

بررسی بسامد نویسه‌های زبان فارسی و مناسبت جایگاه آنها بر صفحه کلید رایانه‌ها

چکیده

نگارش متون فارسی به کمک رایانه یکی از رایج‌ترین اموری است که کاربران ایرانی در همه سطوح با آن سر و کار دارند. به نظر می‌رسد که نحوه قرار گرفتن کلیدهای صفحه کلید رایانه‌ها متناسب با میزان کاربرد آنها تنظیم نشده باشد. در توزیع حروف بر صفحه کلید رایانه‌ها ترتیب الفبایی و تشابه شکلی حروف بیشتر مدنظر قرار گرفته است. این پژوهش ضمن بررسی میزان بسامد حروف الفبای فارسی در پیکره‌ای بیش از یک میلیون و دویست هزار نویسه، مناسبت قرار گرفتن این حروف بر صفحه کلیدها را مورد بررسی قرار داده است. این پژوهش نشان می‌دهد که رابطه معناداری بین ترتیب الفبایی و رتبه بسامدی این نویسه‌ها وجود ندارد. در این بررسی بسامد حروف به صورت سه ردیف و هر ردیف به صورت یک نمودار مورد تجزیه و تحلیل قرار گرفته است. این پژوهش نشان می‌دهد که توزیع حروف الفبای فارسی بر صفحه کلید رایانه‌ها متناسب با میزان کاربرد آنها تنظیم نشده است و برخی از حروف پر بسامد و برخی از حروف کم بسامد در جایگاه نسبتاً یکسان و کنار هم قرار گرفته‌اند. در این پژوهش همچنین صفحه کلیدی متناسب با بسامد نویسه‌ها و بسامد انگشت‌گذاری به صورت آزمایشی و بر اساس داده‌های محدود پیشنهاد شده است.

کلیدواژه‌ها: صفحه کلید، زبان‌شناسی رایانه‌ای، الفبا، زبان‌شناسی پیکره‌ای، زبان فارسی.

۱- مقدمه

۱-۱- طرح مسئله تحقیق

شاید خود بارها متوجه این مهم شده باشیم که صفحه کلید رایانه‌ای که در منزل از آن استفاده می‌کنیم با رایانه دیگری که در دفتر کار استفاده می‌کنیم تا حدودی متفاوت است. همچنین از خود پرسیده باشیم که

کدامیک از این دو مناسب تر است و اساساً آیا هیچ کدام از این دو صفحه کلید به گونه‌ای که هر کلید متناسب با حرف و نویسه‌ای که به آن اختصاص یافته تنظیم شده است. به عبارتی آیا کلیدی که بیشتر در دسترس می باشد به حرفی که بیشترین کاربرد را دارد اختصاص یافته است و اساساً چگونه می توان صفحه کلیدی را طراحی کرد که نقص صفحه کلیدهای امروزی را نداشته باشد؟ این پژوهش می کوشد تا برای پرسش های بالا جوابی قانع کننده بر اساس پیکره ای از داده ها که در این جستار ارائه شده است، فراهم نماید.

ترتیب و تنظیم حروف الفبا بر دستگاه نگارش فارسی و همچنین صفحه کلید رایانه های امروزی از جمله مباحثی است که لازم است به صورتی قابل دفاع به آن پرداخته شود. این بحث با توجه به گسترش نگارش در فضای مجازی و رایانه ها و به خصوص با طرح نگارش بدون کاغذ^۱ اهمیتی مضاعف می یابد. بدیهی است تنظیم نویسه های فارسی بر خانه های صفحه کلید رایانه ها نمی تواند سلیقه ای و بدون هر گونه پشتوانه علمی و پژوهشی باشد. بدون شک روش علمی برخاسته از پژوهش کاربردی است که می تواند تناسب جایگاه هر نویسه بر صفحه کلید را با توجه به میزان کاربرد آن نشان دهد. در این راستا پژوهش حاضر می کوشد تا با طرح روشی علمی، الگوی مناسبی از صفحه کلید رایانه ارائه دهد.

بر همین اساس در این پژوهش با شمارش و تعیین میزان فراوانی هر نویسه، ارزش بسامدی هر حرف تعیین می گردد تا از این رهگذر با اطمینان بتوان جایگاه مناسب هر حرف را تعیین نمود. علاوه بر این با انجام آزمایشی از بسامد نسبی انتخاب خانه های صفحه کلید، مناسبیت انتخاب هر خانه رتبه بندی و با مقایسه آن با بسامد نسبی هر نویسه، صفحه کلیدی برای رایانه ها پیشنهاد گردیده است.

۱-۲- روش تحقیق

در این تحقیق ابتدا پیکره ای از متون روزنامه های کثیرالانتشار (کیهان، ایران، همشهری، جمهوری اسلامی، جام جم) مشتمل بر ۴۲۱۶۶۹ واژه و ۱۲۷۷۵۳۳ (یک میلیون و دویست و هفتاد و هفت هزار و پانصد و سی و سه) نویسه در مدت زمان اسفند ۱۳۸۸ و فروردین ۱۳۸۹ جمع آوری و با استفاده از نرم افزار واژه پرداز مورد شمارش و تحلیل قرار گرفت. متون انتخابی در این بررسی از همه بخش های روزنامه ها از جمله اخبار و تحلیل های سیاسی، اجتماعی، اقتصادی، و ورزشی و همچنین حوادث انتخاب

گردیده اند. همچنین با بررسی نحوه و ترتیب نویسه‌ها بر صفحه کلید رایانه‌ها معمول‌ترین آنها انتخاب و مورد بررسی قرار گرفته است. سپس در این پژوهش بسامد کلی همه نویسه‌ها در بیکره مورد نظر به تفکیک هر نویسه تعیین گردیده است. پس از آن بسامد نویسه‌ها در هر ردیف از کلیدهای صفحه کلید رایانه‌ها مورد تحلیل قرار گرفته است. همچنین به منظور طراحی صفحه کلید متناسب با بسامد نسبی نویسه‌ها، از دوازده نفر خواسته شد تا کلیدها را بدون توجه به نویسه‌های آن، هر کدام به اندازه یک صفحه ماشین‌نویسی کنند. سپس بسامد نسبی هر نویسه تعیین و بر حسب رتبه بسامدی کلیدی متناسب با آن اختصاص داده شده است.

۱-۳- پیشینه تحقیق

طراحی صفحه کلید رایانه (و همچنین تلفن‌های همراه) از جمله مواردی است که پژوهش‌های متعددی را در حوزه‌های مختلف به خود اختصاص داده است. بحث طراحی و اصلاح صفحه کلیدهای رایانه‌ای در زبان‌های مختلف پژوهش‌های متعددی صورت گرفته است، که از آن جمله می‌توان به راهی، هانتز و اسمیت (۲۰۰۲)، واکر (۲۰۰۳)، واگنر و دیگران (۲۰۰۳)، ستار، پتام و امیرعلی (۲۰۰۴)، ملاس، طیفور و ابنده (۲۰۰۸) اشاره کرد. راهی، هانتز و اسمیت (۲۰۰۲) دو عامل را در طراحی صفحه کلیدهای رایانه و گوشی همراه بسیار مهم می‌شمرند: یکی اینکه نویسه‌هایی که از بسامد بیش‌تری برخوردار است می‌بایست در مرکز صفحه قرار گیرد و دیگر اینکه نویسه‌هایی که بیش‌تر باهم می‌آیند کنار هم قرار گیرند. آنها بر این نکته نیز تاکید می‌ورزند که طرح‌هایی که تا کنون در این خصوص انجام گرفته است به بسامد نویسه‌ها توجه ویژه داشته‌اند.

ستار، پتام و امیرعلی (۲۰۰۴) که در خصوص صفحه کلید رایانه‌ها در زبان بنگالی پژوهشی انجام داده و در این خصوص صفحه کلیدی نیز طراحی نموده‌اند، بر این باورند که در صفحه کلیدهای موجود که شرکت‌های رایانه‌ای طراحی کرده‌اند به بسامد نویسه‌ها توجه کافی نشده است و یکی از امتیازات طرح خود را توجه به این مهم بر می‌شمرند.

والکر (۲۰۰۳) و همچنین گوتل و دیگران (۲۰۰۳) که از الگوریتم‌های متفاوتی به منظور اصلاح صفحه کلید رایانه انگلیسی استفاده کرده‌اند، بسامد نویسه‌ها مهم بر شمرده‌اند اگرچه به عوامل دیگری نیز

التفات داشته‌اند. اگنر و همکاران (۲۰۰۳) به بهینه‌سازی صفحه‌کلیدهای رایانه‌های انگلیسی، فرانسوی و آلمانی پرداختند و همچنین داشوال و دب (۲۰۰۳) نیز به طراحی و اصلاح صفحه‌کلید هندی بر اساس الگوریتم مخصوصی پرداختند، بسامد نویسه‌ها را مورد توجه قرار داده‌اند.

ملاس، طیفور و ابنده (۲۰۰۸) نیز که به تحلیل صفحه‌کلیدهای موجود عربی پرداخته و بر اساس یک الگوریتم فاصله، بسامد انگشت‌گذاری و تغییر دست، صفحه‌کلید مناسبی طراحی کرده‌اند، ادعا کرده‌اند که صفحه‌کلیدی که آنها طراحی کرده‌اند، ۳۵٪ درصد سریع‌تر از صفحه‌کلید رایج می‌باشد. آنها علاوه بر اینکه بسامد نویسه‌های عربی را در طرح خود لحاظ کرده‌اند، باهمایی نویسه‌ها را به صورت، دو به دو مد نظر قرار داده‌اند.

در پژوهش‌هایی که در خصوص طراحی صفحه‌کلید رایانه‌ها در سال‌های اخیر صورت گرفته است و در بالا به آنها اشاره شد، بر اهمیت ارزش بسامدی نویسه‌ها توجه کافی شده است.

در خصوص بررسی تناسب نویسه‌ها بر صفحه‌کلید رایانه‌های فارسی تا آنجا که نگارنده بررسی کرده تحقیق جامعی صورت نگرفته است. اما در خصوص بررسی بسامدی صداها و نویسه‌ها چند کار تحقیقی صورت گرفته است که از آن جمله می‌توان به موسوی تربتی (۱۳۴۶)، محمدی فر (۱۳۳۸) منصور (۱۳۸۵، ۱۳۸۸ و ۱۳۸۹) اشاره کرد. همه این پژوهش‌ها این واقعیت را نشان می‌دهند که صداها و نویسه‌ها از وضعیت کاملاً نامتوازنی برخوردار است. این عدم توازن تا حدی است که میزان کاربرد برخی دهها برابر برخی دیگر است.

شمار حروف الفبا در زبان فارسی به صورت متفاوت ذکر شده است. شمار این حروف را ۲۴ یا ۲۵، ۳۲ یا ۳۳ و ۳۵ تا ۳۷ نیز برشمرده‌اند. علت تفاوت در ارقامی چون ۲۴ یا ۲۵ و همچنین ۳۲ یا ۳۳ این است که برخی همزه را جزء حروف به حساب آورده‌اند و برخی نیاورده‌اند. علت تفاوت ۲۴ تا ۳۲ و ۲۵ تا ۳۳ این است که برخی ۸ حرف از حروف الفبا را عربی می‌دانند. در زبان فارسی علاوه بر حروف الفبا نشانه‌هایی وجود دارد که وضعیت آن باید در صفحه‌کلید رایانه مشخص شود. از جمله این نشانه‌ها می‌توان به نشانه‌های تنوین "، -، "، -، " نشانه‌های حرکت یا واکه فتنحه،

کسره و ضمه و هم چنین تشدید اشاره کرد. یکی دیگر از مسائل خط فارسی صورت‌های متفاوت برخی از حروف الفباست.

۲- بررسی و تحلیل داده‌ها

۲-۱- نگاهی به نحوه قرار گرفتن نویسه‌ها بر صفحه کلید

در سال‌های اخیر از طرف جمعی از متخصصین رایانه به همراه یکی از زبان‌شناسان صفحه کلیدی به صورت (۱) پیشنهاد شده است که توانسته است تأییدیه استاندارد صنعتی ایران با شماره ۲۹۰۱ را نیز کسب کند. در این صفحه کلید نیز تغییری که بتواند میزان کاربرد لحاظ شده باشد، دیده نمی‌شود. تنها تغییر قابل ذکر در این صفحه کلید جایگزین کردن "پ" به جای "همزه" است. از ۳۳ نشانه فارسی ۳۱ حرف به صورت گزینه اول و دو حرف "ژ" و "ء" به صورت گزینه دوم که با فشار دادن کلید تبدیل حاصل می‌شود، ارائه شده است. نشانه "آ" نیز در این صفحه کلید به صورت گزینه دوم نشان داده شده است.

(۱). ترتیب قرار گرفتن حروف به صورت گزینه اول (حالت عادی) هر کلید

حالت عادی

اتصال مجازی	۱	۲	۳	۴	۵	۶	۷	۸	۹	۰	-	=	\	پس‌بر	
جهش	ص	ض	ث	ق	ف	غ	ع	ه	خ	ح	ج	چ			
ورود	ش	س	ی	ب	ل	ا	ت	ن	م	ک	گ				
تبدیل	ظ	ط	ز	ر	ذ	د	پ	و	.	/				تبدیل	
مه‌ار	دگرساز	فاصله						دگرساز راست	مه‌ار						

در تجدید استاندارد ۲۹۰۱ مؤسسه ملی استاندارد اختصاص یافته به صفحه کلید فارسی دلایلی جهت تأیید این استاندارد آمده است که توجه به آن ضروری به نظر می‌رسد. در این تأییدیه از ویژگی‌های این صفحه کلید به اختصاص ردیف اول به اعداد و نشانه‌ها اختصاص جایگاه یکسان برای نمادهای مشابه انگلیسی و فارسی حفظ هماهنگی و تشابه با استاندارد ۸۲۰ (قبلی)، توجه به

بسامد حروف فارسی در متن‌های متعارف به منظور صرفه‌جویی در حرکت انگشتان، در کنارهم قرار دادن حروفی که دارای شباهت صوری هستند تا حد امکان مانند ر، ز، ژ اشاره شده است. بدیهی است که صفحه‌کلید استاندارد مانند برخی از صفحه‌کلیدهای دیگر دارای ویژگی‌های مهمی از جمله اختصاص دادن ردیف اول به اعداد و نشانه‌ها و رعایت یکسانی جایگاه‌های نمادهای مشابه انگلیسی و فارسی می‌باشد. اما همچنانکه در بالا اشاره شد یکی از ویژگی‌های صفحه‌کلید استاندارد توجه به میزان کاربرد و بسامد حروف فارسی بر صفحه‌کلید رایانه‌ها می‌باشد که در این تحقیق با استناد به حجم میلیونی داده‌ها مورد نقد قرار گرفته است که امید است با در نظر گرفتن نتایج این تحقیق و پیشنهادی که منتج از این پژوهش ارائه شده است، زمینه توجه جدی‌تر به این مهم با توجه به اهمیت نقش روزافزون رایانه‌ها در زندگی مردم فراهم شود. قبل از ارائه این پژوهش لازم است با نگاهی به نمادهای صفحه‌کلید فارسی ترتیب نسبی این نمادها را مورد توجه قرار دهیم.

آنچه که در قرار گرفتن حروف بر صفحه‌کلید جلب توجه می‌کند، چینش نسبی کلیدها بر اساس ترتیب حروف الفبای فارسی است. در ردیف اول چهار حرف "چ، ج، ح، خ" به ترتیب حروف الفبا پشت سرهم قرار گرفته است. به همین صورت دو حرف "ع" و "غ" دو حرف "ف" و "ق" و دو حرف "ص" و "ض" به ترتیب حروف الفبا و بدون توجه به تفاوت در میزان کاربرد آنها پشت سرهم قرار گرفته‌اند. در ردیف دوم تنها دو جفت "م" و "ن" و "س" و "ش" بر حسب حروف الفبا پشت سرهم قرار گرفته‌اند. دو حرف "گ" و "ک" اگرچه به ترتیب حروف الفبا نیستند ولی بی‌تاثیر از آن نیز نیستند.

در ردیف سوم نیز ترتیب الفبایی نویسه‌ها در چینش کلیدهای صفحه‌کلید بی‌تاثیر به نظر نمی‌رسد. در این ردیف سه جفت "د" و "ذ"، "ر" و "ز" و همچنین "ط" و "ظ" بر حسب ترتیب الفبایی چینش شده است. البته دلیلی که در استاندارد برای این موارد ذکر شده است رعایت کنارهم قرار گرفتن حروفی می‌باشد که به لحاظ شکل یکسان می‌باشند. اما آنچه که در خصوص این صفحه‌کلید دور از ذهن به نظر می‌رسد این است که صفحه‌کلید فارسی به طور همزمان دارای سه ویژگی مهم، یعنی رعایت نسبی ترتیب الفبایی، کنارهم قرار دادن حروف هم‌شکل و رعایت بسامد حروف

باشد، که حتی اگر اولی و دومی به طور تصادفی رعایت شده باشد، رعایت سومی بعید به نظر می‌رسد.

تحلیلی که در این قسمت ارائه شد نشان می‌دهد که چینش نویسه‌ها بر صفحه کلید نه بر حسب تصادف و نه بر حسب میزان کاربرد تنظیم شده، بلکه بیشتر سعی شده است از ترتیب الفبایی و تا حدی کنارهم قرار گرفتن حروف هم شکل استفاده شود که کم و بیش در ماشین‌های تایپ قدیم مد نظر قرار می‌گرفته است. اگرچه به نظر می‌رسد سعی شده است چینش نویسه‌ها متناسب با قرار گرفتن انگشتان دست باشد، ولی با توضیحی که در ادامه خواهد آمد نشان داده خواهد شد که ظاهراً این نکته نیز اساس نویسه‌ها نبوده است.

قبل از پرداختن به تناسب قرار گرفتن حروف بر صفحه کلید رایانه‌ها لازم است به تحلیل بسامدی کاربرد حروف الفبا پرداخته شود تا میزان کاربرد هر حرف مشخص شود.

۲-۲- تحلیل توزیع بسامدی نویسه‌ها

تحقیق حاضر نشان می‌دهد که نویسه‌ها در متون روزنامه‌ای توزیع برابری ندارند (نمودار ۲). نویسه "ا" با ۲۵۵ و ۲۴۴ میزان وقوع بیشترین فراوانی را به خود اختصاص داده است. نویسه‌های "د" و "ن" به ترتیب با ۱۱۵ و ۳۲۷ و ۱۱۴ و ۵۹۸ بار تکرار در رده‌های دوم و سوم قرار گرفته‌اند. کمترین میزان تکرار به نویسه "ژ"، با ۱۵۵۶ تکرار تعلق دارد. نویسه‌های "ظ" و "غ" به ترتیب با ۱۷۵۰ و ۱۷۸۳ تکرار در رده‌های بعدی قرار گرفته‌اند. پربسامدترین نویسه بیش از ۱۶۴ برابر کم بسامدترین نویسه بکار رفته است. این نشان می‌دهد که تا چه اندازه توزیع نویسه‌ها در متون مورد بررسی متفاوت است.

در صورتی که مجموع نویسه‌ها را در سه گروه یازده تایی الف (پر بسامد)، ب (متوسط) و ج (کم بسامد) تقسیم نماییم. به ترتیب "ا، د، ن، ه، و، م، ت، ر، ب، ل، ز" پر بسامد، "س، ش، خ، ج، ح، ف، گ، ع، پ، ق، ی" متوسط و "چ، ص، ک، ط، ذ، ض، ء، ث، غ، ظ، ژ" کم بسامد خواهند بود.

(۲) . بسامد حروف الفبا در پیکره مورد پژوهش

ردیف	حروف	تعداد	ردیف	حروف	تعداد
۱	ا	۲۵۵۲۴۴	۱۷	ص	۵۸۴۷
۲	ب	۷۰۲۰۰	۱۸	ض	۲۵۵۷
۳	پ	۱۰۵۴۷	۱۹	ط	۴۷۶۲
۴	ت	۷۹۰۹۵	۲۰	ظ	۱۷۵۰
۵	ث	۲۳۲۶	۲۱	ع	۱۰۹۸۱
۶	ج	۱۶۹۲۸	۲۲	غ	۱۷۸۳
۷	چ	۵۹۰۱	۲۳	ف	۱۲۱۷۶
۸	ح	۱۵۱۶۱	۲۴	ق	۹۹۸۴
۹	خ	۲۰۳۰۴	۲۵	ک	۵۱۲۹
۱۰	د	۱۱۵۳۲۷	۲۶	گ	۱۱۷۴۹
۱۱	ذ	۳۳۱۹	۲۷	ل	۴۲۳۵۰
۱۲	ر	۷۵۳۲۰	۲۸	م	۹۱۳۴۴
۱۳	ز	۳۴۷۶۵	۲۹	ن	۱۱۴۵۹۸
۱۴	ژ	۱۵۵۶	۳۰	و	۹۳۵۹۸
۱۵	س	۲۷۸۶۷	۳۱	ه	۱۰۴۲۴۵
۱۶	ش	۲۱۹۹۲	۳۲	ی	۶۴۱۱
			۳۳	ء	۲۴۱۷

بر اساس جدول فوق نویسه های "ا، د، ن، ه" در داده های این پژوهش بسامدی بیش صد هزار را دارا می باشند. و به عکس نویسه های "ث، چ، ژ، ذ، ص، ض، ط، غ، ق، ک، ی، ء" دارای بسامدی کمتر از ده هزار می باشند.

ترتیب این نویسه‌ها بر حسب رتبه بسامدی (از پربسامد به کم بسامد) به ترتیب عبارت خواهد

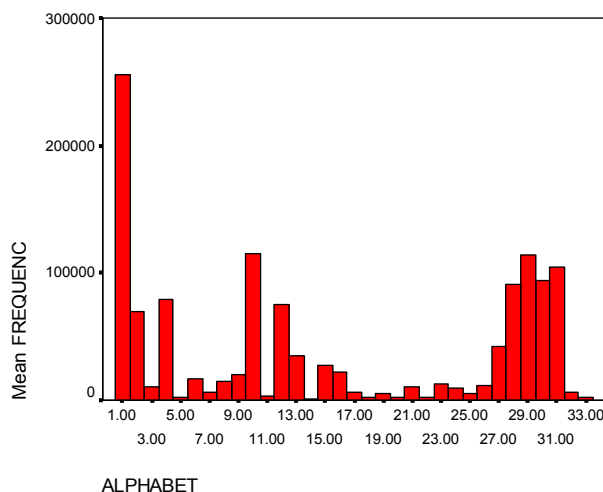
بود از:

۱-۱	۱۰-ل	۱۹-ع	۲۸-ض
۲-د	۱۱-ز	۲۰-پ	۲۹-ء
۳-ن	۱۲-س	۲۱-ق	۳۰-ث
۴-ه	۱۳-ش	۲۲-ی	۳۱-غ
۵-و	۱۴-خ	۲۳-چ	۳۲-ظ
۶-م	۱۵-ج	۲۴-ص	۳۳-ژ
۷-ت	۱۶-ح	۲۵-ک	
۸-ر	۱۷-ف	۲۶-ط	
۹-ب	۱۸-گ	۲۷-ذ	

ترتیب بسامدی حروف الفبای فارسی نشان می‌دهد که حرف "پ" در مرتبه بیستم قرار دارد. این بدان معناست که سیزده حرف دیگر وجود دارند که در رتبه‌های پایین‌تر می‌باشند. در طراحی صفحه کلیدها بعد از "ژ" این حرف بی‌ثبات‌ترین جایگاه را دارد. برخی آن را در اولین کلید سمت راست ردیف اول، برخی آن را در آخرین کلید این ردیف و برخی دیگر در آخرین کلید ردیف سوم قرار داده‌اند. حتی در برخی صفحه کلیدها دیده می‌شود که آن را در انتهای ردیف اعداد قرار داداند.

جدول (۳) نمودار بسامد حروف الفبا را به ترتیب الفبایی: «۱-ا، ۲-ب، ۳-پ، ۴-ت، ۵-ث، ۶-ج، ۷-چ، ۸-ح، ۹-خ، ۱۰-د، ۱۱-ذ، ۱۲-ر، ۱۳-ز، ۱۴-ژ، ۱۵-س، ۱۶-ش، ۱۷-ص، ۱۸-ض، ۱۹-ط، ۲۰-ظ، ۲۱-ع، ۲۲-غ، ۲۳-ف، ۲۴-ق، ۲۵-ک، ۲۶-گ، ۲۷-ل، ۲۸-م، ۲۹-ن، ۳۰-و، ۱-ح، ۳۲-ی، ۳۳-ء» نشان می‌دهد.

(۳). بسامدنویسه‌ها در پیکره پژوهش



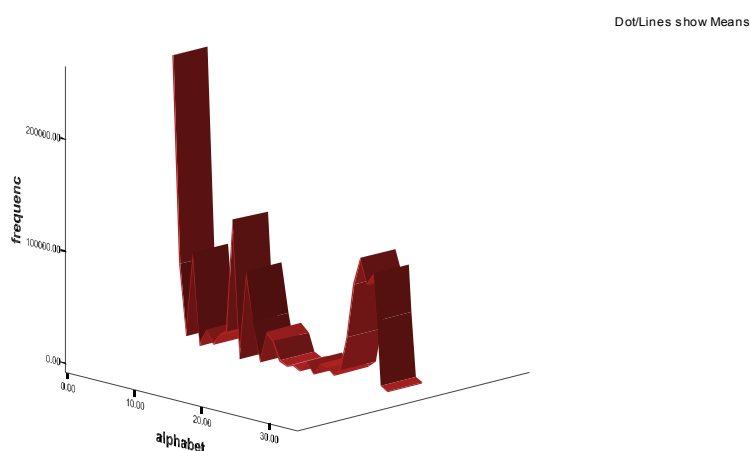
نکته دیگری که لازم است در تحلیل توزیع بسامدی نویسه‌ها مطرح شود این است که هیچ رابطه‌ای بین ترتیب حروف الفبای فارسی و توزیع بسامدی آنها وجود ندارد. مقایسه رتبه بندی نویسه‌ها بر اساس ترتیب معمول در حروف الفبا و رتبه بندی آنها بر اساس توزیع بسامدی نشان می‌دهد که هیچگونه رابطه معناداری بین این دو وجود ندارد. تحلیل آماری ضریب همبستگی رتبه‌ای اسپیرمن نشان می‌دهد که عدد همبستگی بسیار پایین (۰/۰۴ / چهار صدم) و عدد معناداری که کوچک بودن آن همبستگی بیشتر را نشان می‌دهد بسیار بالا و عدد ۰/۸۴ می‌باشد. این عدد همبستگی کمتری را حتی نسبت به حالت تصادفی نشان می‌دهد. دلایل گوناگونی را می‌توان برای این رخداد برشمرد که از جمله آنها می‌توان به تحول واژه‌ها در طول زمان و همچنین اقتباس خط از زبان دیگر (عربی) اشاره کرد.

(۴). ضریب همبستگی بین ترتیب حروف الفبا و رتبه بندی آنها

ضریب اسپیرمن	الفبا	رتبه	عدد معناداری
الفبا	۱	۰/۰۴	۰/۸۲۶
رتبه	۰/۰۴	۱	

این وضعیت را می‌توان به شکل قابل مشاهده به صورت نمودار (۵) نشان داد. این نمودار عدم ارتباط نظم الفبایی و رتبه بسامدی را نشان می‌دهد.

(۵). نمودار خطی بسامد حروف بر اساس ترتیب الفبایی



آمار توصیفی داده‌های فوق که به صورت (۶) ارائه شده است نشان می‌دهد پراکندگی قابل توجهی بین داده‌های این تحقیق وجود دارد.

(۶). آمار توصیفی

تعداد نویسه	حداقل	حداکثر	میانگین	انحراف معیار
۳۳	۱۵۵۶	۲۵۵۲۴۴	۳۸۷۱۳/۱۲	۵۳۸۹۷/۷۹

۴-۲- تحلیل قرار گرفتن حروف بر صفحه کلید

ترتیب قرار گرفتن نویسه‌ها بر صفحه کلید رایانه‌های امروزی برگرفته از ماشین‌های تحریر در گذشته است. در صفحه کلیدهای امروزی که اکثر متون نوشتاری امروز با آنها صورت می‌گیرد (بجز در یکی دو مورد) از الگویی مانند (۷) استفاده می‌شود.

(۷). ترتیب قرار گرفتن حروف بر صفحه کلید معمول

←	+	آ	۰	۹	۸	۷	۶	۵	۴	۳	۲	۱	ز
پ	چ	ج	ح	خ	ه	ع	غ	ف	ق	ث	ص	ض	Tab
ENTER	گ	ک	م	ن	ت	ا	ل	ب	ی	س	ش	Ctrl	Alt
Shift	؟	،	.	و	ء	د	ذ	ر	ز	ط	ظ	Shift	Space
Ctrl	Alt	Space						Alt	Ctrl				

در عموم صفحه کلیدهای امروزی اختلاف اساسا بر سر جایگاه "پ"، "ز" و تا حدودی "آ" می باشد. در برخی از این صفحه کلیدها جایگاه "ز" و "آ" به عنوان گزینه دوم "ز" و "آ" است. جایگاه "پ" نیز در برخی "منتهی الیه" سمت راست ردیف اول می باشد. این تفاوت ها در مجموع مشکل زیادی ایجاد نخواهد کرد؛ چرا که این نویسه ها خود از بسامد زیادی برخوردار نمی باشند و یا به دلیل اینکه شکل کم بسامدتر یکی از نویسه های دیگر می باشد، از بسامد زیادی برخوردار نیستند. در هر حال یکدست شدن برنامه تایپ فارسی که به صورت یک دست شدن صفحه کلیدهای رایانه ای خود را نشان می دهد ضرورتی غیر قابل انکار است. ترتیب قرار گرفتن نویسه ها بر صفحه کلید بر اساس یک قرار و به صورت استاندارد (۸) اگرچه این مشکل را حل می کند، ولی به نظر می رسد قبل از عمل به این قرار لازم است به تحلیل دقیق تر جایگاه هایی که می بایست به حروف اختصاص یابد پرداخت.

(۸). ترتیب حروف بر صفحه کلید استاندارد

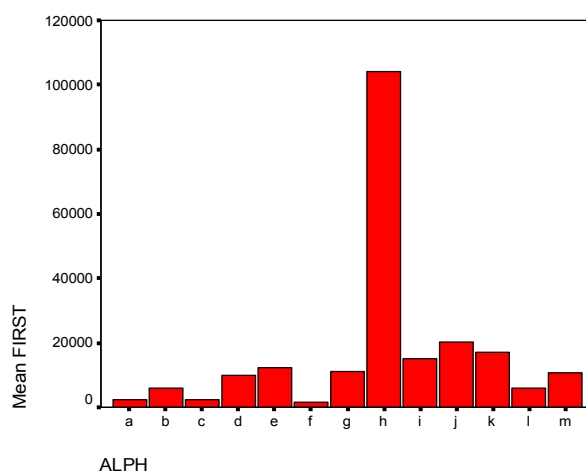
چ ج ح خ ه ع غ ف ق ث ص ض
گ ک م ن ت ا ل ب ی س ش
. / و پ د ذ ر ز ط ظ

اگرچه در طراحی صفحه کلید فوق سعی شده است ترتیب قرار گرفتن کلیدها که هرکدام متعلق به یک یا دو نویسه است به گونه ای باشد که با کارایی انگشتان دست متناسب باشد ولی به دلیل اینکه بیشتر این تشخیص بر اساس حدس و گمان بوده است با واقعیت میزان کاربرد این نویسه ها مطابقت ندارد. در این قسمت با استناد بر توزیع بسامدی نویسه ها که به صورت نمودارهای (۹)،

(۱۰) و (۱۱) نشان داده شده است استدلال خواهد شد که تا چه اندازه ترتیب قرار گرفتن نویسه‌ها با واقعیت توزیع بسامدی مطابقت دارد.

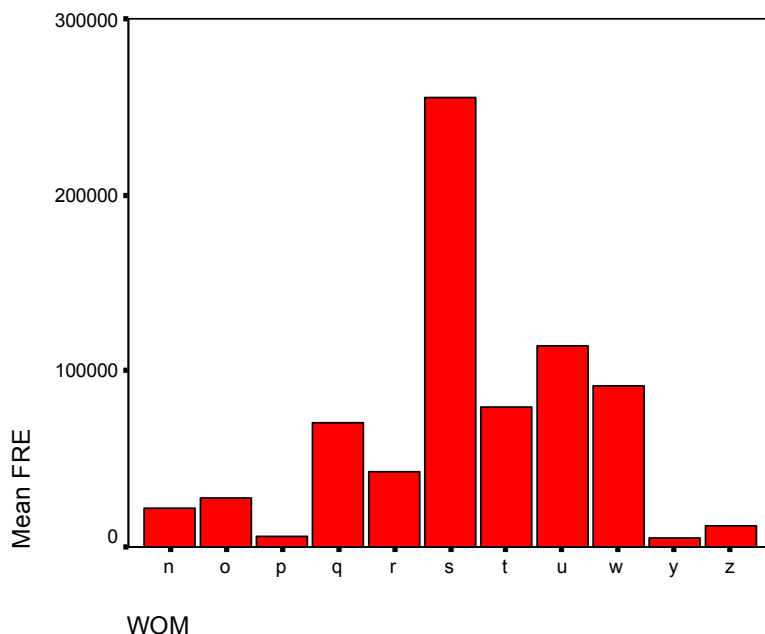
نمودار (۹) ردیف اول از بالا در صفحه کلید را نشان می‌دهد که از راست به چپ عبارتند از «پ، چ، ج، ح، خ، ه، ع، غ، ف، ق، ث، ص، ض». کلیدهای سمت راست زیر انگشتان راست و کلیدهای سمت چپ زیر انگشتان دست چپ در نظر گرفته شده‌اند. با توجه به اینکه عموم افراد راست دست می‌باشند، در مجموع، انگشتان دست راست و در صفحه کلید، کلیدهای سمت راست از اهمیت بیشتری برخوردارند. از تقسیم بندی سه گانه الف (پرسامد)، ب (متوسط) و ج (کم بسامد که در بالا ارائه شد. تنها یک نویسه در این ردیف متعلق به گروه الف می‌باشد. نمودار (۹) نشان می‌دهد که در مجموع بسامد نویسه‌هایی که با کلیدهای سمت راست نوشته می‌شوند بیشتر از نویسه‌های سمت چپ می‌باشند. اما در این ردیف نویسه «غ»، که از کم بسامدترین نویسه‌ها می‌باشد اندکی بعد از نویسه «ه» است که از پر بسامدترین نویسه‌ها می‌باشد. مقایسه میزان وقوع ۱۰۴۲۴۵ نویسه «ه» و ۱۷۸۳ نویسه «غ» که نشان می‌دهد نویسه «ه» بیش از ۵۸ برابر نویسه «غ» به کار رفته است نادرستی چنین چینی را در صفحه کلیدهای معمول را نشان می‌دهد. مهم‌تر اینکه اشکال اصلی تنها به چش در ردیف خلاصه نمی‌شود، بلکه برخی از کلیدهایی که در ردیف خاصی طراحی شده‌اند مناسب ردیف دیگر می‌باشند که در ادامه به آن اشاره خواهد شد.

(۹). نمودار ستونی بسامد حروف ردیف اول



ردیف دوم مهم‌ترین ردیف در صفحه کلید به حساب می‌آید، به این دلیل که به طور عادی و خستگی انگشتان دست روی این ردیف قرار می‌گیرند. بنابراین انتظار می‌رود که پربسامدترین نویسه‌ها با کلیدهای این ردیف نگاشته شوند. پربسامدترین نویسه که در داده‌های این پژوهش بالغ بر ۲۵۰۰۰۰ دویست و پنجاه هزار مرتبه تکرار شده است در این ردیف و درست زیر انگشت نشانه قرار گرفته است. همچنین نویسه‌های "ن، م، ت، ب، ل" به ترتیب در رتبه‌های سوم، ششم، هفتم، نهم و دهم در این ردیف قرار گرفته‌اند که حسن انتخاب پیشینیان را نشان می‌دهد. قرار گرفتن نویسه‌های پربسامد "ن، م، ت" زیر انگشتان دست راست نیز نشان از چینش درست این کلیدها دارد. همچنانکه نمودار (۹) نشان می‌دهد سمت راست این ردیف پربسامدتر از سمت چپ می‌باشد. در عین حال به نظر می‌رسد که نویسه‌های پربسامدتر مانند "د، و، ر" در این ردیف که مهم‌ترین ردیف به حساب می‌آید قرار نگرفته‌اند.

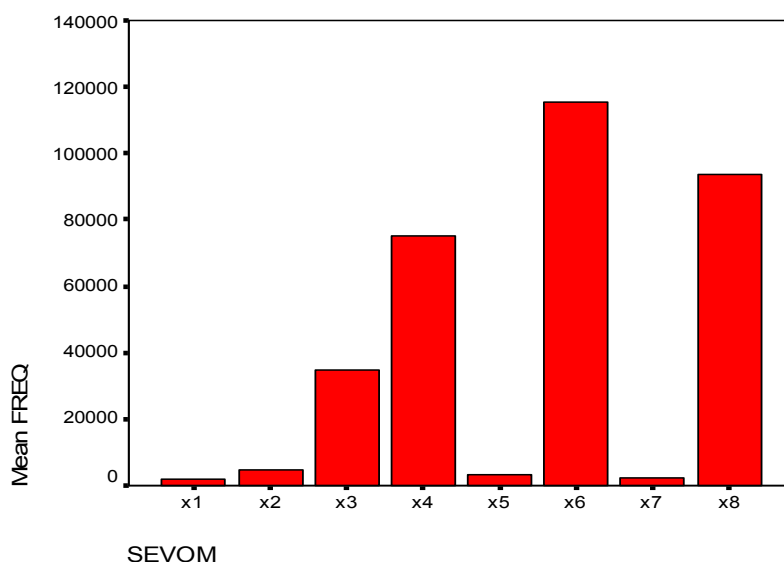
(۱۰). نمودار ستونی بسامد حروف ردیف دوم



ردیف سوم به لحاظ اینکه زیر انگشتان دست قرار نمی‌گیرد و از طرفی زیر کف دست تا حدودی پنهان می‌شود از اهمیت کمتری برخوردار است. بنابراین انتظار می‌رود که نویسه‌هایی که

به لحاظ رتبه از بسامد کمتری برخوردارند در این ردیف قرار گیرند. متناسب با این انتظار، نویسه‌های "ظ، ط، ذ، ء" در این ردیف و عموماً سمت چپ قرار گرفته اند که چینش درست را در این موارد نشان می‌دهد. اما برخلاف انتظار نویسه ,,د، با بیش از ۱۲۰۰۰۰ صد و بیست هزار تکرار و پر بسامدترین نویسه بعد از "ا" در این ردیف قرار گرفته است. نویسه های "ه" و "و" نیز که در رتبه های چهارم و پنجم به لحاظ بسامد می باشند در این ردیف قرار گرفته اند. نمودار (۱۱) این آشفتگی و عدم الگوی مشخص در چینش این ردیف را نشان می دهد.

(۱۱). نمودار ستونی بسامد حروف در ردیف سوم



تحلیلی که در این بخش از رابطه قرار گرفتن حروف بر صفحه کلید رایانه‌ها و میزان کاربرد حروف ارائه شد نشان می دهد که قرار گرفتن حروف متناسب با میزان کاربرد آنها بر صفحه کلید نمی باشد. به نظر می رسد که در چینش کلیدها شکل یکسان و ترتیب الفبایی حروف نقش اصلی را به عهده داشته است. البته چینش کلیدها بر اساس این دو ویژگی از امتیازهای این صفحه کلیدها محسوب می شود، ولی باید به خاطر داشت که یکسانی شکل و ترتیب الفبایی نسبت به میزان کاربرد در مرتبه پایین تری از اهمیت قرار دارد.

۲-۴- پیشنهاد صفحه کلید متناسب با بسامد نسبی

به منظور ارائه پیشنهاد صفحه کلید مناسب همچنانکه در بند ۱-۲ اشاره شد. از چند نفر از فارسی زبانان خواسته شد تا به هر گونه که می‌توانند بر کلیدها صفحه کلید رایانه در حد نگارش یک صفحه بدون توجه به حروف آنها فشار دهند، سپس داده‌های به دست آمده شمارش و تحلیل گردید. جدول (۱۲)، فراوانی به دست آمده را برای کلیدهایی که در صفحه کلید قرار دارند نشان می‌دهد.

جدول (۱۲). بسامد نسبی انگشت گذاری صفحه کلید رایانه

	ض	ص	ث	ق	ف	ه	خ	ح	ج	چ	پ
	۹۲	۳۶۸	۴۶۳	۳۴۹	۲۴۹	۴۵۱	۲۷۳	۱۸۰	۵۱	۲۰	۱۴
ش	س	ی	ب	ل	ا	ت	ن	م	ک	گ	
۱۴۸	۳۶۲	۵۸۷	۶۱۷	۶۱۸	۶۹۲	۷۷۹	۴۵۹	۳۸۲	۱۸۷	۶۱	
	ظ	ط	ز	ر	ذ	د	ئ	و	.	/	
	۱۹	۹۷	۲۶۷	۳۹۷	۲۲۴	۳۴۷	۱۸۴	۶۰	۲۸	۱۴	

بر اساس جدول فوق کلیدهایی که در حاشیه قرار گرفته اند کمتر انتخاب شده اند. برخلاف کلیدهای حاشیه ای کلیدهایی که در مرکز صفحه کلید قرار گرفته اند، از بسامد بیشتری برخوردارند. داده‌های جدول (۱۲) را می‌توان بر حسب رتبه بسامدی به صورت جدول (۱۳) نشان داد.

جدول (۱۳): رتبه بندی بر اساس انگشت گذاری

	۲۴	۱۱	۷	۱۳	۱۷	۸	۱۵	۲۱	۲۷	۲۹	۳۱
۲۲	۱۲	۵	۴	۳	۲	۱	۶	۱۰	۱۹	۲۵	
	۳۰	۲۳	۱۶	۹	۱۸	۱۴	۲۰	۲۶	۲۸	۳۲	

در طراحی صفحه کلیدی متناسب با ارزش بسامدی داده ها لازم است ارزش بسامدی نویسه‌هایی که از رورنامه گردآوری شده است با جدول (۱۳) مقایسه شود. به این صورت که

خانه ای که در جدول (۱۳) رتبه ۱ گرفته است به نویسه ای که بیشترین بسامد را دارد، اختصاص یابد. به همین ترتیب جایگاه سایر نویسه‌ها به این شکل تعیین می‌گردد. جدول (۱۴) صفحه کلید پیشنهادی را نشان می‌دهد.

جدول (۱۴). صفحه کلید پیشنهادی بر اساس بسامد انگشت گذاری

	ص	ز	ت	ش	ف	ر	ج	ق	ذ	ء	غ
ی	س	و	ه	ن	د	ا	م	ل	ع	ک	
	ث	چ	ح	ب	گ	خ	پ	ط	ض	ظ	

در خصوص جدول فوق لازم است به این نکته اشاره شود که این جدول بر اساس انگشت گذاری محدود صورت گرفته است، لذا در صورتی که آزمایش انگشت گذاری وسیع‌تری صورت گیرد، احتمال جابجایی برخی از خانه‌های جدول وجود دارد. بر این اساس می‌توان مدعی شد که این طرح صرفاً آزمایشی می‌باشد و به منظور پیشنهاد به عنوان یک طرح اساسی لازم است داده‌های انگشت گذاری در حجم وسیع‌تری گردآوری شود.

نکته دیگر اینکه علی‌رغم اهمیت بسامد انگشت گذاری، عوامل دیگری دخالت دارند که لازم است با رعایت بسامد نویسه‌ها مورد نظر قرار گیرند. یکی از این عوامل ترتیب حروف الفبا می‌باشد. یکی از مواردی که عموماً افراد تحصیل کرده همواره در ذهن خود دارند (اگرچه در برخی از افراد نه به طور کامل)، ترتیب حروف الفبا می‌باشد. در صورتی که بتوان هر دو عامل بسامد و ترتیب حروف الفبا در طراحی صفحه کلید در نظر گرفت بدون شک، صفحه کلید مناسب‌تری به دست خواهد آمد. به این منظور می‌توان در مواردی که تفاوت بسامد چندان قابل ملاحظه نیست، ترتیب حروف الفبا را در اولویت قرار داد.

نکته آخر اینکه نویسه‌های کم بسامدی که در صفحه کلید جای نمی‌گیرند و به صورت گزینه دوم مطرح است، حتی المقدور بر اساس تشابه شکلی که ترتیب الفبایی نیز می‌باشد، به عنوان گزینه دوم کلید مناسب تعریف شود. بر همین اساس در برخی از صفحه کلیدها نویسه "ز" به عنوان گزینه دوم "ز" تعریف شده است.

۳- نتیجه گیری

در این مطالعه استدلال شد که طراحی صفحه کلید رایانه فارسی بر اساس میزان کاربرد صورت نگرفته است. چینش حروف بر اساس یک سری از مزیت‌های جانبی مانند رعایت نسبی ترتیب الفبایی و شکل یکسان حروف هم‌نشین نمی‌تواند مبنای مناسبی برای عموم کاربران با توجه به لزوم گرایش هرچه بیشتر به "طرح حذف کاغذ" باشد. پژوهش فوق نشان می‌دهد که به منظور صرفه جویی در زمان و انرژی میلیون‌ها نفر از کاربران ایرانی و غیر ایرانی که با صفحه کلید رایانه ای فارسی کار می‌کنند، لازم است چینش حروف الفبای فارسی به گونه ای اصلاح شود که نویسه‌هایی که بیشترین بسامد را در متون نوشتاری دارند بیشتر در دسترس قرار گیرند و حروفی که از بسامد کمتری برخوردارند کمتر در دسترس واقع شوند. بر این اساس در این پژوهش پس از تعیین ارزش بسامدی نویسه‌های زبان فارسی، ارزش بسامدی انگشت گذاری‌های صفحه رایانه تعیین گردیده و بر اساس انطباق این دو ارزش صفحه کلیدی پیشنهاد شده است. با توجه به محدود بودن داده‌ها صفحه کلید پیشنهادی در این پژوهش جنبه مقدماتی داشته و با توجه به اهمیت موضوع، در صورتی که آزمایش گسترده ای به منظور ارزش بسامدی دقیق میزان انگشت گذاری برای هر کلید صورت گیرد، احتمال جابجایی برخی از نویسه‌ها نیز وجود دارد. در این پژوهش همچنین استدلال شد که در مواردی که تفاوت قابل ملاحظه ای بین ارزش بسامدی برخی از نویسه‌ها وجود نداشته باشد، می‌توان برخی از عوامل دیگر از جمله ترتیب الفبایی را در اولویت قرار داد.

کتابنامه

- الماس گنج. (۱۳۸۳). «نرم افزار بازشناسی گفتار پیوسته فارسی شنوا ۲». مجموعه سخنرانی‌ها، گزارش‌ها و چکیده طرح‌های اولین کارگاه پژوهشی زبان فارسی و رایانه. دانشکده ادبیات و علوم انسانی تهران. صص ۱۶۱-۱۶۲.
- آیت الهی، کاظم. (۱۳۸۳). «نرم افزار ترجمه متون از انگلیسی به فارسی و از فارسی به انگلیسی». مجموعه سخنرانی‌ها، گزارش‌ها و چکیده طرح‌های اولین کارگاه پژوهشی زبان فارسی و رایانه. دانشکده ادبیات و علوم انسانی تهران، صص ۶۳-۴۲.

- بحرانی محمد. (۱۳۸۲). «به کارگیری ساختارهای وابسته به بافت در بازشناسی گفتار پیوسته مبتنی بر مدل مخفی مارکوف». پایان نامه کارشناسی ارشد. دانشکده مهندسی کامپیوتر. دانشگاه صنعتی شریف.
- بهبهانی، مزده و عاصی، مصطفی (۱۳۸۶). «استفاده از شیوه نوشتار یک فارسی زبان در طراحی نرم افزار تشخیص دست نویس پیوسته فارسی». مجموعه مقالات هفتمین همایش زبان شناسی ایران. به کوشش محمد دبیر مقدم و همکاران، ۴۷۶-۴۸۹.
- بی جن خان، محمود. (۱۳۸۳). «هم نگاره های خط فارسی». مجموعه سخنرانی ها، گزارش ها و چکیده طرح های اولین کارگاه پژوهشی زبان فارسی و رایانه. دانشکده ادبیات و علوم انسانی تهران. صص ۶۳-۵۳.
- تیمور نژاد، کامران. (۱۳۸۳). «بررسی کاربردی خط فارسی از ایران باستان تا به امروز». مجموعه سخنرانی ها، گزارش ها و چکیده طرح های اولین کارگاه پژوهشی زبان فارسی و رایانه. دانشکده ادبیات و علوم انسانی تهران.
- خداپرستی، فرج الله. (۱۳۸۳). «ترجمه ماشینی فارسی». مجموعه سخنرانی ها، گزارش ها و چکیده طرح های اولین کارگاه پژوهشی زبان فارسی و رایانه. دانشکده ادبیات و علوم انسانی تهران. صص ۲۰-۱۴.
- رضانیا، کیانوش. (۱۳۷۶). بی ریزی طرح کلی واژگان و طراحی پیاده سازی پردازشگر ساختوازی برای زبان فارسی. پایان نامه کارشناسی ارشد. تهران: دانشگاه صنعتی شریف.
- روحی زاده، مسعود و همکاران. (۱۳۸۶). «طراحی شبکه واژگانی افعال زبان فارسی». مجموعه مقالات هفتمین زبانشناسی. به کوشش محمد دبیر مقدم. صص ۵۳۰-۵۱۸.
- دانش کار آراسته، پویان. (۱۳۸۱). نرم افزار تشخیص فعل در زبان فارسی. پایان نامه کارشناسی ارشد. تهران: دانشگاه علامه طباطبایی تهران.
- شجاع مودب، حمید رضا. (۱۳۸۳). «طراحی و پیاده سازی یک آشکار ساز فعال شدن صدا». مجموعه سخنرانی ها، گزارش ها و چکیده طرح های اولین کارگاه پژوهشی زبان فارسی و رایانه. دانشکده ادبیات و علوم انسانی تهران.
- صفایی مهران. (۱۳۸۵). «بکارگیری آرایه میکروفن به منظور بهبود دقت سیستم های بازشناسی گفتار». پایان نامه کارشناسی ارشد. دانشکده مهندسی کامپیوتر. دانشگاه صنعتی شریف.
- طیبی، اکرم. (۱۳۷۴). «کاربرد دستور واژگانی نقشمند در ترجمه ماشینی پاره ای از متون زبان فارسی». پایان نامه کارشناسی ارشد. دانشگاه تهران.

عبادت، علیرضا. (۱۳۸۳). «مترجم پارس: نرم افزار ترجمه متن انگلیسی به جمله های فارسی» مجموعه سخنرانی‌ها، گزارش‌ها و چکیده طرح‌های اولین کارگاه پژوهشی زبان فارسی و رایانه. دانشکده ادبیات و علوم انسانی تهران، صص ۴۱-۳۱.

عاصی، مصطفی. (۱۳۷۶). رایانه و استاندارد سازی در زبان. چهارمین کنفرانس زبان شناسی. دانشگاه علامه طباطبایی ۲۵-۲۳ اسفند.

فاروقی، جلیل اله و نادر جهانگیری. (۱۳۸۷). «استفاده از دستور واژی-تقشی در طراحی ماشین ترجمه انگلیسی به فارسی». مطالعات ترجمه. ش ۲۱. صص ۴۷-۲۵.

محمدی فر، محمد رضا. (۱۳۶۸). «زبان و اطلاعات». مجله زبان شناسی. س ۶ ش ۱: ۲-۲۲.

ممتازی، سعیده و فاضل، مریم (۱۳۸۴). پارس مقاوم در سیستم‌های بازشناسی گفتار پیوسته. پایان نامه کارشناسی. دانشکده مهندسی کامپیوتر. دانشگاه صنعتی شریف.

منصوری، مهرزاد. (۱۳۸۵). «اسامی زنان ایلامی در آیین نامگذاری در ایران». مجموعه مقالات زنان و چالش‌های پیش رو، ۳۶۷-۳۵۷.

----- (۱۳۸۱). نظام بسامدی واج‌ها در زبان فارسی. ایلام: انتشارات دانشگاه ایلام.

----- (۱۳۸۹). «تحلیل بسامدی هجا در اسامی زنان و مردان در ایران». مجله پژوهش‌های زبان شناسی.

سال دوم. شماره اول. صص ۹۶-۸۱

موتق، حامد. (۱۳۸۱). طراحی و پیاده سازی روش جستجوی بهینه برای بازشناسی گفتار پیوسته فارسی با مدل مخفی مارکوف. پایان نامه کارشناسی ارشد. دانشکده مهندسی کامپیوتر. دانشگاه صنعتی شریف.

موزن، علیرضا. (۱۳۷۸). «بررسی مسائل جستجوی موضوعی در فهرست‌های رایانه‌ای با نگاهی به مشکلات رسم‌الخط زبان فارسی». همایش کاربرد و توسعه فهرست‌های رایانه‌ای در کتابخانه‌های ایران.

موسوی تربتی، سعید. (۱۳۴۶)^۱. بسامد کاربرد واج‌ها در زبان فارسی نوشتاری. پایان نامه کارشناسی ارشد. دانشگاه تهران.

۱. هیچ گونه تاریخی در خصوص دفاع از این پایان نامه بر روی جلد و در پایان نامه ذکر نشده است همچنین در اینترنت و پایگاه کتابخانه دانشگاه تهران بدون تاریخ ثبت شده است، ولی باتوجه به داده‌های این پایان نامه که از نشریات ۱۳۴۶ می باشد، به نظر می رسد که سال دفاع ۱۳۴۶ باشد.

موسوی میانگاه، طیبه. (۱۳۸۵). «ترجمه ماشینی و ترجمه انسانی، در تکامل نه در تقابل.» مطالعات ترجمه. ش ۱۳. صص ۴۰-۳۱.

----- (۱۳۸۶). درآمدی بر ترجمه ماشینی. تهران: انتشارات یلدا قلم.

هاشمی گلپایگانی، محمدرضا و دیگران. (۱۳۷۳). «بازشناسی مرزهای نحوی جملات زبان فارسی با استفاده از شبکه‌های عصبی ترکیبی.» سومین کنفرانس سالانه انجمن کامپیوتر ایران، دانشگاه علم و صنعت ۴-۲ دی. ویسی پور، سامان. (۱۳۸۵). «ایجاد توانایی تشخیص کلمات خارج از واژگان به منظور بهبود عملکرد یک سیستم بازشناسی گفتار فارسی.» پایان نامه کارشناسی ارشد. دانشکده مهندسی کامپیوتر. دانشگاه صنعتی شریف. ویسی، هادی. (۱۳۸۴). «روش‌های مبتنی بر مدل برای سیستم‌های بازشناسی گفتار مقاوم به نویز.» پایان نامه کارشناسی ارشد. دانشکده مهندسی کامپیوتر. دانشگاه صنعتی شریف.

ویسی، هادی، نامتی، حسین، حسین زاده، خسرو. (۱۳۸۵). «مقاوم‌سازی سیستم تشخیص گفتار پیوسته فارسی نویسا به نویز محیط و صدای گوینده.» دومین کارگاه پژوهشی زبان فارسی و رایانه. دانشگاه تهران، تهران-ایران، تیرماه

Babaali B., Sameti H. (2004). *The Sharif Speaker-Independent Large Vocabulary Speech Recognition System . The 2nd Workshop on Information Technology & Its Disciplines. (WITID 2004. Feb. 24-26, 2004, Kish Island, Iran.*

Bahrani M., Sameti H., Hafezi N., Movasagh H. (2006). *Building and Incorporating Language Models for Persian continuous speech Recognition system , The fifth international conference on Language Resources and Evaluation (LREC). Genoa. Italy.*

Bakhshandeh, E. and P. Mosallanejad .(2006). "Computer-Aided, Machine Translation in the Age of Communications". *Translation Studies*. 14: 29-46

Megerdooian, K. (2004). "Developing a Persian Part of Speech Tagger" *Proceedings of the 1st Workshop On Persian Language and Computer , the faculty of Letters and Humanities. Tehran University. Tehran- Iran*

Mirzaeian, V. R. (2010). "Challenges of Machine Translation in Persian. Using Three MT Systems" *Translations Studies*. 28: 73-90.

Movasagh H., Sameti H. (2005). Two-step synchronous phoneme based search for continuouse speech Recognition. *First International Conference On Modeling, Simulation and Applied Optimization (ICMSAO'05). Sharjah.UAE.*