

عمق خط فارسی

الهام علایی ابوذر^۱

دانشجوی دکتری زبان‌شناسی دانشگاه تهران

محمود بی‌جن خان

دانشیار دانشگاه تهران

(از ۱ ص تا ۱۹ ص)

تاریخ دریافت مقاله: ۹۱/۵/۱ تاریخ پذیرش قطعی: ۹۱/۹/۱۵

چکیده

هدف از این مقاله تبیین فاصله بین کلمات نوشتاری و صورت واجی متناظر با آنها در خط و زبان فارسی در چارچوب مفهوم «عمق خط» است. برای این منظور رابطه بین صورت نوشتاری و واجی کلمات با استفاده از قواعد آزاد و وابسته به بافت نگاره- واج صورت‌بندی می‌شود. سپس مبانی واج‌شناختی نگاره-های زیروزبری به عنوان ابزار نوشتاری برای کاهش عمق خط فارسی تبیین می‌شود و در خاتمه شیوه محاسبه عمق خط فارسی، فارغ از اطلاعات زبانی، توضیح داده می‌شود.

واژه‌های کلیدی: عمق خط، صورت نوشتاری، صورت واجی، نگاره، نگاره‌های زیروزبری، اطلاعات زبانی.

۱- مقدمه

رویکرد غالب در مطالعه نظام‌های نوشتاری، چگونگی پیوند دادن زبان با نشانه‌های نوشتاری است. در سال‌های اخیر نظام نوشتاری به عنوان حوزه‌ای مجزا در مطالعات زبان‌شناسی از اهمیت خاصی برخوردار شده است و مطالعات بسیاری در زمینه نظام‌های نوشتاری گوناگون انجام شده است که از میان آنها می‌توان به تحقیقات کارینه مگردومیان (۲۰۰۴) تجزیه و تحلیل ساختواژی روش حالت‌های محدود در فارسی، مریم دانای طوس (۱۳۸۴) اثرات شفافیت و تیرگی خط فارسی بر مهارت زبانی خواندن کودکان طبیعی و خوانش‌پیش‌رشدی فارسی زبان، تیم بوکوالتر^۱ (۲۰۰۴) مسائلی درباره خط عربی و تحلیل‌های ساختواژی مربوط به آن، آنتال ون دن بوش^۲، الن کانتنت^۳، والتر دلمانز^۴ و بآتریس دخلدر^۵ (۱۹۹۴) تجزیه و تحلیل عمق نوشتاری زبان‌های مختلف با استفاده از الگوریتم داده-محور، الینرگیبسون^۶، انه پایک^۷، هری اوسر^۸ و مارشا همند^۹ (۱۹۶۲)- نقش انطباق نگاره-واج در فهم کلمات، ریچرد ونزکی^{۱۰} (۲۰۰۴) در جستجوی نظام نوشتاری جامع و رابرت ترمل^{۱۱} (۱۹۹۰) تطابق مختلف نگاره-واج در کلمات ناآشنای چندهجایی، اشاره کرد. به دلیل وجود انبوه اطلاعات موجود در محیط‌های گوناگون از جمله محیط‌های رایانشی مانند اینترنت و رسانه‌ها مانند روزنامه و مجلات، حجم کلماتی که در واژگان ذهنی وجود ندارند و برای اولین بار با آنها مواجه می‌شویم (کلمات خارج از واژگان^{۱۲}) افزایش می‌یابد. حجم قابل توجهی از این کلمات، نام‌های تجاری هستند که حرف‌گردانی^{۱۳} شده‌اند مانند کلمه < ناسیونال > برای برخی دیگر فرهنگستان زبان فارسی معادل فارسی معرفی کرده است مانند کلمه < رایانه >

¹.Tim Buckwalter

².Antal Van den Bosch

³.Alain Content

⁴.Walter Daelemans

⁵.Beatrice de Gelder

⁶.Eleanor J. Gibson

⁷.Anne Pick

⁸.Harry Osser

⁹.Marcia Hammond

¹⁰.Richard Venezky

¹¹.Robert L. Trammel

¹².out of vocabulary words

¹³.transliteration

که معادل کلمه < کامپیوتر > می‌باشد. این کلمات جدیدالورود در متون فارسی را می‌توان از دو دیدگاه مطالعه کرد: (۱) روانشناسی زبان و (۲) پردازش متن فارسی از دیدگاه تبدیل نوشتار به گفتار. روانشناسی زبان در جستجوی یافتن پاسخ برای این مسئله است که علیرغم نبود این کلمات در واژگان ذهنی، چگونه گویشوران یک زبان قادر به تلفظ کلماتی که برای اولین بار با آنها مواجه می‌شوند، هستند و علت وجود شباهت‌ها در نحوه تلفظ این کلمات میان افرادی که در یک طبقه تحصیلاتی قرار می‌گیرند چیست. هدف این مقاله مطالعه مسائل مربوط به خط فارسی از نظر پردازش متن فارسی از دیدگاه تبدیل نوشتار به صورت واجی و سپس به صورت آوایی است. از آنجائی که کلمات جدیدالورود در واژگان تعریف نشده‌اند، در پردازش متن فارسی از دیدگاه تبدیل نوشتار به صورت واجی و سپس به صورت آوایی، باید روشی پیش‌بینی شود که بوسیله آن بتوان تلفظی برای این کلمات ارائه داد. نگارنده با استفاده از مفهوم نگاره^۱ فاصله بین صورت نوشتاری کلمات فارسی و صورت آوایی را تبیین می‌کند و از طریق ارائه روش

ریاضی بخش‌بندی قطعه‌های نوشتاری و فرمول
$$\frac{\sum_{s \in S} \left(v(s) \times \frac{\pi(s)}{\alpha(s)} \right)}{\sum_{t \in S} v(t)}$$
، که برای اولین

بار نگارنده این روش را معرفی می‌کند، به محاسبه عمق خط^۲ فارسی می‌پردازد. عبارت «عمق خط» بر میزان فاصله‌ای که یک نظام نوشتاری از یک نظام آوانگاشتی آرمانی دارد دلالت می‌کند. هر قدر این فاصله کمتر باشد و عدد به‌دست‌آمده به یک نزدیک‌تر باشد، نظام نوشتاری مورد نظر سطحی‌تر است و برعکس هر قدر این فاصله بیشتر باشد، نظام نوشتاری مورد نظر عمیق‌تر است. طبق محاسبات بعمل‌آمده، عمق خط فارسی زیاد است و به منظور تبدیل نوشتار به صورت واجی و سپس صورت آوایی، باید از اطلاعات زبانی موجود در خط فارسی بهره‌مند شد. حاصل چنین پژوهشی می‌تواند در طراحی یک واژگان مناسب برای یک نظام تبدیل نوشتار به گفتار فارسی مفید باشد.

در تحقیق حاضر ابتدا به تفاوت میان حرف^۳ و نگاره و مبانی واج‌شناختی نگاشت نگاره به واج در زبان فارسی، به منظور مطالعه اطلاعات واجی موجود در نظام نوشتاری

^۱.grapheme

^۲.orthographic depth

^۳.letter

زبان فارسی، پرداخته می‌شود. سپس در بخش سوم مفهوم عمق خط معرفی می‌شود. بخش چهارم محاسبه عمق خط فارسی و مباحث مربوط به آن می‌باشد و در نهایت بخش پنجم نتیجه‌گیری مقاله خواهد بود.

۲- حرف و نگاره

با وجود مفهوم «حرف» در نظام نوشتاری، در تحقیق حاضر از مفهوم «نگاره» استفاده می‌شود. نگاره، کوچکترین واحد نوشتاری است که میان یک جفت کمینه نوشتاری تمایز ایجاد می‌کند، (راجرز^۱ ۲۰۰۵: ۱۰). در حالیکه مفهوم «حرف» به مجموعه نگاره‌هایی اطلاق می‌شود که برای نمایش همخوان‌ها و بعضی از واژه‌ها به کار می‌رود و به لحاظ سنتی «الفبا» نامیده می‌شوند. با طرح مفهوم نگاره می‌توان گفت اعداد، علائم نقطه‌گذاری و علائم زیرویری^۲ نیز نگاره محسوب می‌شوند. به عنوان مثال جفت‌های کمینه نوشتاری مانند: «تار» و «نار» تنها در تعداد نقطه‌ها از هم متمایز می‌شوند؛ اگر یک نقطه از قطعه نوشتاری «تار» کم شود تبدیل به قطعه نوشتاری «نار» می‌شود. علائم زیرویری مانند «تنوین»، «مد» و «تشدید» نیز نگاره محسوب می‌شوند. به عنوان مثال در جفت‌های کمینه نوشتاری «بنا» و «بنا»، «اجر» و «اجر»، «حکماً» و «حکماً»، به ترتیب، حضور یا عدم حضور نگاره‌های تشدید، مد و تنوین چنین جفت‌های کمینه‌ای را ایجاد کرده است. بنابراین برای تبدیل نوشتار به گفتار یا حرف به آوا باید از مفهوم نگاره استفاده کرد تا بتوان چنین تمایزاتی را نشان داد؛ در حالی که اگر تنها به مفهوم «حرف» اکتفا شود، نمی‌توان چنین تمایزاتی را نشان داد. به عنوان مثال «ا» یک حرف است اما «أ» حرف نیست بلکه یک نگاره است که به رشته آوایی /a?/ نگاشته می‌شود مانند: «تأیید» /ta?jid/؛ همچنین است نگاره «خو» در ابتدای کلمه که در بسیاری از موارد به واج /x/ نگاشته می‌شود مانند: «خواب» /xɒb/. بنابراین، به کار بردن مفهوم «نگاره» در کنار مفهوم «حرف» به این دلیل است که هدف تبدیل نوشتار به صورت واجی و سپس صورت آوایی است. مفهوم «نویسه»^۳ به رمز رایانه‌ای مربوط می‌شود که برای بازنمایی نگاره‌ها و پردازش آنها در

^۱.Rogers

^۲.diacritics

^۳.character

حافظه رایانه از آن استفاده می‌شود. از آنجایی که تحقیق حاضر رویکردی واج‌شناختی به خط فارسی دارد ذکر انواع نگاره در نظام نوشتاری لازم به نظر می‌رسد. در نظام‌های نوشتاری نگاره‌ها را می‌توان به انواع: نگاره‌های آزاد^۱، نگاره‌های وابسته^۲، نگاره‌های همخوانی^۳ و نگاره‌های واکه‌ای^۴ تقسیم کرد. نگاره‌های آزاد به طور مستقل ظاهر می‌شوند و در بیشتر موارد قابل انطباق با حروف هستند. به عنوان مثال در نظام نوشتاری زبان فارسی در کلمه «امید» نگاره‌های <ا>، <م>، <ی> و <د> نگاره‌های آزاد محسوب می‌شوند زیرا در بافت‌های دیگر (در کلمات دیگر) هر کدام از این نگاره‌ها به طور مستقل ظاهر می‌شوند. نگاره‌های وابسته، که در بعضی از نظام‌های نوشتاری از جمله نظام نوشتاری زبان فارسی وجود دارند، نگاره‌هایی هستند که تنها در ترکیب با دیگر نگاره‌ها ظاهر می‌شوند و به طور مستقل در خط ظاهر نمی‌شوند. به عنوان مثال نگاره مد <~> در نظام نوشتاری زبان فارسی به تنهایی ظاهر نمی‌شود و حتماً همراه با نگاره الف <ا> به صورت <آ> ظاهر می‌شود. نگاره‌های وابسته نوعاً اختیاری‌اند اما در بیشتر موارد حضورشان باعث رفع ابهام از هم‌نگاره‌ها^۵ می‌شود. نگاره‌های همخوانی برای نمایش همخوان‌های موجود در نظام واجی به کار می‌روند مانند: نگاره <پ> که برای نمایش همخوان /p/ به کار می‌رود. نگاره‌های واکه‌ای برای نمایش واکه‌های موجود در نظام واجی به کار می‌روند مانند: نگاره فتحه <َ>. نگاره‌های واکه‌ای اغلب اختیاری‌اند اما حضورشان باعث رفع ابهام از هم‌نگاره‌ها می‌شود. بیشتر زبان‌شناسان معتقدند در یک نظام نوشتاری آرمانی، رابطه یک به یک میان نگاره و واج وجود دارد. اما به دلیل نگاشت‌های چندبه‌یک یا یک‌به‌چند بین نگاره‌ها و واج‌ها، وجود یک نظام نوشتاری آرمانی بعید به نظر می‌رسد. در نظام نوشتاری زبان فارسی واکه‌های کوتاه معمولاً نمود نوشتاری ندارند و تنها تعداد انگشت‌شماری نگاره وجود دارد که برای نشان دادن واکه‌های کوتاه در خط به کار می‌روند؛ از جمله «های بیان حرکت» (<ه>) که برای نشان دادن واکه کوتاه /e/ یا /a/ به کار می‌رود؛ مانند <به> و <نه>؛ و حرف <و> که می‌تواند نمود نوشتاری واکه کوتاه /o/ باشد؛ مانند <تو>. از دیگر مسائلی که در نظام

^۱.free graphemes

^۲.bound graphemes

^۳.consonantal graphemes

^۴.vowel graphemes

^۵.homographs

نوشتاری زبان فارسی حائز اهمیت است، وجود علائم زیروزبری^۱ است که به عنوان نگاره های وابسته محسوب می‌شوند. در نظام نوشتاری زبان فارسی ۱۱ علائم زیروزبری وجود دارد که می‌توان آنها را در پنج طبقه واج‌شناختی دسته‌بندی کرد.

۲-۱- طبقه‌بندی نگاره‌های زیروزبری در نظام نوشتاری زبان فارسی

۲-۱-۱- واکه‌سازی^۲: علائم زیروزبری فتحه یا زبر <َ -َ >، ضمه یا پیش <ُ -ُ > کسره یا زیر <ِ -ِ >، مد < ~ > و الف مقصوره < ِ > در این طبقه قرار می‌گیرند. نتیجه واج‌شناختی ترکیب این علائم با نگاره‌های آزاد این است که آن نگاره آزاد توسط این علائم واکه‌سازی می‌شود.

نگاره همخوانی که با فتحه، کسره و ضمه تلفیق می‌شود به ترتیب بوسیله واکه‌های کوتاه /a/، /e/ و /o/ آواسازی می‌شود. از کسره می‌توان به عنوان پرکاربردترین علامت زیروزبری در واکه‌سازی نام برد که کلمات را برای ساختن سازه‌های بزرگتر نحوی - که ساختار اضافه^۳ نامیده می‌شود - به هم وصل می‌کند. این نوع کسره، کسره اضافه خوانده می‌شود.

در برخی موارد به منظور ابهام‌زدایی از هم نگاره‌ها و در نتیجه فهم بهتر متن می‌توان از واکه‌سازی واکه کوتاه استفاده کرد. به عنوان مثال در یک متن اقتصادی، قطعه نوشتاری <ارزش> می‌تواند دو معنی کاملاً متفاوت داشته باشد: <ارزش> /ʔarzeʃ/ (اسم) و <ارزش> /ʔarzaʃ/ (اسم + ضمیر ملکی سوم شخص مفرد)؛ این مثال کاربرد نگاره وابسته فتحه و کسره را در رفع ابهام از هم‌نگاره <ارزش> نشان می‌دهد.

نوع دیگر واکه‌سازی که در متون فارسی نمونه‌های فراوانی از آن یافت می‌شود، واکه‌سازی نگاره «الف» < ِ > است که در بیشتر موارد در جایگاه ابتدائی قطعه نوشتاری قرار می‌گیرد و به شکل مد < ~ > بالای نگاره الف < ا > ظاهر می‌شود < آ > که در واقع نمود نوشتاری واکه بلند پسین افتاده /ɒ/ است؛ بنابراین، نشانه < آ > توالی واجی /ʔɒ/ است

^۱.diacritics

^۲.vocalization

^۳.Ezafé construction

الف مقصوره <ا> در بالای حرف <ی> در جایگاه انتهایی دنباله نوشتاری، نوع دیگر واکه‌سازی است که تنها در کلمات قرضی عربی یافت می‌شود؛ در این صورت نگاره همخوانی قبل از <ی> بوسیله /v/ واکه‌سازی می‌شود. مثال‌های <عیسی> در مقابل <عیسی>، <موسی> در مقابل <موسی> و <یحیی> در مقابل <یحیی> نشان می‌دهند که الف مقصوره <ا> یک نگاره وابسته است.

۲-۱-۲- تنوین^۱: ترکیب علائم زیروزبری دو زبر <ـَـ>، دو پیش <ـِـ> و دو زیر <ـِـ> با یک نگاره همخوانی، تنوین نامیده می‌شود. نتیجه واج‌شناختی چنین ترکیبی، واکه‌سازی نگاره همخوانی بوسیله یک واکه کوتاه و همخوان خیشومی /n/ است؛ بنابراین، ترکیب‌های <ان> /an/، <ان> /en/ و <ان> /on/ ساخته می‌شوند. رایج‌ترین نوع تنوین در متون فارسی، دو زبر <ان> است که به انتهای اسم یا صفت اضافه می‌شود و در نتیجه قیده‌های واژگانی ساخته می‌شوند. برای اثبات وابسته بودن نگاره دو زبر <ـِـ> می‌توان به تعداد زیادی از جفت‌های کمینه نوشتاری اشاره کرد مانند: <عرفاً> /ʔorfan/ در مقابل <عرفا> /ʔorafɒ/، <عقلاً> /ʔaGlan/ در مقابل <عقلا> /ʔoGalɒ/، <فرداً> /fardan/ در مقابل <فردا> /fardɒ/ . بنابراین، در چنین مواردی اگر دو زبر-جایی که حضور آن الزامی است- حذف شود، ابهام واژگانی را در تبدیل نگاره به واج ایجاد می‌کند و یا به عبارت دیگر، به میزان عمق خط فارسی افزوده می‌شود. حال آنکه در مورد بعضی قیود دیگر چنین مسأله‌ای صدق نمی‌کند؛ به عبارتی دیگر در صورت عدم حضور دو زبر، ابهام واژگانی رخ نمی‌دهد و تنها باعث تفاوت در تلفظ یک کلمه می‌شود مانند: <حقاً> /haGGan/ و <حقا> /haGGɒ/، <ابدأ> /ʔabadan/ و <ابدا> /ʔabadɒ/، <صلاً> /ʔaslan/ و <صلا> /ʔaslɒ/ همانگونه که ذکر شد نوع دیگر تنوین دو پیش <ـِـ> است که تنها در تعداد معدودی از کلمات قرضی عربی دیده می‌شود مانند <مضاف‌الیه> و <مشار‌الیه>. دو زبر <ـِـ> و دیگر علائم زیروزبری عربی تنها در متون مذهبی یافت می‌شوند.

^۱.nunation

۲-۱-۳- تشدید^۱: ترکیب علامت زیروزبری تشدید <ـّ> - که در عربی «شدّاء» نامیده می‌شود- با یک نگاره همخوانی آزاد، تشدید خوانده می‌شود. نتیجه واج‌شناختی چنین ترکیبی، یک بار تکرار آن نگاره همخوانی آزاد است و در نتیجه آن همخوان به لحاظ آواشناسی کشیده تلفظ می‌شود. برای اثبات وابسته بودن نگاره تشدید <ـّ> در نظام نوشتاری زبان فارسی می‌توان به جفت‌های کمینه نوشتاری زیر اشاره کرد:

<کمی> /kammi/ در مقابل <کمی> /kami/ ، <سری> /serri/ در مقابل <سری> /seri/ ، <حلال> /hallal/ در مقابل <حلال> /halal/.

۲-۱-۴- خوشه‌نویسی^۲: ترکیب علامت زیروزبری سکون <ْ> با یک نگاره همخوانی آزاد، خوشه‌سازی نامیده می‌شود. نتیجه واج‌شناختی چنین ترکیبی خوشه‌سازی آن نگاره همخوانی است. در خط فارسی قطعه‌های نوشتاری وجود دارند که در آنها حضور یا عدم حضور سکون <ْ> فت کمینه نوشتاری ایجاد می‌کند مانند: <سبک> /sabk/ در مقابل <سبک> /sabok/ ، <سمت> /samt/ در مقابل <سمت> /semat/. بنابراین، نگاره سکون <ْ> یک نگاره وابسته است.

ذکر نمود نوشتاری کلمات دارای سکون نشان می‌دهد که سکون <ْ> بر روی یک نگاره همخوانی- که خوشه‌سازی در مورد آن صورت گرفته است- نقش تعیین مرز جایگاه پایانه را در هجا به عهده دارد؛ بنابراین، سکون <ْ> ابزار نوشتاری دیگری برای حل مسأله ابهام هم‌نگاره‌ها در متن است.

۲-۱-۵- همزه‌سازی^۳: ترکیب علامت زیروزبری همزه <ء> با <ه> غیرملفوظ، همزه-سازی نامیده می‌شود. <ه> غیرملفوظ نگاره‌گونه‌ای^۴ است از <ه> در جایگاه پایانی قطعه نوشتاری که برای نشان دادن واکنه /e/ به کار می‌رود و همزه بالای <ه> نشان دهنده توالی [je] است؛ بنابراین، نشانه <ه> توالی آوایی [eje] است. علامت زیروزبری همزه <ء> یک نگاره وابسته محسوب می‌شود که برای نشان دادن واژه‌بست^۵

1.gemination

2.non-vocalization

3.hamzeticization

4.allograph

5.enclitic

اضافه به کار می‌رود - که در نحو فارسی برای متصل کردن کلمات در ساختار اضافه به کار می‌رود. بنابراین، عدم حضور همزه <ه> می‌تواند ابهام نحوی ایجاد کند و در تبدیل نوشتار به گفتار مشکل ایجاد کند. در مثال‌های: الف) «خاله او را دید» و ب) «خاله او را دید»

در جمله الف <خاله> فاعل و در جمله ب <خاله> نقش مفعول را دارد. بنابراین، عدم حضور همزه <ه> می‌تواند ابهام نحوی ایجاد کند.

شکل دیگر رایج نمود کسره اضافه در متون فارسی، چسباندن شکل مجزای نگاره آزاد <ی> به آخر کلمه برای ساختن ساخت اضافه است مانند <خاله‌ی>. بنابراین، برای نشان دادن ترکیب <ه> غیرملفوظ و کسره اضافه، سه شکل ممکن نوشتاری در متون فارسی یافت می‌شود: <ه^۱>، <ه‌ی> و <ه‌ی>.

شکل دوم <ه‌ی>، توالی است از <ه> و <ی> که برای نگه داشتن شکل <ه> آخر <ه> نویسه نیم‌فاصله (ZWJN (U+200c^۱ بعد از <ه> اضافه می‌شود تا از اتصال <ه> به <ی> جلوگیری شود (اسفهبود^۲ ۲۰۰۴: ۸). شکل سوم <ه‌ی> همان توالی است و تفاوت آن با شکل دوم این است که یک فاصله بین <ه> و <ی> درج شده است. به طور کلی می‌توان گفت این سه شکل نوشتاری برای یک دنباله نوشتاری که متشکل از ستاک و وند است به کار می‌رود و در واقع بر حسب شکل نگاره آخر ستاک و اولین نگاره وند ایجاد می‌شوند. (مگردومیان^۳ ۲۰۰۴). برای مثال برای توالی <کتاب> - که ستاک است- و <ها> - که وند است - سه شکل نوشتاری وجود دارد: <کتابها>، <کتاب‌ها> و <کتاب‌ها>.

۲-۱-۶-حروف مرکب^۳: در فارسی، حروف مرکب متشکل از دو نگاره است که به صورت یک واحد نوشته می‌شوند. ترکیب‌های متعددی از ترکیب نگاره‌های آزاد با نگاره‌های وابسته در فارسی وجود دارند اما در صورتی حرف مرکب خوانده می‌شوند که فرهنگ لغت فارسی آنها را همانند یک حرف منفرد در الفبا بیاورد. بر اساس چنین

^۱.Zero-width non-joiner

نشانه کنترل کننده‌ایست که در محل اتصال تکواژ به ستاک به کار برده می‌شود تا تکواژ و ستاک به هم وصل شوند و یک کلمه نوشتاری بسازند

^۲.Esfahbod

^۳ ligature

۲-۲-۳- ارتباط نگاره و واج: یک به چند

<ا> → {/v/, /ʔ/, /Φ/} ؛ <و> → {/v/, /u/, /o/} ؛

<ی> → {/j/, /i/, /v/}

رابطه میان <ا> و /ʔ/ بحث برانگیز است. هر واکه ابتدایی در خط فارسی روی کرسی <ا> قرار می‌گیرد. شکل‌های نوشتاری واکه‌های بلند ابتدایی یعنی: (/i/، /u/، /v/) به ترتیب به صورت <ای>، <او> و <آ> هستند. اما از آنجایی که هیچ حرف مازادی در نظام نوشتاری فارسی برای نشان دادن واکه‌های کوتاه به کار برده نمی‌شود، <ا> ابتدایی به صورت [ʔa]، [ʔe] و یا [ʔo] تلفظ می‌شود. بعضی از نویسندگان برای ابهام‌زدایی ارزش واجی <ا> ابتدایی یا هر حرف دیگری که بدنبالش واکه کوتاه بیاید، از نگاره‌های وابسته فتحه <َ>، کسره <ِ> و ضمه <ُ> در بالا یا پایین الف <ا> استفاده می‌کنند که نتیجه آن نویسه‌های <اَ>، <اِ> و <اُ> است. شایان ذکر است که در بسیاری از کلمات قرضی عربی وجود <ا> ابتدایی به صورت <اِ> یا <اُ> جفت کمینه ساخته است. به عنوان مثال: قطعه‌های حاوی <اَ> در جایگاه ابتدایی‌شان، اسم جمع مکسر هستند و قطعه‌هایی که در ابتدایشان <اِ> است به عمل فعل مربوط می‌شوند مانند <اخبار> /ʔaxbər/ و <اخبار> /ʔexbər/ (عمل به اطلاع‌رسانی)، <اسناد> /ʔasnəd/ و <اسناد> /ʔesnəd/، <اشکال> /aʃkəl/ و <اشکال> /ʔeʃkəl/ از آنجایی که در متون فارسی چنین مواردی بدون نمود فتحه یا کسره ظاهر می‌شوند، تنها در بافت نحوی می‌توان از آنها ابهام‌زدایی کرد.

ساختار دیگر در نظام نوشتاری زبان فارسی توالی حرفی <خو> است. در زبان‌شناسی تاریخی زبان فارسی در این‌که آیا <خو> یک نشانه نوشتاری برای واج موجود در زبان فارسی میانه است /xw/ - یا توالی دو واج /x/ و /w/ اتفاق نظر وجود ندارد، (پیسوویچ^۱، ۱۹۸۵: ۱۲۱-۲). در فارسی معاصر کلماتی وجود دارند که با <خوا> شروع می‌شوند و در آنها حرف <و> تلفظ نمی‌شود. بنابراین، <خوا> - در جایگاه ابتدایی - در این کلمات به صورت [XD] تلفظ می‌شود. به طور کلی می‌توان گفت تلفظ <خوا> در جایگاه ابتدایی به دو صورت است: ۱- [XD] : که رایج‌ترین کلماتی که در آنها <خوا>

^۱.Pisowicz

به صورت [Xd] تلفظ می‌شود، شکل‌های تصریف شده فعل <خواستن> و اسم <خواب> است. ۲- [xavɒ]: رایج‌ترین کلمه در این دسته، کلمه <خواص> است.

۳- عمق خط

عمق خط ناظر بر ارتباط میان زبان و نوشتار است (راجرز، ۲۰۰۵: ۱۷۷). با گذشت زمان تغییراتی در زبان رخ می‌دهد اما این امکان وجود دارد که نظام نوشتاری تغییر نکند؛ این مسأله باعث افزایش عمق خط می‌شود. در نظام نوشتاری سطحی، نگاره‌ها نماینده واج‌ها هستند مانند خط فنلاندی، در حالیکه در نظام نوشتاری عمیق، نگاره‌ها نماینده واژواچ‌ها هستند^۱؛ در چنین نظامی نشانه‌های کمتری برای نمایش واج‌ها به کار برده می‌شود. از دیگر عواملی که باعث افزایش عمق خط می‌شود، وجود تکواژگونه‌هایی از یک تکواژ است که نگارش یکسان اما تلفظ متفاوتی دارند؛ به عنوان مثال در زبان انگلیسی تکواژگونه‌های /tʃaɪld/ و /tʃɪld/ به ترتیب در دو کلمه child و children، تکواژگونه‌های /saʊθ/ و /sʌð/ در دو کلمه south و southern باعث افزایش عمق خط انگلیسی می‌شوند زیرا هر جفت تکواژ صورت نوشتاری یکسان ولی تلفظ‌های متفاوت دارند.

تاکنون روش‌هایی برای محاسبه عمق خط در زبانهای گوناگون پیشنهاد شده است که از آن جمله می‌توان به روشی که اسپروت^۲ (۲۰۰۰) برای محاسبه عمق خط انگلیسی ارائه داده است، اشاره کرد. وی تعداد ۱۱۶۹ کلمه را از یک فرهنگ لغت انتخاب کرده که املاء انگلیسی آمریکایی و نمود واجی مربوط به تلفظ انگلیسی آمریکایی معیار آن کلمات نیز آورده شده است. علاوه بر ذکر نمود واجی موجود در فرهنگ لغت که در واقع نمود روساختی محسوب می‌شود، اسپروت بر اساس مدل اس-پی‌ای^۳ چامسکی^۴ و هله^۵ (۱۹۶۸) نمود زیرساختی مربوط به آن کلمات را نیز بازسازی کرده است. وی تعدادی قاعده هم تحت عنوان قاعده عمیق و قاعده سطحی به کار برده است که از نظر تعداد با هم متفاوتند. بنابراین، در مورد هر کلمه در فهرست مذکور با

^۱.morphophoneme

^۲.Sproat

^۳.SPE

^۴.Chomsky

^۵.Halle

توجه به قواعد عمیق و قواعد سطحی دو نمود واجی سطحی و عمیق ارائه شده است و با توجه به تفاوت چشم‌گیر نموده‌های عمیق و سطحی، وی چنین نتیجه‌گیری می‌کند که نظام نوشتاری زبان انگلیسی عمیق در نظر گرفته می‌شود. اسپروت نظام نوشتاری زبان روسی را نیز عمیق در نظر می‌گیرد. وی چنین استدلال می‌کند که در خط روسی به منظور ارائه تلفظ صحیح، خواننده باید از اطلاعات واژگانی و بافتی که کلمات در آن به کار برده شده‌اند، استفاده کند.

۴- عمق خط فارسی

خط فارسی از نقطه نظر اطلاعات واجی موجود در آن تا حدی مطالعه شده است، اما روشی برای محاسبه عمق خط فارسی ارائه نشده است. در این بخش نگارنده روشی را برای تعیین عمق خط فارسی ارائه می‌دهد. برای تعیین عمق خط فارسی، از تعبیر شهودی «دوری از یک‌به‌یک بودن» استفاده می‌شود و به‌ازای هر قطعه نوشتاری مشخصی که داده شده، تعداد حالت‌های ممکن خواننده شدن آن قطعه و سهم آن در «عمیق کردن» پیکره محاسبه می‌شود. برای صورت‌بندی مسئله، فرض کنید پیکره مرجع C داده شده باشد. مجموعه همه قطعه‌های نوشتاری موجود در C با S نشان داده می‌شود. به این ترتیب، مثلاً «مهر» عضوی از S است. در حالی که صورت‌های مختلفی از آن (مانند مهر، مَهر و مَهر) در C یافت می‌شوند.

• فراوانی قطعه‌ای مانند s از عضوهای S با $v(s)$ نشان داده می‌شود، و به صورت حاصل جمع همه فراوانی‌های همه صورت‌های تلفظی s در C تعریف می‌شود.

• به‌ازای هر قطعه مانند s در S ، تعداد حالت‌های قابل قبول خواندن s با $\alpha(s)$ نشان داده می‌شود و به صورت تعداد تلفظ‌های مختلف قطعه نوشتاری s در پیکره تعریف می‌شود (به عنوان مثال، $\alpha(\text{مهر}) = 3$).

• به‌ازای هر قطعه مانند s در S ، تعداد حالت‌های «ممکن» تلفظ s با $\pi(s)$ نشان داده می‌شود و به صورت تعداد تلفظ‌های ممکن قطعه s تعریف می‌شود (به عنوان مثال، $\pi = 12$).

(مهر) Π که شامل: /meh/, /mehor/, /mehar/, /mahr/, /mohr/, /meh/, /mahor/, /mahar/, /maher/, /mahor/, /mohr/, /mohar/, /moher/, /meher/, /mahor/ می‌باشد.

به این ترتیب، نسبت $\frac{\pi(s)}{\alpha(s)}$ میزان «ابهام» در تلفظ قطعه S [= دوری از یک‌به‌یک بودن رابطه قطعه و تلفظ آن] را نشان می‌دهد (به عنوان مثال، این نسبت برای <مهر < برابر ۴ است، در حالی که برای <راز < برابر ۱ است). به همین ترتیب، عبارت $\sum_{t \in S} \frac{v(s)}{v(t)} \times \frac{\pi(s)}{\alpha(s)}$ سهم قطعه S را در کل ابهام نشان می‌دهد و در نتیجه، اندازه ابهام کل پیکره برابر است با
$$\frac{\sum_{s \in S} \left(v(s) \times \frac{\pi(s)}{\alpha(s)} \right)}{\sum_{t \in S} v(t)}$$

در فرمول فوق، مخرج کسر مجموع همه حالت‌های خوانده شده قطعه‌های نوشتاری موجود در پیکره می‌باشد. به عنوان مثال، ممکن است در پیکره قطعه نوشتاری <مهر < ۱۰۰۰ بار با تلفظ /mohr/ ظاهر شده باشد، ۲۰۰۰ بار با تلفظ /mehr/. بنابراین برای همه قطعه‌های نوشتاری موجود در پیکره، حالت‌های تلفظی ظاهر شده مربوط به هر کدام در نظر گرفته می‌شود. در پیکره برای محاسبه تعداد حالت‌های «ممکن» تلفظ قطعه نوشتاری داده‌شده، می‌توان از این نکته استفاده کرد که از آنجائی که به لحاظ واج‌شناختی در هر هجا / بخش، یک واکه وجود دارد، بنابراین اگر بدانیم که چطور می‌توان یک قطعه را «بخش» کرد، تعداد حالت‌های ممکن تلفظ کل قطعه برابر است با حاصلضرب تعداد حالت‌های ممکن تلفظ هر کدام از «بخش»‌ها. اما، چون روش مشخصی برای تعیین این بخش‌ها وجود ندارد، مسئله تحویل می‌شود به تعیین مواضع احتمالی بخش کردن. اشکال کار این است که صرف شکل پشت‌هم قرار گرفتن همخوان‌ها معیار مشخصی برای بخش کردن قطعه به دست نمی‌دهد — حتی قطعه‌هایی مانند <خرد> هم هستند که در زبان، هم به عنوان کلمه‌ای یک‌بخشی ظاهر می‌شوند و هم به عنوان کلمه‌ای دوبخشی. در واقع، تنها قاعده «قطعی» این است که همخوانی که بلافاصله پیش از واکه‌ای بلند قرار دارد، حتماً شروع بخشی جدید است (مثلاً، <الهام>، لزوماً به صورت <ال + بهام> می‌شکند، هر چند که خود <ل> و <ل> ابتدای قطعه ممکن است به <ل + ل> بشکند یا مجموعاً در یک بخش قرار گیرد <ال>). پس برای پیدا کردن مواضع «قطعی» بخش کردن، باید واکه‌های بلند را پیدا کنیم. اشکال کار، این است که نگاره‌های <ی> و <و> می‌توانند نشانه همخوان /j/ برای نگاره <ی> و همخوان /v/ برای نگاره <و>،

واکه کوتاه /o/ برای نگاره <و> یا واکه بلند /i/ برای نگاره <ی> و واکه بلند /u/ برای نگاره <و> باشند و باید این موضوع را هم در شمارش لحاظ کرد.

در نظام نوشتاری زبان فارسی تنها واکه‌های بلند و همخوان‌ها نمایش داده می‌شوند. از آنجا که نوع واکه بلند (<و>، <ی>، <د>) تأثیری در محاسبه تعداد هجاها ندارد، در محاسبه هر کدام از همخوان‌ها را با C و هر کدام از واکه‌های بلند با a نمایش داده می‌شوند. به این ترتیب، قطعه نوشتاری <الهام>/?elhom/ به cccac تبدیل می‌شود (به درج «بست چاکنایی» در ابتدای قطعه توجه کنید). از آنجا که همخوان قبل از واکه بلند در محاسبه تعداد حالت‌های ممکن تلفظ قطعه تأثیری ندارد و با واکه بعدی در یک هجا قرار می‌گیرد، C قبل از a و خود a را با هم به صورت یک A نمایش داده می‌شوند. همچنین از آنجایی که همخوان پایان قطعه نوشتاری نیز تغییری در محاسبه تعداد حالت‌های «ممکن» تلفظ آن قطعه نوشتاری ندارد، همخوان پایانی با C نمایش داده می‌شود تا از همخوان‌های دیگر متمایز شود. به این ترتیب، به‌ازای قطعه <الهام> (که به cccac تبدیل می‌شود) به ccAC می‌رسیم و قطعه نوشتاری <عقل> هم به cCC تبدیل می‌شود.

برای محاسبه تعداد حالت‌های ممکن تلفظ قطعه‌ای که واکه بلند ندارد، باید به این مسئله توجه کرد که همخوان پایانی در تعیین تعداد حالت‌ها نقشی ایفا نمی‌کند، بنابراین، قطعه‌ای مانند «عقل» در حقیقت دارای دو همخوان تعیین‌کننده تعداد حالت‌های ممکن تلفظ این قطعه نوشتاری است. اگر S_n را تعداد تلفظ‌های ممکن چنین قطعه‌ای در نظر بگیریم که در آن n تعداد Cهای این قطعه است، به‌ازای قطعه‌ای متشکل از یک C، مانند <در>، با در نظر گرفتن احتمال حضور سه واکه کوتاه بعد از همخوان اول در تلفظ نتیجه می‌شود که

$S_1=3$. بنابراین، سه حالت تلفظی ممکن برای این قطعه نوشتاری وجود دارد. به‌ازای قطعه‌ای متشکل از دو C، مانند <خرس>، با در نظر گرفتن احتمال حضور سه واکه کوتاه بعد از همخوان اول و احتمال حضور سه واکه کوتاه و ساکن بعد از همخوان دوم، ۱۲ حالت تلفظی ممکن وجود خواهد داشت؛ پس

$S_2=3 \times 4=12$ و برای قطعه‌ای متشکل از سه C، مانند <کشتم>، اگر در تقطیع هجایی C اول را یک هجا در نظر بگیریم (که به‌طور بالقوه می‌تواند با سه واکه‌ی کوتاه همراه شود) و بقیه قطعه را به‌طور مجزا به عنوان قطعه‌ای دیگر (که به S_2 حالت

خوانده می‌شود) در نظر بگیریم، ۳۶ حالت به دست می‌آید و اگر دو C اول را یک هجا و بقیه قطعه را هم یک قطعه در نظر بگیریم، ۹ حالت دیگر به حالت‌های قبل اضافه می‌شود. به این ترتیب، در کل ۴۵ حالت تلفظی ممکن برای چنین قطعه‌هایی می‌توان تصور کرد که یعنی $S_3 = 36 + 9 = 45$ در حالت کلی، به دلیل الگوی هجایی زبان فارسی، نمی‌توان هجایی داشت که از بیش از سه C متوالی تشکیل شده باشد؛ پس بسته به این که چه تعدادی از C‌های ابتدایی در هجای اول قرار می‌گیرند می‌توان نوشت:

$$S_n = 3S_{n-1} + 3S_{n-2} + 3S_{n-3}$$

حال قطعه‌ای را تصور کنید که در آن به غیر از تعدادی C، A هم وجود دارد. در این صورت یا A خودش یک هجا محسوب می‌شود و بقیه قطعه n تا C است و یا A روی یک یا دو C بعد از خود اثر می‌گذارد و آن‌ها را ساکن می‌کند؛ بنابراین، در مورد قطعه‌ای مانند «بناپارت» مراحل تقطیع هجایی به این صورت خواهد بود:

ccacacc بناپارت

caAcC با حذف C‌های قبل از a و بی‌تأثیر بودن C پایانی در شمردن حالت‌ها بنابراین، با در نظر گرفتن این که سه واژه کوتاه می‌تواند با C اول همراه شود ($S_1 = 3$) و جدا کردن A به عنوان یک هجا (که فقط به یک حالت خوانده می‌شود) و AcC به عنوان مرز شروع هجا یا هجاهایی دیگر، نهایتاً به $12 = 3 \times (3+1)$ حالت ممکن تلفظی می‌رسیم.

پیکره مورد استفاده در این تحقیق متنی بود تحت عنوان «مرغ آتش‌خوار» از کتاب فارسی پنجم دبستان که فراوانی، تعداد تلفظ‌های ممکن و تعداد تلفظ‌های قابل قبول قطعه‌های نوشتاری موجود در این پیکره محاسبه شد. به عنوان مثال برای جمله «یک روز یکی از بزرگان از سفری برگشته بود» موارد فوق بدین صورت محاسبه شد:

تعداد حالت‌های قابل قبول تلفظی	تعداد حالت‌های ممکن تلفظی	فراوانی	قطعه نوشتاری
۱	۳	۱	یک
۱	۱۳	۴	روز
۱	۳	۵	یکی
۱	۳	۱۹	از
۱	۴۵	۱	بزرگان
۲	۵۷	۱	سفری
۱	۷۱۱	۱	برگشته
۲	۱۳	۱۱	بود

و در فرمول ارائه شده (یعنی $\frac{\pi(s)}{\alpha(s)} \times \frac{v(s)}{\sum_{t \in S} v(t)}$) که قبلاً دیدیم) برای اندازه‌گیری

سهام هر کدام از قطعه‌ها در ابهام متن قرار داده شدند و در نهایت، اندازه ابهام کل پیکره

با توجه به فرمول $\frac{\sum_{s \in S} (v(s) \times \frac{\pi(s)}{\alpha(s)})}{\sum_{t \in S} v(t)}$ به دست آمد. نتیجه دور از انتظار، عدد ۸۷/۷

است که نشان می‌دهد ارتباط نظام نوشتاری زبان فارسی با زبان گفتاری، به‌میزان قابل توجهی از «تناظر یک به یک» دور است و به‌این ترتیب، نظام نوشتاری زبان فارسی را باید عمیق در نظر گرفت.

محدودیت‌های حاکم بر این محاسبه شامل لحاظ نکردن تشدید و کسره اضافه می‌باشد، سؤال مطرح در این مبحث این است که آیا در نظر گرفتن این موارد تأثیری در نتیجه‌گیری نهایی مبنی بر در نظر گرفتن نظام نوشتاری زبان فارسی به عنوان نظامی عمیق دارد؟ در پاسخ به این سؤال باید گفت عدد ۸۷/۷ آنقدر از عدد یک دور است که لحاظ کردن مواردی از قبیل کسره اضافه و تشدید باعث بالاتر رفتن این عدد و در نهایت افزایش بیشتر عمق نظام نوشتاری زبان فارسی می‌شود. زیرا تایپیست‌ها از تایپ نگاره‌های وابسته معمولاً خودداری می‌کنند. با در نظر گرفتن چنین عمق زیادی برای خط فارسی، این سؤال مطرح می‌شود که چگونه گویشوران زبان فارسی قادر به خواندن خط فارسی هستند؟ به نظر می‌رسد در فرآیند خواندن، گویشوران زبان فارسی از اطلاعات زبانی خود استفاده می‌کنند و بدینوسیله عمق خط فارسی را کم می‌کنند. این اطلاعات زبانی شامل اطلاعات تکواژی، قواعد حرف‌نویسی مربوط به اضافه شدن وندها به ستاک که در مرز تکواژها عمل می‌کنند، آشنایی با صورت نوشتاری و تلفظ کلماتی که منشأ عربی دارند استفاده از قیاس با تلفظ‌ها و صورت‌های نوشتاری در واژگان ذهنیشان است. بنابراین می‌توان نتیجه گرفت که عدد ۸۷/۷ که بدون لحاظ کردن چنین اطلاعات زبانی بدست آمده - یک حد بیشینه برای عمق خط فارسی است و گویشوران زبان فارسی با استفاده از اطلاعات زبانی عمق خط فارسی را کم می‌کنند.

۵- نتیجه

در این بخش سعی بر آن است که به سؤال مطرح‌شده در تحقیق پاسخ داده شود. همانگونه که ذکر شد در یک نظام نوشتاری آرمانی، رابطه یک‌به‌یک میان نشانه‌های نوشتاری و واج‌ها وجود دارد. برای بررسی میزان نزدیکی نظام نوشتاری زبان فارسی به نظام نوشتاری آرمانی ابتدا نظام نوشتاری زبان فارسی به تفصیل بررسی شد. با توجه به عدم حضور واژه‌های کوتاه در نظام نوشتاری زبان فارسی، وجود علائم زیروزبری که نگاره‌های وابسته محسوب می‌شوند، رابطه نگاره و واج (چندبه‌یک و یک‌به‌چند) و وجود ساختار <خو> که به صورت [xv] و یا [xavv] تلفظ می‌شود و محاسبه به عمل آمده برای تعیین عمق خط فارسی (عدد ۸۷/۷ که بسیار دور از عدد یک است)، می‌توان گفت نظام نوشتاری زبان فارسی، نظامی عمیق است و بدین ترتیب از نظام نوشتاری آرمانی دور است. عدد بدست‌آمده از محاسبه عمق خط فارسی بدون در نظر گرفتن اطلاعات زبانی است که گویشوران زبان فارسی در فرآیند خواندن استفاده می‌کنند و همانگونه که ذکر شد، عدد حاصل یک حد بیشینه است و گویشوران زبان فارسی با استفاده از اطلاعات زبانی مانند اطلاعات تکواژی، آشنایی با صورت نوشتاری وندهای تصریفی و اشتقاقی، دانستن قواعد حرف‌نویسی مربوط به اضافه شدن وندها به ستاک که در مرز تکواژها عمل می‌کنند، آشنایی با صورت نوشتاری و تلفظ کلماتی که منشأ عربی دارند و استفاده از قیاس با کلمات موجود در واژگان ذهنی خود، سعی در کم کردن عمق خط فارسی دارند تا بدین ترتیب بتوانند کلمات نوشتاری را که اولین بار با آنها مواجه می‌شوند، بدرستی بخوانند. همچنین سیستم‌های نوشتار به گفتار از این رهگذر قادر خواهند بود کلمات نوشتاری را که در واژگان خود ندارند، بدرستی به رشته واج‌ها تبدیل کنند. نویسندگان در تحقیقات آتی خود با استفاده از اطلاعات واجی، ساخت‌واژی و نحوی در نظام نوشتاری فارسی به مطالعه کاهش عمق خط فارسی خواهند پرداخت تا از این رهگذر بتوان بازشناسی صورت واجی نوواژه‌ها را تبیین کرد.

منابع

دانای طوس، مریم. ۱۳۸۴. اثرات شفافیت و تیرگی خط فارسی بر مهارت زبان خواندن کودکان طبیعی و خوانش‌پریش رشدی فارسی زبان. رساله دکتری، دانشگاه تهران.

Buckwalter, T. 2004. Issues in Arabic orthography and morphology analysis.

- In *computational approaches to Arabic script-based languages in conjunction with COLING 2004*. Switzerland.
- El_Imam, Y. 2003. Phonetization of Arabic: rules and algorithms. *Computer speech and language*. 18(2004):339-373. Elsevier Ltd.
- Esfahbod, B. 2004. Persian computing with unicode. In *25th internationalization and unicode conference*, Washington, DC.
- Gibson, E.J., A. Pick, H. Osler, and M. Hommond, 1962. The role of grapheme-phoneme correspondence in the perception of words. *The American journal of psychology*, Vol.75, No.4: 554-570.
- Megerdooian, K. 2004. Finite state morphological analysis of Persian. In *computational approaches to Arabic script-based languages*. Switzerland.
- Pisowicz, A. 1985. Origins of the new and middle Persian phonological system. *Nakladem Uniwersytetu Jagiellonskiego*.
- Rogers, H. 2005. *Writing systems: A linguistics approach*. MA: Blackwell textbooks in linguistics. Blackwell publishing.
- Sproat, R. 2000. *A computational theory of writing systems*. Cambridge University Press. Stanford, CA.
- Trammel, R.L. 1990. Variant grapheme-phoneme correspondences in unfamiliar polysyllabic words. *Language and speech*. 33(4): 293-323.
- Van den Bosch, A., A. Content, W. Daelemans, and B. De Gelder, 1994. Analysing orthographic depth of different languages using data-oriented algorithms. In *2nd international conference on quantitative linguistics*. Moscow.
- Venezky, R.L. 2004. In Search of the perfect orthography. *Written language and literacy* :139-163. Amsterdam. John Benjamins publishing company.

