyatda *oxunabilirlik testləri* (*ing.readability tests*), *oxunabilirlik metrikası* (*ing.readability metrics*) kimi də ifadə olunur.

Oxunabilirlik düsturları, sadəcə, reqressiya analizi yolu ilə alınmış riyazi düsturlardır. Bu prosedur iki dəyişən arasında münasibəti ifadə edən tənliyi tapır. Belə ki, bu dəyişənlərdən biri verilmiş mətnin oxunması zamanı insanların çəkdikləri çətinliyin ölçüsünü, o biri isə həmin mətnin linqvistik xarakteristikalarının ölçüsünü göstərir [6].

Qeyd olunmalıdır ki, 1920-ci ildən 1980-ci ilədək ingilis dilindəki mətnlər üçün 200-dən artıq oxunabilirlik düsturu işlənib hazırlansa da, onların heç də hamısı kütləvilik qazana bilməmişdir. Onların içərisində aşağıdakı düsturlar daha geniş istifadə olunur [8]:

- Fleşin oxuma asanlıq düsturu (*ing. Flesch Reading Ease Formula*);
- Fleş-Kinsaydın təhsil səviyyəsi düsturu (*ing.Flesch-Kincaid Grade Level Formula*);
- Frayın oxunabilirlik düsturu (*ing.Fry Readability Formula*);
- Qanningin dumanlıq indeksi (*ing.Gunning Fog Index*);
- Deyl-Çellin oxunabilirlik düsturu (*ing.Dale-Chall Readability Formula*);
- SMOG düsturu (*ing.SMOG Formula*);
- Speşin oxunabilirlik düsturu (*ing.Spache Readability Formula*);
- Povers-Samner-Kerl oxunabilirlik düsturu (*ing.Powers-Sumner-Kearl Readability Formula*);
- FORCAST düsturu (*ing.FORCAST Formula*).

Nəzərə alınmalıdır ki, oxunabilirlik düsturları mətnin qavranılması üçün yeganə keyfiyyət göstəricisi deyildir. Mətnin qavranılması və materialın başa düşülməsinə cümlələrin quruluşu, səhifədə olan sözlərin sayı, illüstrasiyaların olması, çarpaz istinadlar, səhifədəki yeni sözlərin sayı, mücərrəd sözlərin sayı və bu kimi başqa faktorlar da təsir göstərir. Oxunabilirliyin qiymətləndirilməsi düsturları müəllif üslubunun incəliklərini dəyərləndirə bilmir, ancaq onlar sadə, aydın mətni mürəkkəb mətndən dəqiq fərqləndirir.

3. Oxunabilirliyin ölçülməsində istifadə olunan əsas dəyişənlər

Bəs mətnin hansı xarakteristikalarının onun mürəkkəbliyinə təsir etməsi necə müəyyən olunur? Başqa sözlə, mətnin mürəkkəblik dərəcəsini qiymətləndirmək üçün hansı amillər nəzərə alınmalıdır?

Mürəkkəblik amillərinin aşkarlanması prinsipi çox sadədir: bəziləri sadə, digərləri isə mürəkkəb olmaqla bir neçə mətn götürülür. Bu mətnlər ayrı-ayrı göstəricilər – məntiqi quruluş, mövzu, cümlələrin uzunluğu və başqa parametrlər üzrə müqayisə olunur. Əgər xarakteristikanın qiyməti asan mətnlərdən mürəkkəb mətnlərə keçdikcə dəyişirsə, demək, bu xarakteristika mətnin mürəkkəblik amillərindən biridir. Məsələn, sadə və mürəkkəb mətnlərin müqayisəsi zamanı mürəkkəb mətnlərdə daha çox naməlum sözlərin və uzun cümlələrin olmasını görmək olar. Deməli, sözlərin tanış olması və cümlələrin uzunluğu mətnin mürəkkəbliyi ilə bağlıdır [2].

Ümumiyyətlə, tədqiqatçılar mətnin mürəkkəbliyinə ən çox təsir edən 17 parametri qeyd edirlər (cədvəl 1) [2, 3].

Cədvəl 1

Mətnlərin mürəkkəbliyinə ən çox təsir edən parametrlər

№	Parametrin adı (azərbaycanca)	Parametrin adı (ingiliscə)
1.	Orta cümlə uzunluğu (sözlərlə)	Average sentence length in words
2.	“Asan” sözlərin faizi	Percentage of “easy” words
3.	6-cı sinif şagirdlərinin 90%-nin bilmədiyi sözlərin sayı	Number of words not known to 90% of sixth-grade students
4.	“Asan” sözlərin sayı	Number of “easy” words
5.	Fərqli “çətin” sözlərin sayı	Number of different “hard” words
6.	Minimum hecalı cümlə uzunluğu	Minimum syllabic sentence length
7.	Tam aydın cümlələrin sayı	Number of explicit sentences
8.	Birinci, ikinci və üçüncü şəxs əvəzlilərinin sayı	Number of first, second, and third-person pronouns
9.	Maksimum hecalı cümlə uzunluğu	Maximum syllabic sentence length
10.	Orta cümlə uzunluğu (hecalarla)	Average sentence length in syllables
11.	Birhecalı sözlərin faizi	Percentage of monosyllables
12.	Bir abzasdakı cümlələrin sayı	Number of sentences per paragraph
13.	6-cı sinif şagirdlərinin 90%-nin bilmədiyi fərqli sözlərin faizi	Percentage of different words not known to 90% of sixth-grade students
14.	Sadə cümlələrin sayı	Number of simple sentences
15.	Fərqli sözlərin faizi	Percentage of different words
16.	Çoxhecalı sözlərin faizi	Percentage of polysyllables
17.	Sözönlü ifadələrin sayı	Number of prepositional phrases

Sözsüz, bu 17 amilin bəzilərinin hesablanmasında başqa gərəkli göstəricilərdən də istifadə olunur (cədvəl 2).

Cədvəl 2

Oxunabilirlik düsturlarında istifadə olunan digər dəyişənlər

№	Parametrin adı (azərbaycanca)	Parametrin adı (ingiliscə)
1.	Mətdəki cümlələrin sayı	Number of sentences in the text
2.	Mətdəki sözlərin sayı	Number of words in the text
3.	Mətdəki “uzun” (3 və daha çoxhecalı) sözlərin sayı	Number of “long” words in the text
4.	Mətdəki birhecalı sözlərin sayı	Number of monosyllabic words in the text
5.	Hecaların sayı	Number of syllables
6.	Mətnin sözlərindəki simvolların sayı	Number of symbols in words of the text
7.	Mətdəki simvolların sayı (boşluq və durğu işarələri də daxil olmaqla)	Number of symbols in the text
8.	Çətin cümlələrin (durğu işarələri, bağlayıcılar, əvəzlilər, sifətlər və s. ilə zəngin olan cümlələrin) faizi	Percentage of hard (difficult) sentences
9.	Sözlərin orta uzunluğu (hecalarla)	Average word length in syllables
10.	Sözlərin orta uzunluğu (simvollarla)	Average word length in symbols

4. Fleşin oxuma asanlığı düsturu

Hər bir dilin özünəməxsus sintaksisi, yəni söz və cümlə quruluşu vardır. Başqa sözlə, bu parametrlərin qiymətləri müxtəlif dillər üçün eyni deyildir. Belə ki, məsələn, ingilis dilində sözlərin orta uzunluğu Azərbaycan, yaxud rus dili ilə müqayisədə qısadır. Bundan başqa, müxtəlif dillərdə sözlərdə olan hecaların miqdarı fərqli qaydada hesablanır: dillərin əksəriyyətində sözdəki hecaların sayı həmin sözdəki saillərin miqdarı ilə müəyyən olunsada, bəzi dillərdə bu say tələffüzdən asılı olur. Məsələn, ingilis dilində “analyse” sözü üç hecadan ibarətdir, ancaq “analyses” sözü dörd hecalıdır. Ona görə də bir dildəki mətnlər üçün hazırlanmış oxunabilirlik düsturunu başqa dildəki mətnlərə tətbiq etdikdə, düzgün olmayan nəticələr alınabilir. Daha dəqiq desək, ingilis dilinin sintaksisinə uyğun hazırlanmış və bəziləri ilə yuxarıda tanış olduğumuz oxunabilirlik düsturlarını Azərbaycan dilində olan mətnlərə tətbiq etmək üçün həmin düsturdakı əmsallar korrektə edilməlidir.

Bu məqalədə biz bu düsturlardan birini – Fleşin oxuma asanlığı düsturunu Azərbaycan dilində olan mətnlər üçün uyğunlaşdırmaq niyyətindəyik.

Fleşin oxuma asanlığı düsturunda (ing. Flesch Reading Ease Formula) "oxuma asanlığına" təsir edən faktorlar olaraq iki dəyişən – məndəki cümlə uzunluğu və sözdəki heca sayından istifadə edilir:

$$K = 206.835 - (1.015 \times S) - (84.6 \times W), \quad (1)$$

burada K – mətnin mürəkkəbliyinin (oxuma asanlığı) göstəricisi, S – orta cümlə uzunluğu (sözlərin sayının cümlələrin sayına nisbəti), W isə sözdəki hecaların orta sayıdır (hecaların sayının sözlərin sayına nisbəti).

Düsturun asan tətbiqi üçün bu yazıda belə bir metodika təklif olunur: hər hansı mətndən 100 söz götürülür, oradakı cümlələrin orta uzunluğu və həmin yüz sözdəki orta heca sayı hesablanır. Oxuma asanlığı qiyəti 100-dən (çox asan mətn) 0-dək (çox çətin mətn) dəyişə bilər (cədvəl 3).

Cədvəl 3

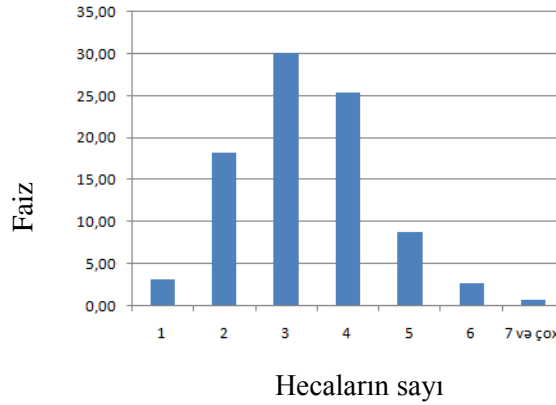
Fleşin oxuma asanlığı qiymətləri

Oxuma asanlığı qiyəti (K)	Təsviri
0 – 29	çox çətin mətn
30 – 49	çətin mətn
50 – 59	nisbətən çətin mətn
60 – 69	standart mətn
70 – 79	nisbətən asan mətn
80 – 89	asan mətn
90 – 100	çox asan mətn

Nəzəri olaraq, oxuma asanlığı qiyəti göstərilən intervaldan kənara da çıxa bilər. Belə ki, ən yüksək oxunabilirlik balı (ən asan mətn) təxminən 120 ola bilər: məsələn, belə mətnin hər bir cümləsi iki birhecalı sözdən ibarət olur. Oxunabilirlik balının nəzəri olaraq aşağı sərhədi (ən aşağı qiyəti) yoxdur, çünki mətnə ixtiyari sayda çoxhecalı söz əlavə etməklə onun oxunabilirlik balını istənilən qədər aşağı salmaq olar.

Bir daha vurğulamaq lazımdır ki, bu düstur təcrübi yolla ingilis dilində olan mətnlər üçün alınmışdır. Ancaq məlumdur ki, ingilis dilindən fərqli olaraq, Azərbaycan dili morfoloji tip baxımından şəkilçili (aqqütinativ) dil olduğundan bu dildə orta cümlə uzunluğu ingilis dili ilə müqayisədə az, orta söz uzunluğu isə, əksinə, çoxdur. Belə ki, bəzən Azərbaycan dilində bir sözlə ifadə olunan fikir ingilis dilində bir neçə sözdən ibarət cümlə şəklində olur; məsələn: "Evdəydim." ("I was at home"). Ona görə də, (1) düsturunu Azərbaycan dilində olan mətnlərə tətbiq etmək üçün həmin düsturdakı əmsallar korrektə edilməlidir.

"Azərbaycan dilinin izahlı lüğəti"ndəki [6] sözlərin statistikasına görə, sözlərin orta uzunluğu təxminən 3 hecadan ibarətdir: onların 3.18%-i birhecalı, 18.29%-i ikihecalı, 30.19%-i üçhecalı, 25.38%-i dördhecalı, 8.77%-i beşhecalı, 2.68%-i altıhecalı, 0.73%-i isə yeddi və daha çox hecalıdır (şəkil 1).



Şəkil 1. "Azərbaycan dilinin izahlı lüğəti"ndəki sözlərin heca sayına görə faizi

İndi orta cümlə uzunluğunun və orta söz uzunluğunun müxtəlif qiymətlərində (1) düsturuna görə oxuma asanlığının qiymətini hesablayaq (cədvəl 4).

Cədvəl 4

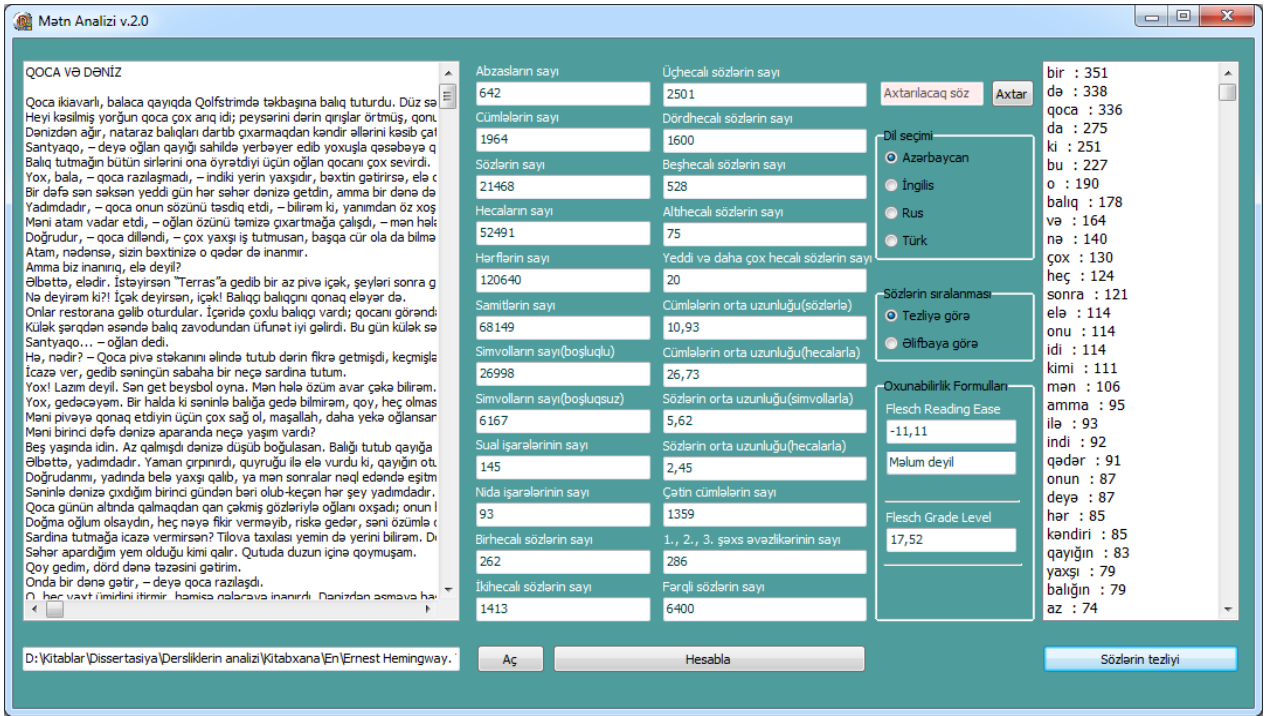
Fleş düsturunda oxuma asanlığı əmsalının sözlərin və hecaların sayından asılılığı

Sözdəki hecaların sayı (W)	Cümlədəki sözlərin sayı (S)								
	4	5	6	7	8	9	10	11	
2.1	25	24	23	22	21	20	19	18	
2.2	17	16	15	14	13	12	11	10	
2.3	8	7	6	5	4	3	2	1	
2.4	0	-1	-2	-3	-4	-5	-6	-7	
2.5	-9	-10	-11	-12	-13	-14	-15	-16	
2.6	-17	-18	-19	-20	-21	-22	-23	-24	
2.7	-26	-27	-28	-29	-30	-31	-32	-33	
2.8	-34	-35	-36	-37	-38	-39	-40	-41	
2.9	-43	-44	-45	-46	-47	-48	-49	-50	
3	-51	-52	-53	-54	-55	-56	-57	-58	

Cədvəldən görüldüyü kimi, orta heca uzunluğu 2.3 hecadan çox olduqda, Fleş düsturunda oxuma asanlığının qiyməti (K) mənfi olur. Başqa sözlə, Fleşin oxuma asanlığı düsturu Azərbaycan dilindəki mətnlər üçün işləmir. Bu problemi həll etmək üçün, yuxarıda da deyildiyi kimi, həmin düsturda əmsalları korrekt etmək lazımdır. Bu məqsədlə öncə Azərbaycan və ingilis dillərindəki orta söz uzunluqlarını müqayisə edək.

Araşdırmalara görə [8], V.K.Müllerin redaktəsi ilə hazırlanmış "İngiliscə–rusca lüğət"də ("Англо-русский словарь") [9] olan ingiliscə 41977 sözün orta uzunluğu 2.97 hecaya bərabərdir. Azərbaycan dilindəki sözlərin orta uzunluğunu hesablamaq üçün mənbə olaraq 48 mindən artıq sözü əhatə edən "Azərbaycan dilinin izahlı lüğəti"ni [10] götürəcəyik.

Azərbaycan dilindəki mətnlərin (sözlərin) kəmiyyət xarakteristikalarını hesablamaq üçün xüsusi proqram təminatı hazırlanıb. "Mətn analizi" proqram təminatı .doc, yaxud .docx formatlı fayllarla işləyir (şəkil 2). Nəticələr cədvəl 5-də əks olunub.



Şəkil 2. "Mətn analizi" proqram təminatının ekran görüntüsü

Cədvəl 5

İngilis və Azərbaycan dillərindəki sözlərin müqayisəli xarakteristikaları

Mənbənin adı	Sözlərin sayı	Hecaların sayı	Sözdəki hecaların sayı	Sözdəki samitlərin sayı	Çoxhecalı sözlərin sayı
İngiliscə–rusca lüğət	41977	124518	2.97	4.56	13495
Azərbaycan dilinin izahlı lüğəti	48457	153443	3.17	4.72	32830

Göründüyü kimi, ingilis dilində sözlərin orta uzunluğu 2.97 hecadır, Azərbaycan dilində isə bu göstərici 3.17-yə bərabərdir. Deməli, Fleş düsturunda sözün orta uzunluğu ilə bağlı əmsal 1.07 dəfə korrekte edilməlidir. Bundan başqa, "Azərbaycan dilinin izahlı lüğəti"ndə çoxhecalı sözlər, yəni üç və daha çox hecalı sözlər ümumi sözlərin 68%-ni təşkil etdiyi halda, ingilis dilində bu göstərici 32%-dir.

5. Azərbaycan dilində müxtəlif mətnlərin statistik təhlili

Yuxarıda deyildiyi kimi, Azərbaycan dili iltisafı dil olduğundan sözlərlə bağlı orfoqrafiya lüğətindən hesablanmış qiymətlər əsas olaraq götürülə bilməz. Daha dəqiq qiymətləri müxtəlif mətnlərin statistik təhlili əsasında almaq mümkündür. Belə bir metodikadan istifadə etməyi təklif edirik: öncə Azərbaycan dilində müxtəlif mətnlərin və həmin mətnlərin ingilis dilindəki orijinallarının statistik göstəriciləri və bizə lazım olan göstəricilərin hər iki dil üzrə qiymətlərinin nisbəti hesablanacaq.

Sonra bu göstəricilər üzrə orta qiymətlər çıxarılıb Fleş düsturundakı uyğun əmsalların hansı nisbətdə korrekte edilməli olduğu müəyyənləşdiriləcək. Hər iki dildəki mətnlərin məzmun və üslub baxımından səviyyəli olmasına əmin olmaq üçün əsas olaraq ingilis dilindəki bəzi məşhur bədii ədəbiyyat nümunələrindən və onların Azərbaycan dilindəki tərcümələrindən istifadə ediləcək.

Ancaq bədii əsərlər və onların tərcümələri yazıçının və tərcüməçinin üslubundan çox asılı olduğundan *azerbaijan.az* portalından və Azərbaycan Respublikası Prezidentinin rəsmi İnternet saytından (*www.president.az*) götürülmüş müxtəlif məzmunlu akademik mətnlər və onların ingilis dilindəki tərcümələri analoji qaydada tədqiq ediləcək. Tədqiqatı daha dolğun etmək üçün ingilis dilində ayrı-ayrı cümlələrin və onların Azərbaycan dilinə hərfi tərcümələrinin statistik göstəriciləri də nəzərə alınacaqdır.

İngilis dilindəki mətnlərin təhlili üçün <http://countwordsworth.com> və <http://www.wordcalc.com> ünvanlarındakı proqramlardan, Azərbaycan dilindəki mətnlərin təhlili üçün isə xüsusi olaraq bu tədqiqat üçün işlənilib hazırlanmış "Mətn analizi" proqramından istifadə ediləcək.

Bədii əsərlərin ingilis dilindəki elektron mətnləri pdbooks.ca/books/english/, www.fullbooks.com və <http://etc.usf.edu/lit2go/> veb-ünvanlarından götürülmüşdür. Həmin əsərlərin Azərbaycan dilindəki tərcümə mətnlərinin elektron versiyası isə "Altun kitab", "Şərq-Qərb", "Bakı" nəşriyyatlarından və www.translit.az saytından alınmışdır.

"Mətn analizi" proqramı mətnlərin aşağıdakı kəmiyyət xarakteristikalarını hesablayır:

- abzasların, cümlələrin, hecaların, hərflərin, samitlərin, simvolların, sual işarələrinin, nida işarələrinin sayı;
- birhecalı, ikihecalı, üçhecalı, dördhecalı, beşhecalı, altıhecalı, yeddi və daha çox hecalı sözlərin sayı;
- cümlələrin orta uzunluğu (sözlərlə), cümlələrin orta uzunluğu (hecalarla);
- sözlərin orta uzunluğu (simvollarla), sözlərin orta uzunluğu (hecalarla);
- birinci, ikinci və üçüncü şəxs əvəzliliklərinin sayı;
- fərqli sözlərin sayı.

Bu zaman aşağıdakı məqamlar nəzərə alınır:

1. Daxilində boşluq simvolunun və durğu işarələrinin olmadığı simvollar (hərflər, rəqəmlər, xüsusi işarələr) ardıcılığı bir söz kimi qəbul olunur (məsələn: oxunabilirlik, tərcümeyi-hal, BMT, 1962-ci).
2. Sözdəki *hecaların* sayı saiflərin sayına görə müəyyən olunur.
3. *Cümlə* böyük həflə başlayan və nöqtə, nida, yaxud sual işarəsi ilə bitən simvollar yığınıdır.

Cədvəl 6-da ingilis dilindəki ədəbiyyatın Mark Tven, Cek London, Herbert Uels, Ernest Heminquey, Con Qolsuorsi və başqa görkəmli nümayəndələrinin Azərbaycan dilinə tərcümə edilmiş bəzi əsərlərinin (yaxud fraqmentlərinin) və onların ingilis dilindəki orijinal mətnlərinin kəmiyyət xarakteristikaları göstərilib.

Həmin cədvəldən də göründüyü kimi, ingilis dilində sözlərin hecaya görə uzunluğu Azərbaycan dilindəki sözlərin uzunluğundan kiçikdir. Belə ki, ingilis dilində sözlərin orta uzunluğu $\bar{x} = 1.27$ hecadır (minimal uzunluq $-x_{\min} = 1.15$, maksimal uzunluq $-x_{\max} = 1.36$ hecadır, dispersiya isə $\sigma^2 = 0.005$ -dir). Həmin mətnlərin Azərbaycan dilinə tərcümələri üzrə qiymətlər belədir: sözlərin orta uzunluğu $- \bar{x} = 2.45$ heca, sözlərin minimal uzunluğu $- x_{\min} = 2.23$ heca, sözlərin maksimal uzunluğu $- x_{\max} = 2.58$ heca, dispersiya $- \sigma^2 = 0.01$.

İngilis və Azərbaycan dillərində eyni (identik) bədii ədəbiyyat nümunələrinin kəmiyyət xarakteristikalarının müqayisəsi

1	Əsərin adı	Cümlə sayı	Söz sayı	Heca sayı	Orta cümlə uzunluğu, OCU (söz)	Orta söz uzunluğu, OSU (heca)	OCU _{in} / OCU _{az}	OSU _{in} / OSU _{az}	
1	2	3	4	5	6	7	8	9	
1	Ernest Hemingway. Nobody Ever Dies (1939)	540	5264	7063	9.75	1.34	0.67	1.84	
	Ernest Hemingway. Heç kim heç vaxt ölmür (tərcümə: Kamran Nəzirli)	594	3865	9489	6.51	2.46			
2	Gabriel García Marquez. Monologue of Isabel Watching It Rain in Macondo (1955)	178	2902	3529	16.30	1.22	0.72	2.04	
	Qabriel Qarsia Marques. İsabel Makondada yağışa baxır (tərcümə: Natiq Səfərov)	157	1847	4602	11.76	2.49			
3*	Mark Twain. The Adventures of Tom Sawyer (1876)	142	1946	2371	13.70	1.22	0.72	1.83	
	Mark Tven. Tom Soyerin macəraları . (Tərcümə: Şəfiqə Ağayeva)	158	1549	3460	9.80	2.23			
4*	John Galsworthy. Beyond (1917)	352	7119	9001	20.22	1.26	0.76	2.02	
	Con Qolsuorsi. Ölümdən güclü (tərcümə: Aslan Quliyev)	366	5566	14199	15.27	2.54			
5	Herbert Wells. The Crystal Egg (1897)	301	6878	9324	22.85	1.36	0.62	1.90	
	Herbert Uels. Büllur yumurta	346	4889	12603	14.13	2.58			
6	Jack London. Grit of Women (1900)	362	5675	7428	15.68	1.31	0.69	1.82	
	Cek London. Qadın cəsarəti (tərcümə: Sevdə Abuzərli)	400	4344	10337	10.86	2.38			
7	John Steinbeck. The Chrysanthemums (1937)	448	4220	5742	9.42	1.36	0.88	1.87	
	Con Steynbek. Xrizantemlər (tərcümə: Ramiz Abbaslı)	369	3077	7743	8.34	2.52			
8	William Somerset Maugham. Louise (1969)	149	2118	2437	14.21	1.15	0.77	2.03	
	Uilyam Somerset Moyem. Luiza (tərcümə: Kamran Nəzirli)	196	2158	5040	11.01	2.34			
9*	Agatha Christie. Murder on the Orient Express (1934)	334	4737	6047	14.18	1.28	0.63	1.99	
	Aqata Kristi. Şərq ekspessində qətl (tərcümə: Pərviz Cəbrayıl)	425	3795	9683	8.93	2.55			
10*	Oscar Wilde. The Picture of Dorian Gray (1890)	363	4950	6097	13.63	1.23	0.76	1.93	
	Oskar Uayld. Dorian Qreyin portreti (tərcümə: Kamran Nəzirli)	422	4369	10414	10.35	2.38			
Y e k u n									
							orta	0.72	1.93
							dispersiya	0.0059	0.0076
							min	0.62	1.82
							max	0.88	2.04

$$\text{Dispersiya_söz_uzunluğu}_{ing} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = 0.005.$$

$$\text{Dispersiya_söz_uzunluğu}_{aze} = 0.01.$$

Cədvəlin 8-ci sütununda hər bir əsər üçün aşağıdakı nisbət hesablanıb:

$$\frac{\text{Cümlədəki sözlərin orta sayı (ingilis dilindəki mətn)}}{\text{Cümlədəki sözlərin orta sayı (Azərbaycan dilinə tərcümə mətni)}}$$

Cümlədəki söz sayı bir tərəfdən dilin sintaktik özəllikləri ilə bağlı olsa da, digər tərəfdən yazıcının və tərcüməçinin müəllif üslubu ilə bağlıdır. Ona görə də bu göstərici ayrı-ayrı əsərlərdə əhəmiyyətli dərəcədə fərqlənir. Cədvəldən görüldüyü kimi, cümlədəki söz sayı üzrə səpələnmə (variasiya) 0.62-dən 0.88-dək aralıqda yerləşir, ingilis dilindəki cümlələr Azərbaycan dilindəki cümlələrdən təxminən 0.72 dəfə uzundur, dispersiya və orta kvadratik sapma isə uyğun olaraq 0.0059 və 0.0768-dir:

$$\text{Sapma miqyası (ing. range of deviation): } R = x_{\max} - x_{\min} = 0.88 - 0.62 = 0.26$$

$$\text{Örnək orta (ing. sample mean) yaxud seçimin ədədi ortası: } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 0.72$$

$$\text{Dispersiya (ing. variance): } \sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = 0.0059$$

$$\text{Orta kvadratik sapma (ing. standard deviation): } \sigma = 0.0768$$

Cədvəlin 9-cu sütununda isə hər bir əsər üçün aşağıdakı nisbət hesablanıb:

$$\frac{\text{Sözdəki hecaların orta sayı (ingilis dilindəki mətn)}}{\text{Sözdəki hecaların orta sayı (Azərbaycan dilinə tərcümə mətni)}}$$

Bu ədədi qiymətlər ingilis dilindəki mətnlərdə bir sözdəki orta heca sayının Azərbaycan dilindəki mətnlərdəki uyğun göstəricidən orta hesabla 1.93 dəfə kiçik olduğunu göstərir. Bu qiymətlər üzrə səpələnmə (variasiya) 1.82-dən 2.04-dək aralıqda yerləşir, dispersiya və orta kvadratik sapma isə uyğun olaraq 0.0076 və 0.0872-dir.

Statistik təhlil olunan bu mətnlər, görüldüyü kimi, bədii mətnlərdir. İndi akademik (elmi) xarakterli bir neçə mətn nümunəsini eyni qayda ilə təhlil edək (cədvəl 6*).

Cədvəl 6*

İngilis və Azərbaycan dillərində akademik xarakterli eyni (identik) mətnlərin kəmiyyət xarakteristikalarının müqayisəsi

	Mətn	Cümlə sayı	Söz sayı	Heca sayı	Orta cümlə uzunluğu, OCU (söz)	Orta söz uzunluğu, OSU (heca)	OCU _{in} /OCU _{az}	OSU _{in} /OSU _{az}	
1	2	3	4	5	6	7	8	9	
1	Azerbaijan (<i>president.az</i>)	40	840	1405	21.00	1.67	0.82	1.77	
	Azərbaycan	40	691	2047	17.28	2.96			
2	Karabakh (<i>president.az</i>)	160	3650	5812	22.81	1.59	0.78	1.72	
	Qarabağ	160	2844	7788	17.78	2.74			
3	Bloody Memories (<i>president.az</i>)	382	8734	14542	22.86	1.66	0.76	1.72	
	Qanlı yaddaş	382	6595	18824	17.26	2.85			
4	History of Military in Azerbaijan (<i>azerbaijan.az</i>)	90	2148	3736	23.87	1.74	0.82	1.63	
	Azərbaycanın hərbi tarixi	105	2058	5816	19.60	2.83			
5	Sport History in Azerbaijan (<i>azerbaijan.az</i>)	237	3949	6334	16.66	1.60	0.88	1.77	
	Azərbaycanın idman tarixi	235	3450	9750	14.68	2.83			
6	Tourism Policy (<i>azerbaijan.az</i>)	91	2533	4286	27.84	1.69	0.88	1.77	
	Turizm siyasəti	86	2102	6281	24.44	2.99			
7	Nakhchivan Autonomous Republic (<i>azerbaijan.az</i>)	179	3225	4946	18.02	1.53	0.80	1.76	
	Naxçıvan Muxtar Respublikası	169	2441	6582	14.44	2.70			
8	Public Health System in Azerbaijan (<i>azerbaijan.az</i>)	37	954	1690	25.78	1.77	1.03	1.76	
	Azərbaycanda səhiyyə	33	875	2733	26.52	3.12			
9	Science of Azerbaijan (<i>azerbaijan.az</i>)	133	3045	5941	22.89	1.95	0.88	1.59	
	Azərbaycan elmi	130	2610	8104	20.08	3.10			
10	Azerbaijani Diaspora (<i>azerbaijan.az</i>)	76	1643	2966	21.62	1.80	1.02	1.68	
	Azərbaycan diasporu	67	1482	4475	22.12	3.02			
Y e k u n									
							orta	0.87	1.72
							dispersiya	0.0087	0.0041
							min	0.76	1.59
							max	1.03	1.77

Daha bir növ mətnin statistik təhlilinə baxaq (cədvəl 6**).

Cədvəl 6**

İngilis dilində ayrı-ayrı cümlələrin və onların Azərbaycan dilində tərcümələrinin kəmiyyət xarakteristikalarının müqayisəsi

Mətn	Cümlə sayı	Söz sayı	Heca sayı	Orta cümlə uzunluğu, OCU (söz)	Orta söz uzunluğu, OSU (heca)	OCU _{in} / OCU _{az}	OSU _{in} / OSU _{az}
İngilis dilində ayrı-ayrı cümlələr	208	2312	2636	11.11	1.14	0.73	2.09
Bu cümlələrin Azərbaycan dilində hərfi tərcümələri	208	1682	4009	8.09	2.38		

6. Azərbaycan dilindəki mətnlər üçün Fleş düsturu

Bu üç növ mətn (cədvəl 6, 6* və 6**) üzrə alınmış nəticələr üçün orta qiymət çıxarsaq:

$$\frac{\text{Cümlədəki sözlərin orta sayı (İngilis dilindəki mətn)}}{\text{Cümlədəki sözlərin orta sayı (Azərbaycan dilinə tərcümə mətni)}} = \frac{0.72 + 0.87 + 0.73}{3} = 0.77$$

$$\frac{\text{Sözdəki hecaların orta sayı (İngilis dilindəki mətn)}}{\text{Sözdəki hecaların orta sayı (Azərbaycan dilinə tərcümə mətni)}} = \frac{1.93 + 1.72 + 2.09}{3} = 1.91$$

Beləliklə, (1) – Fleşin oxuma asanlıqı düsturunda orta cümlə uzunluğunun əmsalı (S) 0.77 dəfə, hecalarla orta söz uzunluğunun (W) əmsalı isə 1.91 dəfə korrektə edilməlidir. Onda *Azərbaycan dilindəki mətnlər üçün Fleşin oxuma asanlıqı düsturu* aşağıdakı şəkildə olacaq:

$$K = 206.835 - (1.318 \times S) - (44.3 \times W), \quad (2)$$

burada K – mətnin mürəkkəbliyinin (oxuma asanlıqının) göstəricisi, S – orta cümlə uzunluğu, W isə sözdəki hecaların orta sayıdır.

İndi orta cümlə uzunluğunun və orta söz uzunluğunun müxtəlif qiymətlərində oxuma asanlıqının qiymətini (2) düsturuna görə hesablasaq, cədvəl 4 aşağıdakı şəkildə olacaq (cədvəl 7).

Yuxarıda qeyd olunan (1) və (2) düsturlarından istifadə etməklə cədvəl 6-da verilmiş bədii əsər nümunələrinin hər iki dildəki mətnlərinin oxunabilirlik qiymətlərini hesablayıb müqayisə etsək (cədvəl 8), orijinal mətnlərin və tərcümə mətnlərinin mürəkkəblik səviyyəsinin, demək olar ki, eyni olduğunu görürük (korrelyasiya əmsalı – 0.722).

Cədvəl 7

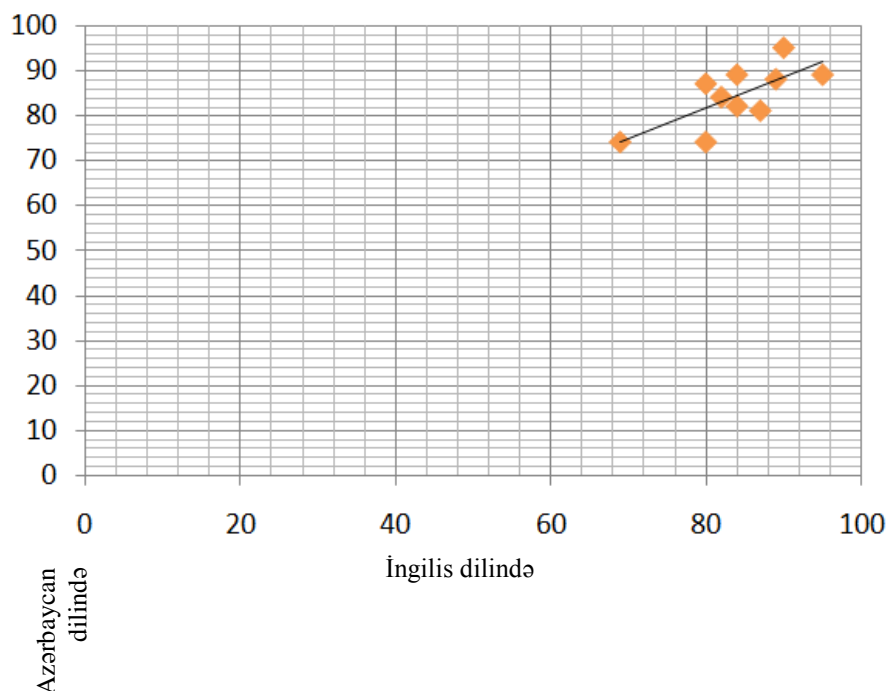
Azərbaycan dilindəki mətnlər üçün uyğunlaşdırılmış Fleş düsturunda oxuma asanlıqı əmsalının sözlərin və hecaların sayından asılılığı

	Cümlədəki sözlərin sayı (S)								
	4	5	6	7	8	9	10	11	
Sözdəki hecaların sayı (W)	2.1	109	107	106	105	103	102	101	99
	2.2	104	103	101	100	99	98	96	95
	2.3	100	98	97	96	94	93	92	90
	2.4	95	94	93	91	90	89	87	86
	2.5	91	89	88	87	86	84	83	82
	2.6	86	85	84	82	81	80	78	77
	2.7	82	81	79	78	77	75	74	73
	2.8	78	76	75	74	72	71	70	68
	2.9	75	72	70	69	68	67	65	64
	3	69	67	66	65	63	62	61	59

İngilis və Azərbaycan dillərində eyni (identik) mətnlərin oxunabilirliyinin Fleş düsturuna görə qiymətlərinin müqayisəsi

	Ə s ə r i n a d ı	Fleşin oxuma asanlığı əmsalı – K (ingilis)	Fleşin oxuma asanlığı əmsalı – K (azərb.)	Fərq
1.	Ernest Heminquey. Həç kim həç vaxt ölmür	84	89	5
2.	Qabriel Qarsia Marques. İsabel Makondada yağışa baxır	87	81	6
3.	Mark Tven. Tom Soyerin macəraları	90	95	5
4.	Con Qolsuorsi. Ölümdən güclü	80	74	6
5.	Herbert Uels. Büllur yumurta	69	74	5
6.	Cek London. Qadın cəsarəti	80	87	7
7.	Con Steynbek. Xrizantemlər	82	84	2
8.	Uilyam Somerset Moyem. Luiza	95	89	6
9.	Aqata Kristi. Şərq ekspessində qətl	84	82	2
10.	Oskar Uayld. Dorian Qreyin portreti	89	88	1
	korrelyasiya	0.722		

İngilis və Azərbaycan dillərində eyni (identik) mətnlərin oxunabilirliklərinin korrelyasiya qrafiki şəkil 3-də verilmişdir.



Şəkil 3. İngilis və Azərbaycan dillərində eyni (identik) mətnlərin oxunabilirliklərinin korrelyasiya qrafiki

7. Nəticə

Şəkil 3-dən görüldüyü kimi, ingilis dilində olan mətnlərin və onların Azərbaycan dilinə tərcümə mətnlərinin oxunabilirlikləri arasında müsbət korrelyasiya var. Bu isə Fleşin oxuma asanlığı düsturundakı əmsalların düzgün korrektə edildiyini göstərir. Deməli, Azərbaycan dilindəki mətnlərin mürəkkəblik səviyyəsini müəyyənləşdirmək üçün (2) düsturundan və cədvəl 3-dən istifadə etmək olar.

Ədəbiyyat

1. Abdullayev A.S. Azərbaycan dilinin tədrisi tarixindən. Bakı: Maarif, 1966.
2. Sadıqov İ.C. Mətnlərin mürəkkəbliyi və onun qiymətləndirilməsi yolları. Ekspres-informasiya. İnformasiya cəmiyyəti seriyası, Bakı: “İnformasiya Texnologiyaları” nəşriyyatı, 2012, 72 səh.
3. DuBay W.H. The Principles of Readability – Costa Mesa, California: Impact Information, 2004.
4. Klare G.R. The measurement of readability. Ames: Iowa State University Press, 1963.
5. Hargis G., Hernandez A.K., Hughes P., Ramaker J., Rouiller S., Wilde E. Developing quality technical information: A handbook for writers and editors. Upper Saddle River, NJ: Prentice Hall, 1998.
6. McLaughlin G.H. SMOG grading – a new readability formula // Journal of reading, 1969, vol.12, no.8, 639–646.
7. Dale E., Chall J.S. The concept of readability. Elementary English, 1949, 26: 23.
8. Оборнева И. В. Автоматизация оценки качества восприятия текста // Вестник Московского городского педагогического университета, 2005.
9. Мюллер В.К. Англо-русский словарь, М.: Русский язык, 1995.
10. Azərbaycan dilinin izahlı lüğəti. Dörd cildə, Bakı: Şərq-Qərb, 2006.

УДК 004.89

Садыгов Исмаил Дж.

Институт Информационных Технологий НАНА, Баку, Азербайджан
ismayil.sadigov@gmail.com

Модифицированная формула Флеша для оценивания сложности текстов на азербайджанском языке

В статье затрагивается проблема применения формулы легкости чтения по Флешу для текстов на английском языке к текстам на азербайджанском языке. Для решения этой проблемы описывается методика корректировки коэффициентов в формуле, соответствующих таким параметрам, как среднее количество слов в предложении и среднее число слогов в слове. На основе данной методики исследуются идентичные тексты на английском и азербайджанском языках (образцы художественных произведений, академические (научные) тексты, отдельные предложения и их дословные переводы). В результате корректируются коэффициенты формулы Флеша для текстов на азербайджанском языке.

Ключевые слова: сложность текста, формулы читабельности, формула легкости чтения по Флешу.

Ismayil J. Sadigov

Institute of Information Technology ANAS, Baku, Azerbaijan
ismayil.sadigov@gmail.com

Modified flesch formula for evaluating the complexity of texts in azerbaijani language

The article touches upon the issue of applying the Flesch reading ease formula for texts in English to texts in Azerbaijani language. To solve this problem, we describe a technique for adjusting the coefficients in a formula corresponding to parameters such as average sentence length in words and average word length in syllables. Based on this methodology, identical texts in English and Azerbaijani languages (samples of works of art, academic (scientific) texts, individual sentences and their verbatim translations) are studied and, as a result, the coefficients of the Flesch formula for texts in Azerbaijani language are adjusted.

Keywords: text complexity, readability formulas, Flesch reading ease formula.