

مطالعه آواشناسی^۱ آماری زبان فارسی
کاربرد آن در مطالعات اکوستیکی

دکتر هوشنگ پاکزاد
بخش آکوستیک - گروه فیزیک - دانشگاه تهران

چکیده

یکی از عواملی که برای ارزیابی ویژگی های آکوستیکی تالارها - تلفن ها و سایر وسایل ارتباط صوتی مورد لزوم است تعیین کمی میزان دریافت کلمات گویندگان است. بدین منظور روشی آماری بکار می برند که در آن از جدولهای لگاتوم^۲ استفاده می گردد. لگاتوم عبارت از مجموعه ای از هجاهای بدون معنی در زبان است که معمولا " از سه واج^۳ تشکیل می یابد. واج اول و آخر هم خوان^۴ (بی صدا) و واج میانی وا که^۵ (با صدا) است. لگاتوم هر زبان باید دارای ویژگیهای صوتی همان زبان بوده و بر خصوصیات ساختمان زبان منطبق باشد. باین سبب اکثر زبانهای زنده دنیا لگاتوم مخصوص بخود آماده کرده اند. کار تحقیقی حاضر بمنظور تهیه لگاتوم برای زبان فارسی انجام یافته و با بررسی حدود ۳۱۱۱۱ حرف از متون قدیم، ادبیات معاصر، زبان رایج مردم و مطالب روزنامه ها، اطلاعاتی از ساختمان زبان فارسی رایج بدست می دهد. این بررسی بدو صورت انجام یافته است:

- ۱- استفاده از نیروی انسانی شامل ۸ دانشجو که حدود ۲۰۰ ساعت کار انجام داده اند.
- ۲- استفاده از کامپیوتر الکترونیکی. بدین منظور برنامه ای بزبان فورترن^۶ برای محاسبه بسامد^۷ تکرار هم خوانها، واکه های زبان فارسی و متوسط طول کلمه ها در این زبان و سایر اطلاعات لازم از یک متن فارسی، نوشته شد و محاسبات لازم در کامپیوتر IBM ۳۷۰ دانشگاه تهران اجرا گردید. نتایج حاصل بصورت جدول و نمودار در متن مقاله مندرج است. دو لیست لگاتوم فارسی در انتهای مقاله ارائه شده است. (جدول ۸)

1- Phonétique

2- Logatomes

3- Phoneme

4- Consonne

5- Voyelle

6- Fortran

7- Fréquence

پیش‌گفتار

در تالیفات و آثاری که راجع به زبان فارسی وجود دارد اشاراتی در مورد آواشناسی زبان فارسی مشاهده می‌شود ولی از نظر آواشناسی آماری و ریاضی یعنی بررسی کمی هجاها، حرف‌ها و واجهای تشکیل‌دهنده زبان فارسی مطلب مهمی وجود ندارد. هر چند راجع به حروف زبان و واجها و هجاها، زبان فارسی مطالبی می‌توان یافت (۱ و ۲ و ۱۶) ولی راجع به بسامد تکرار واجها و هجاها و حروف مطالبی یافت نمی‌شود. در صورتیکه مهندسی ارتباط صدائی بویژه تلفن و همچنین آکوستیک دانهائی که به امر آکوستیک معماری و آواشناسی می‌پردازند احتیاج مبرمی به این اطلاعات دارند. چه هدف این افراد تراکسیل و تقویت واجهای است که از ترکیب آنها زبان لفظی رایج بین مردم بوجود می‌آید. مهندسی مخابرات لیستهای از واج‌ها و یا بهتر بگوئیم مجموعه‌ای از واجها بنام لگاتوم بدون در نظر گرفتن ساختمان زبان استفاده کنندگان از وسائل و متکی به زبان اسپرانتو ایجاد کرده و در آزمایشهای مربوط به تلفن از آنها استفاده می‌کردند (لگاتوم عبارت از صورتی از هجاها در زبان است که در قرارداد مهندسی مخابرات از سه واج تشکیل یافته واج اول هم خوان واج میانی واکه و بالاخره واج سوم نیز هم خوان است. یک نمونه از این لیستها در آخر مقاله مندرج است جدول شماره ۱). این صورت‌ها حتی بر ساختمانهای زبانهای اروپائی که منشاء زبان اسپرانتو است منطبق نیست تا چه رسد به زبان فارسی. واضح است که تشابه قابل ملاحظه‌ای بین ساختمان و خصوصیات زبان فارسی و زبان اسپرانتو وجود ندارد و استفاده از این لیستها برای زبان فارسی فاقد ارزش علمی و ذوقی است.

شناخت ساختمان زبان فارسی و خصوصیات آن که دارد مخصوصاً از نظر بسامد تکرار حروف، هم خوانها، واکه‌ها، طول متوسط کلمات (متوسط تعداد حروف بکاررفته در یک کلمه)، نسبت هم خوانها به واکه‌ها، طویلترین کلمه در آثار نویسندگان، چگونگی تلفظ واکه‌ها حائز کمال اهمیت است. هر چند در باره بسیاری از موضوع‌های بالا در زبان فارسی هنوز شرحی در کتابها نمی‌توان یافت و کار دقیق علمی و آماری بر روی آنها انجام نشده است ولی اطلاعات مفیدی از کتاب تاریخ زبان فارسی (۱) می‌توان بدست آورد.

در این مقاله مطالب فوق را در ۶ قسمت شرح داده و در انتهای آن دولیست لگاتوم زبان فارسی را ارائه می‌دهیم.

۱- کلیات - برای مطالعه شرایط دقیق تراکسیل صدا در وسایل صوتی و بناها غالباً ساختمان آواشناسی زبان رایج مردم مورد لزوم است. کاربرد فوری این اطلاعات در ساختن لیستهای از واجها است که برای مطالعه میزان دریافت صدا بکار می‌رود و باین وسیله می‌توان سهولت انتقال صدا و تغییر شکل آنرا بین گوینده و شنونده در مکالمات تلفنی و یا خطابه‌های ایراد شده در سالن‌ها بدست آورد. مطالعات لازم برای اندازه‌گیری بسامد تکرار واکه‌های زبان خارجی توسط مراکز تحقیق بعمل آمده است از جمله فلت چر (۳) اطلاعاتی در مورد آواشناسی زبان انگلیسی و شاواس (۴) مطالعه حامی در باره زبان فرانسه ارائه نموده‌اند. نتایج این بررسی‌ها، تفاوت قابل ملاحظه‌ای در بسامد تکرار واج و حرف برای زبانهای مختلف نشان می‌دهد. مثلاً " بسامد تکرار حرف a در زبان اسپانیولی ۱۲/۹٪ و در زبان آلمانی ۴/۶٪ است و یا بسامد تکرار e در زبان آلمانی ۱۸٪ در زبان اسپانیولی ۱۴/۲٪ است و همچنین بسامد حرف w در زبانهای اسپانیولی، ایتالیائی، فرانسه صفر و در زبان آلمانی ۱/۶٪ می‌باشد (۴). در جدول ۲ نتایج این بررسی برای زبانهای فرانسه (۴) آلمانی - ایتالیائی - انگلیسی - اسپانیولی و سوئدی (۲۲ و ۲۱) آورده شده است.

هر چند استفاده از آمار و ریاضیات در زبان فارسی توسط محمد نوری عثمانف (۵) برای تهیه واژنامه بسامدی دیوان عنصری بکاررفته و تحقیق جالبی در توانائی زبان فارسی توسط جناب آقای دکتر حسابی (۷) بعمل آمده ولی در مورد بسامد تکرار واج‌ها - واکه‌ها - هم خوانها و نسبت آنها تاکنون مطالعه‌ای در زبان فارسی مشاهده نگردیده و تا آنجاکه نویسنده تحقیق نموده است مطالعه در زبان که منجر به ایجاد خط "تندنویسی" ^۱ در زبان فارسی گردد نیز وجود نداشته است.

جدول شماره ۲ نسبت درصد بسامد حروف در ۷ زبان اروپائی

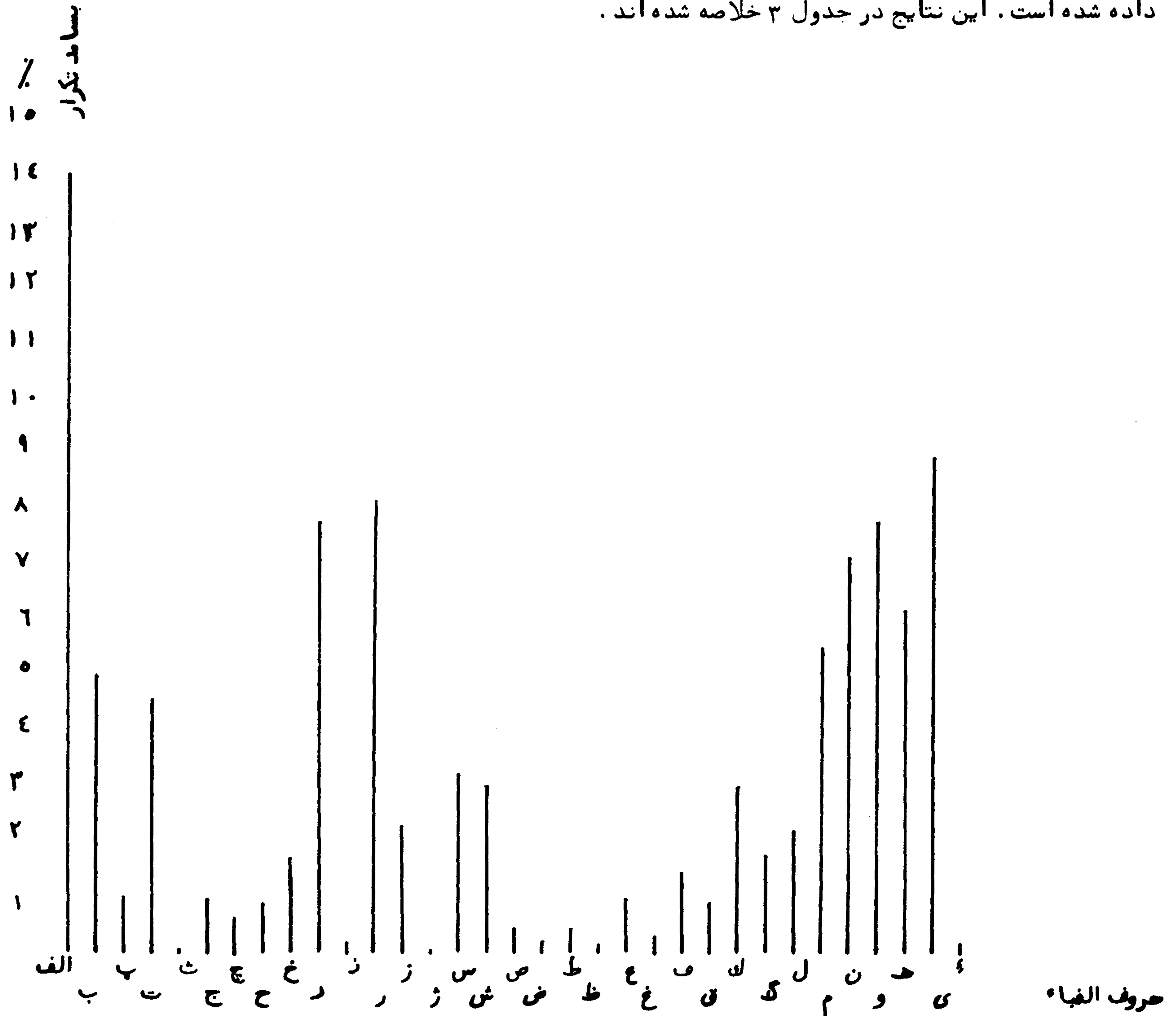
زبان سوئدی	زبان اسپانیولی	زبان ایتالیائی	زبان انگلیسی	زبان آلمانی	زبان فرانسوی	حروف
۹/۴۲	۱۲/۹	۱۰/۳	۷/۸	۴/۶	۷/۴۷	۶/۹۴
۱/۷۰						۰/۴۵
۲/۱۸						۰/۰۸
۱/۲۹	۱/۰	۰/۹	۱/۳	۱/۹	۰/۸۷	b
۱/۲۲	۴/۲	۴/۳	۲/۹	۳/۱	۳/۱۶	۱/۶۵
						۱/۴۴
۴/۵۵	۴/۶	۳/۸	۴/۱	۵/۵	۳/۶۷	۰/۰۷
						۳/۶۷
۹/۵۹	۱۴/۳	۱۲/۶	۱۳/۱	۱۸/۰	۱۷/۶۶	۵/۰۹
						۱/۸۹
						۰/۴۰
۲/۱۵	۰/۷	۰/۸	۲/۸	۱/۵	۱/۱۱	۰/۲۸
						۱/۱۱
۳/۱۴	۱/۰	۲/۰	۱/۴	۳/۰	۰/۷۷	۰/۴۴
						۰/۳۳
۲/۰۷	۰/۹	۱/۱	۵/۹	۴/۴	۰/۸۱	h
۵/۵۹	۷/۰	۱۱/۶	۶/۸	۷/۲	۷/۳۸	i
۰/۶۳	۰/۳	۰	۰/۱	۰/۶۰	۰/۶۰	j
۳/۲۲	۰	۰	۰/۴	۱/۳	۰	k
۵/۲۷	۵/۵	۶/۵	۳/۶	۳/۷	۵/۶۹	l
۳/۶۴	۲/۵	۲/۶	۲/۶	۲/۰	۳/۰۴	m
۸/۸۴	۶/۴	۶/۶	۷/۳	۹/۴	۷/۲۴	n
۳/۹۹	۸/۸	۸/۷	۸/۲	۲/۵	۵/۴۰	۵/۳۷
۱/۵۴						۰/۰۳
۱/۶۰	۳/۳	۳/۲	۲/۲	۰/۷	۲/۷۶	p
۰/۰۰۱	۱/۵	۰/۶	۰/۱	۰	۱/۳۴	q
۸/۳۰	۷/۰	۶/۷	۶/۷	۷/۶	۶/۳۳	r
۶/۴۱	۷/۶	۶/۱	۶/۵	۶/۳	۸/۵۰	s
۸/۸۰	۴/۴	۶/۱	۹/۰	۶/۶	۷/۰۸	t
۱/۷۶	۴/۰	۳/۰	۲/۸	۵/۱	۶/۸۳	u
۲/۲۸	۰/۷	۱/۵	۱/۰	۱/۰	۱/۵۲	v
						۰
۰/۰۰۲	۰	۰	۱/۵	۱/۶	۰/۰۰۲	w
۰/۰۰۹	۰/۱	۰	۰/۳	۰	۰/۳۷	x
۰/۰۶۷	۱/۰	۰	۱/۵	۰	۰/۲۶۷	y
۰/۰۱۵	۰/۳	۰/۹	۰/۱	۲/۴	۰/۱۳۱	z

امید است نتایج این تحقیق شروع مناسبی برای رفع این کمبود ها در زبان فارسی باشد .
 برای تعیین بسامدهای تکرار یعنی نسبت درصد استعمال و تکرار هر حرف - واک - واکسه - هم خوان در زبان فارسی چهار رشته آزمایش و بررسی در آزمایشگاه بخش آکوستیک گروه فیزیک انجام گرفت . هر رشته شامل ۷ متن و هر متن نیز از حدود ۱۰۰۰ حرف تشکیل یافته است . با وجود آنکه پژوهشگران خارجی فقط به متون خاصی توجه نموده اند مثلا " شاواس (۴) از کتابهای نویسندگانی مانند لامارتین - بالزاک - ویکتور هوگو . . . ، یعنی ادبیات کلاسیک فرانسه استفاده کرده و یا ویارایس (۲۵) سه صفحه از مجله تایم را انتخاب و بررسی نموده ولی در این مقاله متون هائی بصورت نظم و نثر از ادبیات قدیم (شاهنامه - گلستان - چهارمقاله) و ادبیات معاصر (سگ ولگرد از صادق هدایت - جادو از علی دشتی جامی بگیر از منصور برمکی - زمستان از اخوان ثالث - تلخ و شیرین از جمال زاده و تنگسیر از صادق چوبک . . .) و روزنامه های عصر زبان فارسی (کیهان و اطلاعات) و مجلات مختلف (دانشمند - شکار . . .) و بالاخره محاوره افراد از رادیو و مردم انتخاب شدند تا بتوان نتایج حاصله را برای زبان فارسی رایج امروز از عامیانه تا خطابه ای تعمیم داد . از انتخاب متن های علمی که در آنها اصطلاحات خارجی بکار برده شده است چشم پوشی گردید زیرا در آنها تعداد زیادی اصطلاح های خارجی وجود دارد که مورد استفاده عموم نیست و نمی توان آنها را جزء زبان فارسی دانست ولی لغات و اصطلاحات خارجی که در زبان فارسی رایج و متداول شده و اکثر مردم آنها را می شناسند و با تلفظ فارسی بیان می کنند مانند تلفن - تلویزیون - سینما و . . . را در محاسبات منظور کرده ایم . اسامی خاصی که در متن های انتخابی وجود داشتند در محاسبات منظور شده اند تا هیچگونه تغییری در نمونه انتخابی داده نشود . همچنین انتخاب متن ها نیز کاملا " اتفاقی انجام گرفته و هیچگونه گزینش خاصی مورد توجه نبوده است . این روش توسط عده ای از پژوهشگران نیز انجام شده است (۸) .

در مجموع ۳۱ متن اختیار شده جمعا " ۳۱۱۱۱ حرف بررسی شده است که شامل ۱۷۵۱۱ واکه و ۸۲۰۳۹ کلمه بوده است .
 ۲ - بررسی از نظر حروف و واجها . بین استادان زبان فارسی در مورد تعداد حروف این زبان اختلاف نظر وجود دارد . برهان قاطع (۱۵) مبنای زبان فارسی را بر ۲۴ حرف قرار می دهد چه فاسی زبانها ۸ حرف زبان عربی (ث - ح - ص - ض - ط - ظ - ع - ق) را که تلفظ آنها برایشان ثقیل بوده حذف کرده و به ۲۰ حرف باقیمانده چهار حرف مخصوص زبان فارسی (پ - چ - ژ - گ) را اضافه نمودند . در برهان قاطع چنین آمده است " . . . و از حسن اتفاق مبنای کلام فارسی بر ۲۴ حرف است و ساعت شبانه روزی نیز ۲۴ است و چهار حرف دیگر از حروف بیست و هشتگانه که آن (ح - ط - ع - ق) بر سیل ندرت در لغت ما و اءالنهر آمده است " (۱۵) این موضوع شاید در زمان تحریر برهان قاطع درست بوده است و امروز هم این اظهار دست کم در متن های بررسی شده ما از شاهنامه تأیید می گردد .

چه از حدود ۳۰۰۰ حرف بررسی شده از شاهنامه حرف " ص " سه بار در کلمه " صدا " و حرف " ط " یکبار در کلمه " طوس " و حرف " ث " یکبار در کلمه " ثریا " و بالاخره حرف " غ " سه بار در کلمه های " غم و تیغ و میغ " آمده است و حروف " ح - ض - ط - ع - ق " اصلا وجود نداشته است . با توجه به چاپ کتاب شاهنامه مورد استفاده که در سال ۱۳۲۳ هـ - ق . و در دوره قاجاریه بوده است احتمال اینکه اصل این کلمات با املاء دیگری بودند زیاد است . استادان معاصر تعداد حروف فارسی را بین ۳۲ و ۳۳ ذکر می کنند یعنی عده ای " ۳۲ " را نیز بعنوان یک حرف بحساب می آورند که از آن جمله اند . . . دستور امروز (۲) دستور زبان فارسی (۹ و ۱۰ و ۱۱ و ۱۲ و ۱۳) و استور نامه (۱۴) در مقابل عده ای نیز تعداد حروف را ۳۲ حرف بشمار می آورند مانند دستور زبان فارسی (۱۶ و ۱۷ و ۱۸) و دستور تعلیم الفباء (۱۹) . با وجود آنکه تعداد " ۳۲ " در بررسی ما نیز زیاد نیست ولی در این مطالعه به پیروی از اکثر استادان زبان فارسی تعداد حروف را ۳۳ اختیار نمودیم . خوشبختانه در بررسی واکها اشکالات موجود در زبانهای اروپائی در زبان فارسی کمتر وجود دارند . مثلا " در زبان فرانسه واک " ح " به سه صورت " ملایم " ۱ در کلمه " célèbre " و " سخت " ۲ در کلمه " canon " و بالاخره " سدیل " ۳

در کلمه " Garçon " وجود دارد . گو اینکه بعضی از واکهای زبان فارسی به دو طریق تلفظ می‌گردند مثل واک " ن " که در کلمه های " جان " و " نماز " هنگام ادای آن سر زبان به پشت دندانهای پیشین بالامی چسبد و در کلمه‌های مانند " چنگ و بانگ " هنگام تلفظ " ن " زبان به میانکام تکیه می‌کند (۲) ولی با توجه به اینکه اختلاف زیادی از نظر شنیدن ندارند و فارسی زبان واک " ن " را در هر دو قسم تشخیص می‌دهد یک واک منظور شد . همچنین واک " ی " در مواردی که صدای الف می‌دهد مانند " کسری و عیسی " الف منظور شده است . حروفی که در زبان فارسی صدای یکسانی دارند مانند " ت " و " ط " در بررسی اولیه جداگانه حساب شده اند ولی از نظر صدا این گونه حروف یک واک محسوب گردیدند . در جدول ۳ نتایج این بررسی برای چهار سری درج شده است . بطوریکه از این جدول پیدا است بزرگترین بسامد تکرار را حرف " الف " با ۱۴/۵۳ % و کمترین بسامد تکرار را حرف " ژ " با ۰/۲۲ % دارد یعنی با وجود آنکه حرف ژ از حروف مخصوص زبان فارسی است و در زبان فارسی نیز استعمال می‌شود ولی شمار آن بسیار کم است بطوریکه در حدود ۲۱۰۰۰ حرف بررسی شده سه بار آنهم در یک متن شاهنامه دیده شده است و باید توجه داشت که نظیر این حرف در سایر زبانها هم دیده می‌شود . مثلاً " در حدود ۴۰۰۰۰ حرف بررسی شده در زبان فرانسه یک نمونه K هم وجود نداشته است و یا نسبت در صد تکرار حرف W فقط ۰/۵۰۲ بوده (۴) و یا حرف Q در زبان سوئدی دارای نسبت درصد تکرار ۰/۵۰۱ است (۲۱) در شکل ۱ نمودار میانگین بسامد تکراری حروف زبان فارسی نشان داده شده است . این نتایج در جدول ۳ خلاصه شده اند .



شکل ۱ - بسامد تکرار حروف زبان فارسی، آمار بر روی ۳۱۱۱۱ حرف گرفته شده است .

حروف الفبا	میانگین سری اول شامل ۷ متن	میانگین سری دوم شامل ۷ متن	میانگین سری سوم شامل ۷ متن	میانگین سری چهارم شامل ۱۰ متن	میانگین کل
الف	۱۴/۶۰	۱۳/۷۷	۱۲/۹۶	۱۴/۸۰	۱۴/۰۳
ب	۵/۶۵	۴/۱۳	۵/۷	۴/۶۳	۵/۰۳
پ	۰/۸۶	۰/۹۶	۱/۲۶	۰/۹۱	۱/۰۰
ت	۴/۴۹	۵/۲۹	۴/۰۸	۴/۲۷	۴/۵۳
ث	۰/۰۴	۰/۱۳	۰/۱	۰/۱	۰/۰۹
ج	۱/۱۱	۱/۰۴	۱/۰۳	۰/۸	۰/۹۹
چ	۰/۶۲	۰/۶	۰/۷	۰/۵۶	۰/۶۲
ح	۱/۱۷	۰/۵۶	۰/۹۴	۰/۷۹	۰/۸۹
خ	۱/۶۶	۱/۲۳	۲/۱۸	۱/۵۸	۱/۶۶
د	۸/۴۷	۷/۷۷	۷/۳۵	۷/۴۷۵	۷/۷۶
ذ	۰/۱	۰/۱۷	۰/۱۷	۰/۲۳	۰/۱۷
ر	۷/۶۷	۷/۹۴	۸/۸۲	۸/۴۵	۸/۲۲
ز	۲/۲۰	۱/۹۷	۲/۶	۲/۲۸	۲/۲۶
ژ	۰	۰	۰/۰۷	۰/۰۲	۰/۰۲
س	۳/۲۲	۳/۶۹	۳/۵۴	۲/۷۸	۳/۱۸
ش	۲/۵۵	۲/۶۷	۳/۴۷	۳/۰۷۵	۲/۹۴
ص	۰/۴۸	۰/۴۳	۰/۴	۰/۵۳	۰/۴۶
ض	۰/۲۲	۰/۱۷	۰/۲۶	۰/۱۳	۰/۱۹۵
ط	۰/۳۵	۰/۵	۰/۳۹	۰/۴۹	۰/۴۳
ظ	۰/۲۱	۰/۱۶	۰/۱	۰/۱۷	۰/۱۶
ع	۱/۰	۰/۹۹	۰/۸	۱/۰۲	۰/۹۵
غ	۰/۱	۰/۳۳	۰/۴۳	۰/۳۴	۰/۳
ف	۱/۳۴	۱/۶۶	۱/۳۴	۱/۴۹	۱/۴۶
ق	۰/۷۹	۰/۹۷	۰/۸۷	۰/۸۵	۰/۸۷
ک	۲/۹۳	۲/۸۹	۳/۰۶	۳/۱	۳/۰
گ	۱/۴۶	۱/۶۷	۲/۱۴	۱/۶۷	۱/۷۴
ل	۲/۲۹	۲/۱	۲/۴۳	۱/۸۶	۲/۱۷
م	۵/۲۰	۵/۸۵	۵/۴۴	۵/۴۶	۵/۴۹
ن	۷/۲۸	۷/۲۳	۷/۰۵	۷/۰۱	۷/۱۴
و	۶/۶۸	۶/۸۹	۶/۹	۶/۴۳	۶/۷۲۵
ه	۵/۸۲	۶/۶۹	۵/۴۷	۶/۷۰	۶/۱۷
ی	۸/۹۷	۹/۳۹	۷/۸	۹/۵۴۵	۸/۹۳
ء	۰/۳۸	۰/۱۴	۰/۲	۰/۱۱	۰/۲۱

۳- بررسی واکه ها - مطالعاتی که در زبان فارسی از نظر آواشناسی انجام گرفته نتایج یکنواخت و یکسانی ارائه نمی دهند یعنی در مورد واکه ها - هم خوانها و چگونگی تلفظ آنها و تقسیم بندیهای انجام شده تفاوتی وجود دارد. از آنجائی که این مقاله می تواند مورد استفاده مهندسين و اکوستیک دانها باشد لذا چند تعریف از زبان فارسی درج می شود. گو اینکه این تعریف ها برای علمای زبان فارسی کاملاً " شناخته شده هستند .

واکه ها طبق تعریف تاریخ زبان (۱) واجهائی هستند که برای ادای آنها گذرگاه هوا از گلو به بعد آزاد می ماند. یعنی برای ادای آنها ابتداهان باز شده و سپس بسته می گردد. تعداد واکه های زبان فارسی را بعضی از محققین (۲، ۱۰، ۱۲، ۱۳، ۱۴) ۸ واکه ذکر می کنند یعنی علاوه بر سه حرکت زیر - زبر - پیش - ۵ حرکت دیگر عبارتند از .

آ در کلماتی چون یاد - باد - داد

او در کلماتی چون هوش - بود - سود

ای در کلماتی چون بید - ایران - ایرج

او (OW) در کلماتی چون جو - مو - نو - خسرو

ای (EY) در کلماتی چون پی - پیک - پیام - کی - ری

۶ واکه اولی را اصلی و دو واکه آخری را مرکب می گویند. در مقابل این عده گروهی نیز (۱ - ۹ - ۱۶ - ۱۸ - ۱۹) واکه ها را فقط ۶ واکه اصلی می دانند و دو واکه آخری را ترکیب دو واکه ساده ذکر می کنند. در این کار ما بررسی را بر روی ۸ واکه انجام دادیم. هر چند تعداد واکه های مرکب چندان زیاد نیست.

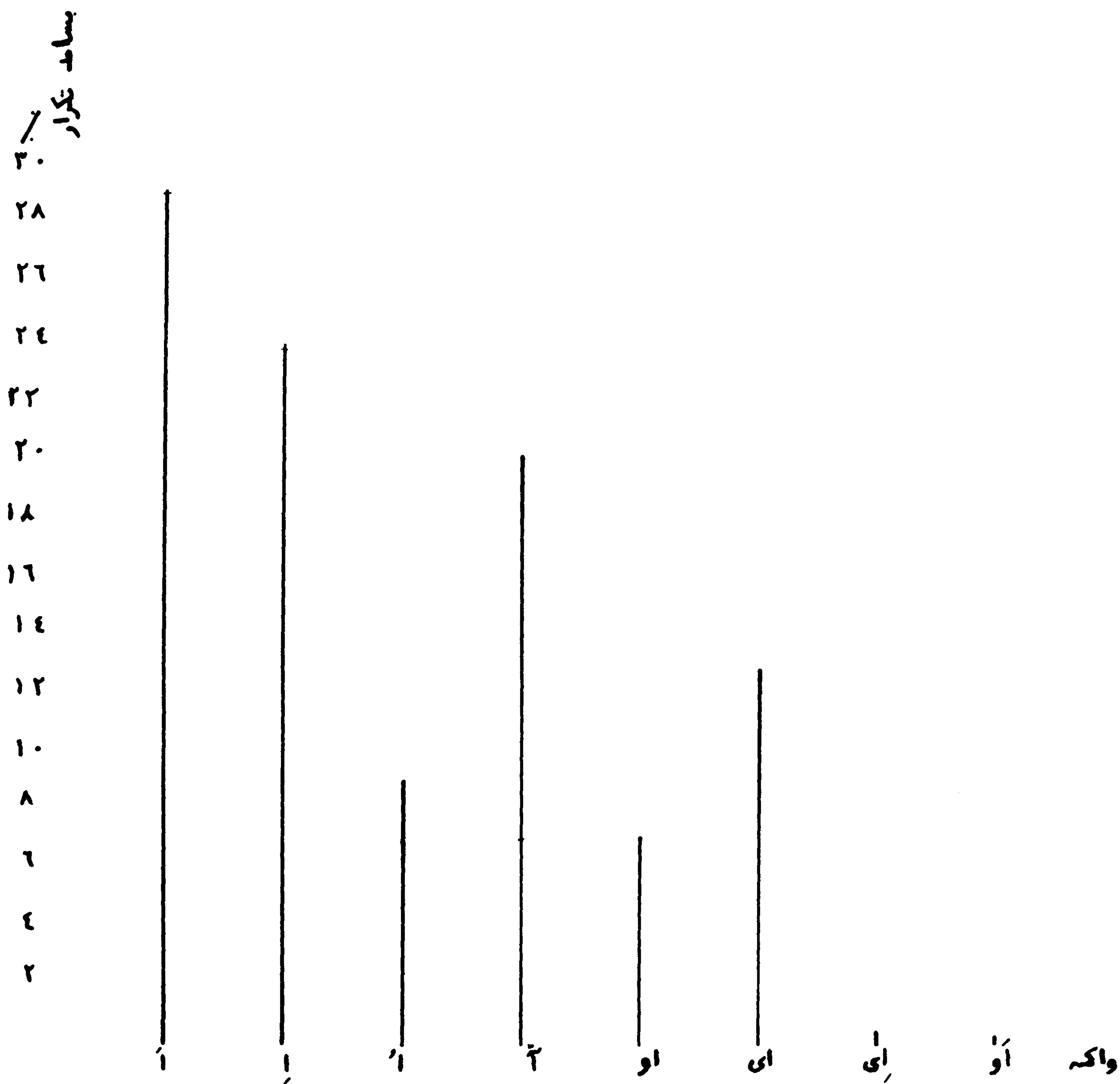
کوچکترین واحد قابل تلفظ در زبان فارسی را هجا می گویند. هجاها را از نظر کشش به چهار قسم تقسیم می کنند (۲) کوتاه، متوسط، بلند، دراز. واکها را از نظر چگونگی تلفظ در دستگاه تولید صدا به لبی، لبی دندانی، پیشکامی، میان کامی، پسکامی، ملازی و گلوئی تقسیم می کنند. بحث جامع و قدیمی در این مورد را می توان در کتاب مخارج الحروف ابوعلی سینا (۶) جستجو کرد. در بین استادان معاصر که به این موضوع پرداخته اند مطالب دستور امروز (۲) و تاریخ زبان فارسی (۱) مختصر و مفید است. بررسی های انجام یافته بر روی ۴۸۱۲۲ واک نشان می دهد که نسبت درصد تعداد واکه ها به تعداد هم خوانها برابر است با ۵۴/۷ که با تقریب ۲/۸ ± در تمام متن ها ثابت مانده این نسبت در زبانهای اروپائی طبق بررسی والرو (۲۲) چنین است که با نسبت ۲ ± صحیح است.

زبان	فرانسه	آلمانی	انگلیسی	ایتالیائی	اسپانیولی
نسبت واکه به هم خوان	۴۴/۵	۳۵	۴۰	۴۶	۴۸

در شکل ۲ نمودار تغییرات بسامد تکرار واکه های زبان فارسی طبق بررسی های این مقاله داده شده است. این نتایج در جدول ۴ خلاصه گردیده است.

باید توجه داشت هر چند در زبان عربی از نظر آواشناسی و تلفظ بین حروفی مانند (ت و ط) و (ح و ه) تفاوتی وجود دارد ولی فارسی زبان این حروف را مانند اعراب تلفظ نمی کند. مثلاً " در زبان فارسی حروف ظ - ض - ذ - ز بیک شکل و با صدای " ز " تلفظ می شوند. یعنی کلمات (ظاهر)، (ضبط)، (ذره) و (زبر) همه با صدای " ز " ادا می کردند. لذا ما حروفی که از نظر صدا برای فارسی زبانها مشابه بودند در یک دسته قرار دادیم. باین ترتیب تعداد حروف تلفظی و یا واکه های زبان فارسی به ۲۳ تقلیل می یابد^۱. بسامد تکرار این واجها در جدول شماره ۵ و شکل ۳ نموده شده اند.

(۱) در زبان فرانسه نیز بعضی حروف چنین هستند مثلاً " حرف (X) در کلمه (Soixante) با صدای (S) و در کلمه (Deuxieme) با صدای (Z) و در (Extra) با صدای (CS) و بالاخره در کلمه (Cxacte) با صدای (gs) تلفظ می شود لذا شاواس (۴) حروف X را از حروف تلفظی زبان فرانسه حذف نموده است.

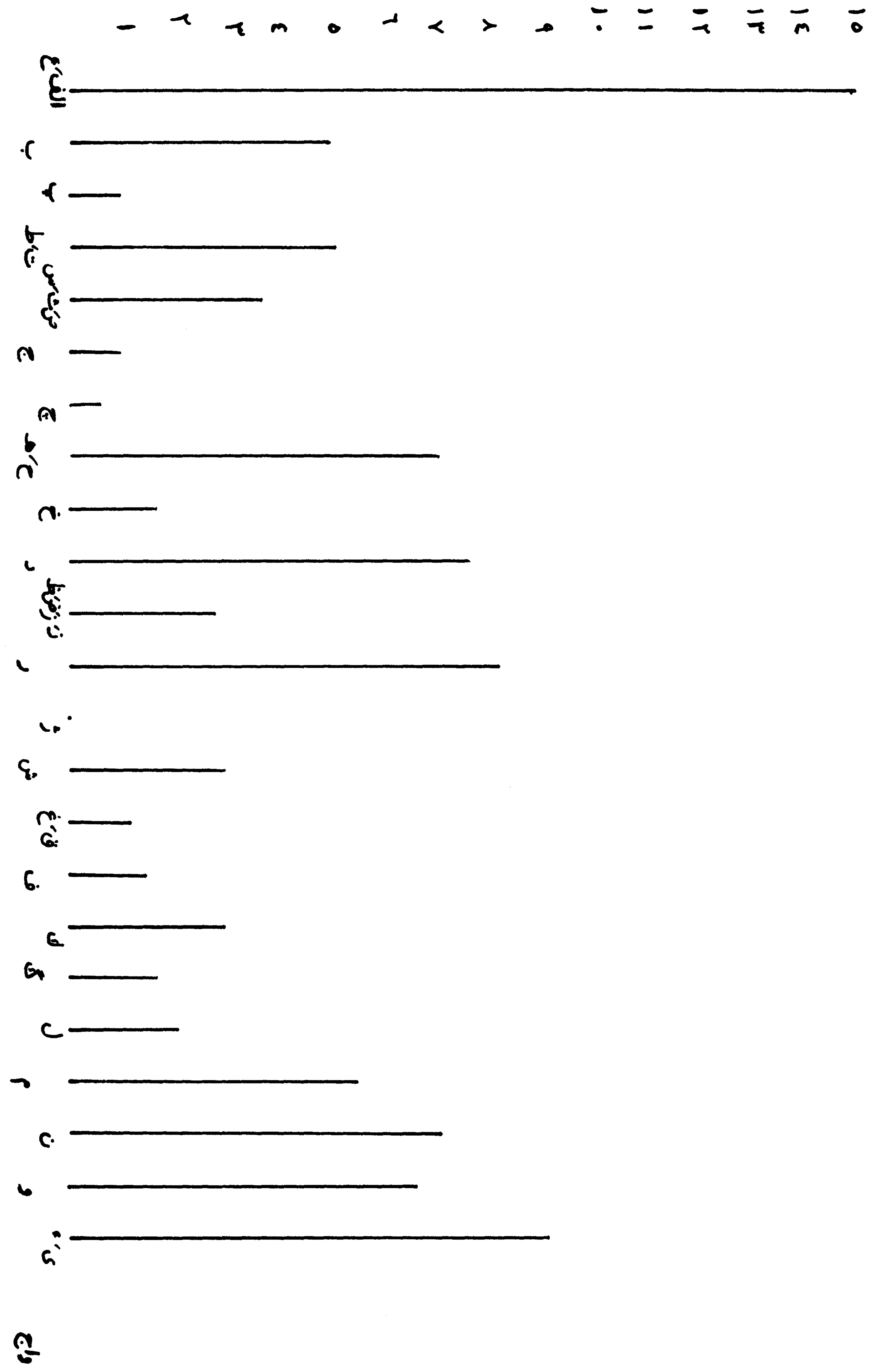


شکل ۲ - بسامد تکرار واکه‌های زبان فارسی، آمار بر روی ۱۷۰۱۱ واکه.

واکه	میانگین سری اول شامل ۷ متن	میانگین سری دوم شامل ۷ متن	میانگین سری سوم شامل ۷ متن	میانگین سری چهارم شامل ۱۰ متن	میانگین کل
ا	۲۹/۷۳	۲۹/۶	۲۹/۴۸	۲۷/۸۵	۲۹/۱۶
ا	۲۲/۷۳	۲۴/۱۳	۲۳/۷۵	۲۳/۸۴	۲۳/۶۱
ا	۹/۴۱	۸/۹۴	۹/۷۷	۷/۹۱	۹/۰۱
آ	۱۹/۱۷	۱۹/۳۴	۲۰/۰۲	۲۱/۹۱	۲۰/۱۱
او	۵/۴۷	۳/۹	۵/۳۳	۴/۹۵	۴/۹۱
ای	۱۳/۲۷	۱۳/۷۷	۱۰/۸۵	۱۲/۵۴	۱۲/۶۱
ای	۰/۰۹	۰	۰/۵۷	۰/۶۵	۰/۳۳
او	۰/۱۳	۰/۳۲	۰/۲۳	۰/۴۱	۰/۲۷

جدول ۴ - میانگین نسبت درصد بسامد واکه‌های زبان فارسی

بساطه تکرار



جدول شماره ۵ نسبت در صد بسامد گروه حروف هم صدا در زبان فارسی

بسامد تکرار بطور غیر مستقیم	بسامد تکرار معدل ۳۱ متن	تعداد حروف بررسی شده	گروه حروف هم صدا در زبان فارسی
۱۵/۰۷	۱۴/۹۸	۴۶۹۰	الف - ع
۴/۹۷	۵/۰۳	۱۵۴۵	ب
۰/۹۹	۱/۰	۳۰۷	پ
۵/۱۰	۴/۹۶	۱۵۸۷	ت - ط
۳/۶۸	۳/۷۳	۱۱۴۶	س - ث - ص
۰/۹۸	۰/۹۹	۳۰۶	ج
۰/۶۱	۰/۶۲	۱۹۱	چ
۷/۱۰	۷/۰۶	۲۲۱۰	ح - ه
۱/۶۶	۱/۶۶	۵۱۸	خ
۷/۷۳	۷/۷۶	۲۴۰۶	د
۲/۸۱	۲/۷۹	۸۷۳	ذ - ز - ض - ظ
۸/۲۶	۸/۲۲	۲۵۷۱	ر
۰/۰۲	۰/۰۲	۷	ژ
۲/۹۴	۲/۹۴	۹۱۳	ش
۱/۱۸	۱/۱۸	۳۶۸	ق - غ
۱/۴۶	۱/۴۶	۴۵۵	ف
۳/۰۱	۳/۰	۹۳۶	ک
۱/۷۲	۱/۷۴	۵۳۷	گ
۲/۱۳	۲/۱۷	۶۶۴	ل
۵/۵۱	۵/۴۹	۱۷۱۳	م
۷/۱۲	۷/۱۴	۲۲۱۵	ن
۶/۷	۶/۷۲	۲۰۸۴	و
۹/۲۲	۹/۱۴	۲۸۶۹	ی - ی'
۹۹/۹۷	۹۹/۹۵	۳۱۱۱۱	جمع

۴- بررسی از نظر ویژگی‌های کلمه. در این بررسی کلمه را مجموعه‌ای از هجاها که یک معنی و مفهوم خاص را بدهند اختیار نمودیم. تشخیص کلمات مجرد معمولاً بدون اشکال است. در مورد کلمات مرکب و انتخاب آنها بعنوان یک یا چند کلمه به نوع تلفظ و معنی آنها توجه شده است. مثلاً "در عبارت" از سرگذشت "اگر تاء کید بر هجای (سر) باشد در نتیجه عبارت سه کلمه "از" "سر" و "گذشت" دارد و اگر هجای سر بدون تاء کید تلفظ شود عبارت شامل دو کلمه "از" و "سرگذشت" است (۱). همچنین کلمات مرکبی که معنی واحدی دارند مانند "سرخوردگی" یا "دلتنگی" یک کلمه حساب شده اند و کلماتی با پیشوند یا پسوند را نیز یک کلمه بحساب آوردیم. در این بررسی مانند نوری عثمانف (۵) حرف ربط و اضافه و... را یک کلمه محسوب نمودیم. کلماتی که دارای تلفظ واحد ولی املاء و معنی متفاوت بودند مانند "صفر" و "سفر" جداگانه بحساب آوردیم.

طول کلمه - منظور از طول کلمه شماره حروف موجود در آن می‌باشد. در این بررسی ۸۲۰۵ کلمه بررسی شده در نتیجه هر کلمه بطور متوسط شامل ۳/۷۹ حرف می‌باشد. این طول متوسط کلمه در شاهنامه ۳/۵۵ در گلستان سعدی ۳/۹۰ است و حداکثر طول کلمه در متنهای مختلف بین دو مقدار ۸ و ۱۴ قرار دارند. در جدول ۶ نتایج این بررسی خلاصه شده است.

۵- روش بررسی. نتایج آماری این بررسی با دوروش استفاده از نیروی انسانی و کامپیوترهای الکترونیکی بدست آمده است.

الف - استفاده از نیروی انسانی. این بررسی بکمکتی چند از دانشجویان رشته فیزیک در آزمایشگاه آگوستیک گروه فیزیک دانشگاه تهران بعمل آمد. بدین ترتیب که سه رشته آزمایش توسط سه گروه ۲ تا ۳ نفری انجام گردید. در هر گروه نتایج هر بررسی بوسیله عضو گروه بازبینی می‌گردید و نتایج هر گروه توسط گروه دیگر دوباره بازبینی و بالاخره بررسی نهائی روی سه رشته کار فوق انجام می‌گرفت. روش بررسی بدینقرار بوده است که برای هر رشته ۷ متن از ادبیات قدیم - ادبیات کنونی - مطالب روزنامه‌ها و سرانجام مطالب گفتگو و مکالمه افراد با گویندگان رادیو و تلویزیون انتخاب می‌نمودیم. در هر متن حدود ۱۰۰۰ حرف را بوسیله شمارش مشخص می‌نمودیم. آنگاه ابتدا شمارش برای تعیین حروف و سپس شمارش برای تعیین بسامد تکرار واژه‌ها و سرانجام شمارش برای تعیین تعداد کلمات انجام می‌گرفت و در جدولهای درج می‌شد. در ضمن طویلترین کلمه هر متن انتخاب و یادداشت می‌گردید. نتایج کلی هر یک از آزمایشهای فوق از جمع شمارش‌های جزئی و محاسبات لازم بدست می‌آمد.

ب - استفاده از کامپیوتر الکترونیکی - بمنظور بازبینی محاسبات دستی و همچنین احتراز از اشتباه و استفاده از تسهیلات کامپیوترها برنامه‌ای بزبان فورترن ۴ برای شمارش کیفیت‌های ذکر شده نوشتیم. این برنامه قادر است تک تک حروف بی صدا و باصدا را از یک متن فارسی جدا کرده و نسبت درصد آنرا در متن تعیین نماید. همچنین با شمارش تعداد کلمات طول متوسط هر کلمه را مشخص کند. چون در زبان فارسی تعدادی از واژه‌ها بدون آنکه نوشته شوند تلفظ می‌گردند لذا برای آنها علامتهائی که با امکانات ماشین وفق دهد (جدول ۷) انتخاب نمودیم. واضح است که این علامات هیچگونه ارتباطی با خط‌آوانویسی که در آواشناسی بکار می‌رود (۲۳) ندارد.

در ضمن علامت یک فاصله سفید برای شمارش کلمه‌ها و عدد ۶ برای همزه انتخاب گردید. متن مورد نظر را با رعایت این قرارداد نوشته و منگنه (پانچ) نمودیم. برنامه‌ها در کامپیوتر IBM مدل ۱۳۵ دانشگاه تهران اجرا نمودیم. ذیلاً یک متن را بعنوان نمونه با ذکر نتایج حاصله از کامپیوتر درج می‌نمائیم.

جدول شماره ۶ - مشخصات کلمه ها در متون فارسی

نام نویسنده	نام کتاب	تعداد متن	تعداد حروف	تعداد کلمه	طول متوسط کلمه	ماکزیم طول
نظامی عروضی	چهارمقاله	۱	۸۱۹	۲۳۹	۳/۴۲	۹
سعدی	گلستان	۳	۲۹۷۶	۷۶۲	۳/۹۰	۱۱
فردوسی	شاهنامه	۳	۲۸۶	۸۰۷	۳/۵۵	۱۲
حافظ	دیوان حافظ	۱	۱۰۳۱	۲۸۷	۳/۵۹	۹
صادق هدایت	سگ ولگرد	۱	۱۰۰۲	۲۳۶	۴/۲۴	۱۱
منصور برمکی	جامی بگیر	۱	۱۰۱۵	۳۰۲	۳/۳۶	۱۰
علی دشتی	جادو	۱	۹۹۷	۲۷۳	۳/۶۵	۹
اخوان ثالث	دفترهای زمانه، زمستان	۲	۲۰۸۰	۵۷۲	۳/۶۴	۱۲
جمالزاده	تلخ و شیرین	۱	۱۰۰۰	۲۷۲	۳/۶۷	۱۱
صادق چوبک	تنگسیر	۱	۱۰۰۰	۲۷۴	۳/۶۵	۹
فروغ فرخزاد	تولد دیگر	۱	۱۰۵۹	۲۵۶	۴/۱۳	۹
محمود دولت آبادی	سفر	۱	۱۰۲۵	۲۹۸	۳/۴۴	۸
دهخدا	اخبار شهری	۱	۸۵۴	۲۶۴	۳/۲۳	۱۰
دکتر زرین کوب	فرار از مدرسه	۱	۱۰۶۶	۲۷۸	۳/۸۳	۱۴
	نقل از روزنامه ها و مجلات	۸	۸۲۰۶	۲۰۱۱	۴/۰۸	۱۳
	نقل از گفتار رادیو ایران	۳	۳۱۸۳	۷۹۴	۴/۰۱	۱۲
	نقل از مکالمه دونفر	۱	۹۳۰	۲۸۳	۳/۲۹	۹
جمع		۳۱	۳۱۱۱۱	۸۲۰۸	۳/۷۹	

جدول شماره (۷)

واکه در زبان فارسی	زیر	زیر	پیش	آ	او	ای	او	ای
علامت اختیاری برای ماشین	/	—	*	۱	۲	۳	۴	۵

۶- تشکیل لیستهای لگاتوم . برای تهیه لیستهای لگاتوم تاکنون از روشهای مختلفی استفاده شده است . یک روش نسبتاً " قدیمی " که مورد استفاده اداره پست و تلگراف فرانسه واقع گشته چنین است :

ابتدا یک متن غیرمشخص از زبان موردنظر را انتخاب نموده و آنرا به جاهای تقسیم می‌کنند و سپس ۵۰ یا صد عدد از این هجاها را در یک لیست لگاتوم قرار می‌دهند . چون زبان رایج مردم از ترکیب این هجاها ایجاد می‌گردد می‌توان گفت که لگاتوم ساخته شده به ساختمان زبان نزدیک است ولی چون یک متن خاص نمی‌تواند معرف ساختمان کلی زبان باشد و همچنین بعضی از این هجاها دارای معنی بوده و یا در شنونده معنی کلمه اصلی را که این هجا با آن تعلق دارد تداعی می‌نماید لذا از این روش کمتر استفاده می‌شود .

روش دیگر . ساختن مصنوعی هجاهای بارعایت نسبت بین هم خوانها و واژه‌های زبان رایج بین مردم است این روش به دو صورت انجام می‌شود :

الف - در این حالت یک یا دو هم خوان در اول هجا قرار گرفته و یک واکه بعد از آن وجود دارد .

ب - در این روش هجای ساخته شده بترتیب از یک هم خوان اولی - یک واکه میانی و بالاخره یک هم خوان نهائی تشکیل می‌شود . لگاتوم حاصل همان مشخصات و مشکلات لگاتوم اسپرانتو را دارد (۴)

برای ایجاد لگاتوم در زبان فارسی ما از ۸ واکه و ۲۳ هم خوان زبان لفظی فارسی که بسامد تکرار آنها در جدول ۴ و ۵ آمده است استفاده کردیم . یعنی از تمام حروفی که صدای مشابهی دارند یکی را انتخاب نمودیم ولی بسامد تکرار آنها برابر مجموع بسامد تکرار حروف مشابه اختیار نمودیم . مثلاً " بین س - ث - ص حرف س را انتخاب و با بسامدی برابر مجموع بسامد های تکرار سه حرف فوق بکار بردیم همچنین نسبت هم خوانها به واکه ها در یک لیست لگاتوم را برابر با این نسبت در زبان فارسی قرار دادیم . با توجه به طول متوسط کلمه‌ها در زبان فارسی که در حدود ۳/۸ حرف است مالکاتوم را به صورت هجاهای سه و چهار واجی ساختیم و چون این هجاها نباید دارای معنی باشند لذا هجاهای معنی دار را از لیست حذف کردیم پس لگاتوم سه تایی ما از یک هم خوان اولی و یک واکه میانی و بالاخره یک هم خوان نهائی تشکیل می‌شود و لگاتوم چهار تایی از یک هم خوان یک واکه یک هم خوان و یک واکه تشکیل می‌گردد . بسامد تکرار هر واج در یک لیست لگاتوم برابر بسامد تکرار واجهای زبان فارسی است که ما در این مقاله با محاسبات آماری بدست آوردیم . در جدول ۸ دو لیست لگاتوم که هر یک از ۵۰ هجا بدون معنی تشکیل یافته نشان داده شده است .

تشکر

از جناب آقای دکتر مهدی برکشلی استاد محترم دانشگاه که مشوق و راهنمای اینجانب در انجام این تحقیق بودند و همچنین خانم سهیلا مگری دانشجوی فیزیک بخاطر همکاری در کارهای آماری صمیمانه تشکر می‌کنم .

جدول شماره ۸

لیست لگاتوم بزبان فارسی شماره ۱

أب - هُن - رَت - شُد - سَخ - نَد - مَز - قَر - رُخ - اِب - پِب - رِد - کِن - کِج - مِد - پِک - رَس - اُر - دُن - هُش -
 تال - واب - اِپ - هاز - دای - پاک - فاد - لور - روک - رین - میف - ایت - لیه - ناش - واس - هات - مود -
 ویت - رین - پس - اس - چمد - اُهم - وِهِن - وِن - اِک - اِهت - اُپم - هز - یُشب

لیست لگاتوم بزبان فارسی شماره ۲

هت - گد - نک - وُم - وُف - هُج - یُش - رُن - نل - اِد - اِخ - پَز - مِت - وِد - زِد - گِس - اِس - رُن - اُت - هُن -
 هاک - پان - هاب - یاخ - مان - راک - ناش - اوش - اور - تید - بیس - سید - میدیاه - هام - واش - نوک - هیک -
 دیم - یزم - کُرت - رِد - یِب - اِج - اِهَر - بَقَب - تِسِر - الر - رو - یُز .

جدول شماره ۱

SLENG	DRIP	GRUNG	RUB	SIV
GED	ZEK	KRUL	VET	TRIN
PLUT	KOV	SOC	SRUS	HEK
VLEZ	LOB	NUG	JAFT	CIST
PUD	STRAM	LAC	PSOS	FOR
SIF	STEF	MUST	FLES	BLAN
TORS	BIM	RAR	BUR	SPAS
GEP	SIL	FRAL	NOG	GLIM
DAFT	BREC	STOC	VOS	PRAC
TURS	MIZ	CAN	KLIT	SLUS

منابع مطالعه

- ۱- تاریخ زبان فارسی - دکتر پرویز ناتل خانلری جلد اول چاپ پنجم ۱۳۵۳ - انتشارات بنیاد فرهنگ ایران
- ۲- دستور امروز - دکتر خسرو فرشید ورد ۱۳۴۸
- 3- Fle-Echer-Speach and hearing.
- 4- Pirr-Chavasse, Essai sur la phonetique, annales des Telecommunication, Tome, 3. Nol 1948
- ۵- واژه نامه بسامدی دیوان عنصری- محمد نوری عثمانف چاپ ادبیات خاور مسکو ۱۹۷۰ میلادی
- ۶- مخارج الحروف یا اسباب حدوث الحروف - شیخ الرئیس ابوعلی سینا - تصحیح و ترجمه ناتل خانلری چاپ بنیاد فرهنگ ایران ۱۳۴۸
- ۷- توانایی زبان فارسی - دکتر محمود حسامی - نشریه انجمن فرهنگ ایران باستان مهرماه ۱۳۵۰
- 8- A-MALECOT-Fréquence d'occurrence des mots dans la conversation - revue d'Acoustique No 38-1976.
- ۹- دستور زبان فارسی - عبدالعظیم قریب چاپ نوزدهم ۱۳۲۶
- ۱۰- دستور زبان فارسی - دکتر طلعت بصری چاپ اول ۱۳۴۵
- ۱۱- دستور زبان فارسی - ع خیامپور چاپ دوم ۱۳۳۴
- ۱۲- دستور زبان فارسی - محمد جواد شریعت چاپ فردوسی ۱۳۴۵
- ۱۳- دستور زبان فارسی - قاسم انصاری انتشارات متین
- ۱۴- استوار نامه - دکتر محمد جواد شکور ۱۳۳۸
- ۱۵- برهان قاطع - محمد حسین بن خلف تبریزی متخلص به برهان ۱۰۶۲
- ۱۶- دستور زبان فارسی - دکتر محمد خزائلی و دیگران انتشارات متین
- ۱۷- دستور زبان فارسی - غلامحسین کاشف - سال ۱۳۲۸
- ۱۸- دستور جامع زبان فارسی استاد عبدالرحیم همایونفرخ
- ۱۹- دستور تعلیم الفباء باغچه بان (میرزا جبار عسگرزاده) سال ۱۳۱۴
- 20- VIARIS-L'art de déchiffrer les dépêches secrètes - ۲۰
- 21- G.FANT, KOMPENDIUM; TALÖVERFÖRING, Royal Institute of Tecknology- ۲۱
stockholm 1969.
- 22- VALÉRIO- Journal de Sciences militaires 1892. - ۲۲
- ۲۳- آواشناسی علی محمد حق شناس انتشارات آگاه ۱۳۵۶