

فصلنامه تخصصی سبک‌شناسی نظم و نثر فارسی (بهار ادب)

علمی-پژوهشی

سال هشتم-شماره دوم-تابستان ۱۳۹۴-شماره پیاپی ۲۸

چالش‌های سبک نوشتار در نویسه‌گردانی ماشینی خطوط فارسی و سرلیک

تاجیکی

(ص ۵۱۴-۴۹۳)

محسن رحیمی^۱

تاریخ دریافت مقاله: ۱۳۹۴/۰۳/۰۶

تاریخ پذیرش قطعی مقاله: ۱۳۹۴/۰۵/۲۷

چکیده

نویسه‌گردانی یا (Transliteration) معمولاً برای نمایش صحیح آوایی کلمات بیگانه در نوشتار به کار گرفته می‌شود. در صورتیکه یک زبان به دو خط مختلف نوشته شود، خوانش محتوای نوشته شده از طریق نویسه‌گردانی قابل انجام خواهد بود. خطوط فارسی و سرلیک از جمله خطوطی هستند که دارای یک زبان مشترک و دو شیوه نمایش (خط) متفاوت هستند. این مقاله در آغاز به تاریخچه تغییر خط در تاجیکستان از فارسی به لاتین و سرلیک پرداخته و به مقایسه‌نویسه‌های فارسی و سرلیک اشاراتی داشته است. سپس با بررسی یک ابزار نویسه‌گردانی رایانه‌ای جهت تبدیل سریع و هوشمند محتوای فارسی در دو سوی خطوط فارسی و سرلیک، نقاط قوت و ضعف این روش را برشمرده و چالش‌های سبک نوشتار فارسی تاجیکی و ایرانی را در این فرایند تبدیلی مقایسه نموده است. همچنین ضمن اشاره به ویژگی‌های هر دو خط در این مسیر، راهکارهایی برای توسعه و تکمیل چنین ابزارهایی ارائه داده است. نتایج این پژوهش نشان می‌دهد برخی دشواریهای منفصل و متصل‌نویسی فارسی بدلیل استفاده از خطوط سامی چگونه در نسخه سرلیک خط فارسی مرتفع شده و همچنین تاثیر شیوه نگارش روسی تا چه حد بر سبک نوشتار فارسی تاجیکی بر جای مانده است.

کلمات کلیدی: سبک نوشتار، نویسه‌گردانی ماشینی، فارسی تاجیکی، سرلیک

^۱مربی گروه زبان و ادبیات فارسی دانشگاه پیام نور M_rahimi@pnu.ac.ir

۱. مقدمه

در سامانه‌های ترجمه ماشینی که معمولاً برای دو زبان مختلف با دو خط متفاوت طراحی میشوند، تغییرات عناصر واژگانی، نحوی و معنایی بصورت همزمان در نظر گرفته میشوند، اما بهر میزان از تفاوت‌های بین دو زبان کاسته شود نتیجه بدست آمده از کیفیت بیشتری برخوردار خواهد بود. بعنوان مثال در بین دو زبان هم خانواده که از خط و ریشه زبانی یکسانی برخوردارند، به نسبت دو زبان با دو خط متفاوت، دگرگونی‌های نحوی و واژگانی، کمتر بچشم می‌خورد.

در میان زبانهایی که به دو خط متفاوت نوشته میشوند، دشواری انتقال، بسیار کمتر است؛ زیرا با نویسه‌گردانی حروف و کلمات دو زبان، امکان خوانش متن برای کاربران هر دو زبان فراهم شده و تنها واژگان خارجی آن متن نیازمند ترجمه هستند. نتیجه در این حالت، با در نظر گرفتن اندکی اصلاح در رسم‌الخط و برخی نکات نحوی کاملاً قابل استفاده خواهد بود.

در میان زبانهای زنده جهان، بسیاری از اقوام هم‌زبان بدلیل پاره‌ای ملاحظات سیاسی و یا اجتماعی دارای شیوه‌های نگارش متفاوتی هستند. از جمله میتوان به اقوام ترک‌زبان و فارسی‌زبان اشاره کرد. پراکندگی این اقوام در سطح جغرافیایی کشورهای مختلف باعث بروز اختلاف در شیوه نگارش آنها بر اساس خط معیار شده است. خط فارسی و خط سرلیک فارسی تاجیکی یکی از نمونه‌های قابل توجه در زبان فارسی هستند.

از قرن ۱۶ میلادی که در ایران حکومت شیعی مذهب صفویه و در ماوراءالنهر حکومت سنی مذهب شیبانی (از اقوام ازبک) تشکیل شد، و بدلائل مذهبی و ملی رابطه سیاسی ایران با جهان خارج محدود گردید، رابطه فرهنگی بین ایرانیان و تاجیکان نیز تقلیل یافت. از اینزمان زبان و ادب فارسی در ایران و ماوراءالنهر به دو مسیر جداگانه افتاد و انقلاب اکتبر ۱۹۱۷ باعث گردید که ایرانیان و تاجیکان بکلی از یکدیگر بی اطلاع بمانند. (کلباسی، ۱۳۷۴: ۳۶)

وجود منابع متنوع فرهنگی، ادبی، علمی و آموزشی در زبان فارسی تاجیکی با خط سرلیک که بیشتر برگرفته از منابع زبان روسی و ادبیات غنی فارسی‌زبانان تاجیکستان میباشد و حجم بی بدیل منابع زبان فارسی به خط فارسی که برآمده از فرهنگ چندهزارساله ایرانیان است بدلیل تفاوت در خط، دور از دسترس علاقمندان هر دو سوی این موضوع قرار گرفته است. فارسی دری، امروز در افغانستان، تاجیکستان و ایران رایج است. در هر سه کشور از اوایل قرن بیستم میلادی، وضعی برای فارسی‌زبانان پیش آمد که باعث شد فارسی رایج در هر یک از این کشورها به راهی بیفتد که به تدریج آنها را از هم جدا خواهد کرد. (ابوالقاسمی، ۱۳۸۵: ۲۶۱)

هدف این مقاله بررسی استفاده از ابزارهای نوین رایانشی جهت انتقال سریع داده‌ها بین دو خط از زبان فارسی است. این ابزارها در راستای گسترش خود بدلیل اختلاف در شیوه نگارش دچار

چالشهای متفاوتی میشوند. در این مقاله، موضوع نویسه‌گردانی ماشینی در دو خط فارسی و تفاوت‌های سبک نوشتار در این دو خط، مورد مطالعه و بررسی قرار گرفته است.

۲. داده‌های پژوهش

داده‌های این پژوهش شامل عبارات ساده و رایج گویشوران دو کشور (جمع‌آوری شده توسط نویسندگان)، به‌همراه مطالبی از آرشیو خبری نسخه آنلاین روزنامه ملت تاجیکستان تاریخ ۲۰۱۲-۰۹-۰۱ و پایگاه اطلاع‌رسانی حزب نهضت اسلامی تاجیکستان به آدرس (<http://www.nahzat.tj>) میباشد. نگارنده با داشتن تسلط بر خطوط فارسی و تاجیکی منابع متفاوت از هر دو خط را انتخاب و مورد تحلیل و بررسی قرار داده است. برای ارزیابی نویسه‌گردانی خودکار بین دو خط فارسی و سرلیک از برنامه رایانه‌ای تحت وب در نشانی اینترنتی (<http://www.persian-tajik.ir>) استفاده شده است. برای تبدیل آوانوشتی سرلیک به لاتین از نویسه‌گردان مورد تایید [lexilogos.com](http://www.lexilogos.com) استفاده شده است. سایر مثالهای کاربردی از محاورات روزمره اخذ شده است.

۳. پیشینه موضوع و اهمیت پژوهش

تاکنون در کنار تمام پیشرفتهای علوم رایانه، تلاش ناچیزی برای تبدیل و یا نویسه‌گردانی خودکار بین دو خط فارسی و تاجیکی صورت گرفته است. یکی از این تلاشهای قابل توجه، توسط جلال حجتی فهیم با همراهی دو زبانشناس در قالب سامانه نویسه‌گردانی خط فارسی و سرلیک انجام شده است. مطالعات این موضوع از سال ۱۳۸۶ شمسی آغاز شده و تا پایان سال ۱۳۹۱ دو مرحله اصلی فارسی به سرلیک و سرلیک به فارسی محقق شده است. اهمیت موضوع دستیابی سریع و خودکار به دانش نهفته در متون هر خط در دو سوی جغرافیای زبان فارسی بر کسی پوشیده نیست. بسط و گسترش دانش و فرهنگ ایرانی و اسلامی و آشنایی با ادبیات و منابع غنی ادبیات تاجیکی که طی ۷۰ سال از طریق زبان روسی به سرلیک ترجمه شده، بزرگترین ویژگی این فرایند است که از طریق فناوری رایانشی و نویسه‌گردانی خودکار در حداقل زمان، محقق خواهد شد. اما تاکنون کسی به موضوع چالشهای سبک نوشتار فارسی تاجیکی در این فرایند نپرداخته است.

۴. بحث و بررسی

زبانهای ایرانی که در حقیقت انشعاب ایرانی از شاخه هندوایرانی و با ریشه هندواروپایی است شامل فارسی، تاجیکی، پشتو، آسی، کردی، طالشی، تاتی و بلوچی، زبانهای پامیری (شغنی و روشانی) و برخی دیگر زبانهای باستانی است که فقط از طریق اسناد مکتوب از آنها اطلاع داریم. دسته حاضر همگی دارای خط یکسان نبوده و برخی فقط از طریق گویشوران آن باقی مانده‌اند. بر اساس نظر ای. ام. آرانسکی در کتاب زبانهای ایرانی، اشتراک واژه‌های بنیادی و اصول ساخت دستوری این زبانها، برهان خویشاوندی تباری آنهاست. (آرانسکی، ۱۳۷۸ : ۲۶) براین اساس خط فارسی و سرلیک دو نمود متفاوت از یک زبان واحد هستند. تفاوت در این خطوط بر اساس نظام نوشتاری آنها قابل

بررسی است. به این مفهوم که هر کدام از خطوط در درون ساختار خود با هماهنگی نظام نوشتاری خود نمایش داده میشوند. بنابراین در نگارش این هر دو خط، نظام نوشتاری آنها بیشتر باید مورد توجه قرار گیرد. مطالعه *در زمانی*^۱ خطوط سامی نشان میدهد که شاخه شمالی خطوط سامی به خطوط کنعانی، آرامی و فینیقی انشعاب پیدا کرده است. از خط آرامی نبطی و از آن خط سینایی نو پدید آمده که باعث پیدایش خط عربی در قرن اول پیش از میلاد مسیح شد. خط فارسی کنونی، یکی از خطوط برگرفته از خط عربی است. از سوی دیگر از خط فینیقی، بهترین خط ممکن برای نظام آوایی و معادل سازی واجها را یونانیان پایه گذاری کردند که از همین خط «اتروسکی» و از آن خط رومیایی بعنوان پایه گذار خطوط لاتینی شکل گرفت. خط سرلیک در انشعابی مستقیم از خط یونانی اخذ شده است.

در کتاب درآمدی بر چگونگی شیوه خط فارسی و در توصیف فرآیند ساخت الفبای سامی آمده است: ایشان حرفهای صامت را جسم متصور داشته‌اند که دیدنی است و آنها را نشان دادند و مصوتها را روح که دیدنی نیست و آنها را نشان ندادند. برای نخستین بار یونانیان در حدود نهمصد سال پیش از میلاد، روح را نیز به جسم مبدل ساختند و وارد حرفهای الفبا گرداندند. (ادیب سلطانی، ۱۳۷۸: ۲۲) برخی زبانشناسان معتقدند، خط فارسی نمیتواند نمایش دهنده مناسبی برای گفتار فارسی باشد. این گروه، به ریشه‌های تاریخی زبان و خط فارسی اشاره کرده و معتقدند: زبان عربی از خانواده زبانهای سامی و زبان فارسی از خانواده زبانهای هندواروپایی است. بنابراین دو زبان تفاوتی ریشه‌ای در ساخت واجی، صرفی، و نحوی دارند و بنا به همین دلایل خط عربی نتوانسته و نمیتواند بازتاب دقیقی از گفتار فارسی باشد. (خلخالی نازیلا، ۱۳۷۵: ۲۲)

مطابق نظر دکتر کوروش صفوی در کتاب آشنایی با نظامهای نوشتاری، نویسه معادل (grapheme) بوده و هر نشانه‌ای که در خط بکار میرود به یک نویسه دلالت میکند. این نویسه‌ها را میتوان در دوگروه نویسه‌های اصلی و فرعی تقسیم کرد. (صفوی، ۱۳۸۶: ۲۹) او معتقد است نویسه‌های اصلی همان حروف الفبا در شکل سنتی هستند ولی نویسه‌های فرعی شامل علائم نقطه-گذاری، اعداد، و نمادهای قراردادی همچون \$ یا @ میشوند. ضمناً نویسه‌گونه‌ها^۲ در اصل صورتهای مختلف یک نویسه‌اند که معمولاً دارای تفاوتی در ظاهر هستند.

نویسه‌گردانی یا (Transliteration) معادل نوشتن یا هجی کردن واژه‌ها و عبارات یک زبان به زبان دیگر با الفبای زبان مقصد است بنحوی که لحنی مشابه از آن اراده گردد. به تعبیر دیگر

^۱ diachronic

^۲ Allograph

نویسه‌گردانی عبارت است از معادل‌یابی و معادل‌سازی واژه‌ها و اصطلاحات، بنحوی که بتوان بار آوایی آن را از زبان مبدأ، به زبان مقصد به قدر ممکن، منتقل کرد. این عمل با صرف وقت کافی و بررسی تمامی ملاحظات دو زبان مبدأ و مقصد صورت می‌پذیرد. در برخی موارد نویسه‌گردانی واژگان، جایگزینی برای لغات غیر قابل ترجمه محسوب می‌شود. هر واج در هر زبان دارای یک بار آوایی است که صرف‌نظر از ویژگی‌های خاص آن زبان در سایر زبانها دارای معادلی مشخص و یا ترکیبی است. در یک متن نوشته شده به خط فارسی هنگام مواجهه با نام فرد یا مکان خاص که از یک زبان دیگر وام گرفته شده و به فارسی برگردانده شده، خواننده فارسی زبان که معمولاً از اعراب استفاده نمی‌کند، نمیتواند از شم زبانی خود کمک بگیرد، بنابراین معادل نویسه‌گردانی شده آن برای خواننده متن، راهگشای خوانش صحیح واژه مورد نظر است. اما در دو گویش فارسی و تاجیکی از زبان فارسی که اختلاف در سطح زبانی نبوده و تنها اشکال مختلفی برای آواهای یکسان وجود دارد، این شیوه میتواند نمایش دهنده یک واژه از خط مبدأ در خط دیگر باشد. بظاهر معادل‌یابی و معادل‌سازی آواها در دو خط و تبدیل خودکار آنها بدون هیچ مشکلی میتواند متن مبدأ را به مقصد منتقل نماید. اما این مقاله نشان میدهد این فرایند در دو سوی خطوط مختلف در زبان فارسی دارای چه ابهامات و نقاط ضعف و قوتی است.

از طریق مطالعه جدول شماره ۱ هر خواننده فارسی زبان ایرانی قادر خواهد بود متن سرلیک از همین زبان را با اندکی ممارست دریابد. نویسه‌های دو خط فارسی و سرلیک به‌مراه ارزش آوایی آنها که از استاندارد (IPA) استفاده میکنند در این جدول نمایش داده شده است.

جدول شماره ۱: نویسه های فارسی - سرلیک - لاتین و آوانویسی

سرلیک	لاتین	فارسی	(IPA)	نمونه
همخوان‌ها				
Б б	B b	ب	/b/	барг = برگ = بَرگ
П п	P p	پ	/p/	панҷ = پَنج
Т т	T t	ت، ط	/t/	точик = طَلَب = талаб تاجیک،
Ч ч	Ç ç	چ	/dʒ/	чанг = چَنگ
Ч ч	C c	چ	/tʃ/	чӣ = چی
Х х	H h	ح، ه	/h/	хофиз = حَافِظ
Х х	X x	خ	/x/	хондан = خواندَن
Д д	D d	د	/d/	модар = بغداد = Бағдод = مادر = مادر،
Р р	R r	ر	/r/	ранг = رَنگ
З з	Z z	ز، ذ، ض	/z/	баъз = захоб = نظر، = назар بعض،
Ж ж	Z z	ژ	/ʒ/	жола = کژدم = کَزْدَم = کاجدوم = زاله،
С с	S s	س، ث	/s/	сар = ثُرَيَا = сурайё = صُبْح، = субх سر،
Ш ш	Sh sh	ش	/ʃ/	шаб = شَب
Ъ	'	ع	/ʔ/	таъриф = تعريف
Ғ ғ	Q q	غ	/ɣ/	ғор = بغداد = بَغْداد = Бағдод = غار،
Ф ф	F f	ف	/f/	фурӯғ = فُرُوع
Қ қ	Q q	ق	/q/	қадам = قَدَم

سرلیک	لات	فارسی	(IPA)	نمونه
К к	K k	ک	/k/	кадом = کدام
Г г	G g	گ	/g/	санг = سنگ = سنگ
Л л	L l	ل	/l/	лола = لاله
М м	M m	م	/m/	мурдагӣ = مُردگی
Н н	N n	ن	/n/	нон = نان
В в	V v	و	/v/	номвар = نامور = نامور
Й й	J j	ی، ی	/j/	май = می
واکه‌ها				
Е е	E e	!	/e/	шер = شیر = میروم = میروم = меравам = شیر، شیر
Ё ё	Jo jo	یا	/jɔ/	дарё = آسیاب = осиеб = دریا،
Й й	Ī ī	ئی	/i/	зебой = زیبائی، زبائی
Я я	Ja ja	یه، ی	/ja/	ягонагӣ = یگانگی
А а	A a	اَ، ا	/a/	санг = سنگ = سنگ
И и	I i	اِ، ا	/i/	ихтиёр = اختیار
У у	U u	اُ، ا	/u/	дуд = دود
О о	O o	اَ، ا	/ɔ/	орзу = آرزو
Ў ў	Ū ū	اووی	/ø/	Наврӯзи = اُوو = نُوروزی،
Э э	E e	!	/e/	эй = ای

سرلیک	لات	فارسی	(IPA)	نمونه
Ю ю	Ju ju	یو، ی	/ju/	июн = یون ^۱

در خط سرلیک ۲۴ همخوان و بیش از ۶ واکه مطابق آنچه در فارسی ایرانی مشاهده میشود، وجود دارد. در فارسی تاجیکی برای همخوانهای ث، ص، س تنها یک نویسه (C) و برای همخوانهای ز، ذ، ض و ظ نویسه (3) و برای همخوانهای ح و ه از نویسه (X) استفاده میشود. اما در مورد (غ) نویسه E و برای (ق) نویسه K وجود دارد. بنابراین در مورد تبدیل واژگان عربی واردشده به زبان فارسی تاجیکی، حروف دخیل عربی وجود ندارد. این موضوع باعث تاثیر پذیرفتن کمتر خط مذکور از زبان عربی به نسبت تاثیر آن بر روی خط فارسی شده است. در زمینه واکه‌ها اما تاثیر کلمات روسی و خط روسی بر خط سرلیک زبان فارسی بسیار زیاد بوده است. بعنوان مثال نویسه Ю ю برای (یو) در واژه‌هایی که از روسی وام گرفته شده‌اند استفاده میشود و برخی آنرا جزء واکه‌های سرلیک بحساب نمی‌آورند. ضمناً در زبان تاجیکی علامتهای И و Й و Е به ترتیب برای صدای کسره اضافه و یای نسبت و یای نکره استفاده میشود. وجود برخی از حروف در الفبای تاجیکی از آنجاست که واضعان خط میخواستند کلیه حروف الفبای روسی را در بافت خط تاجیکی به مصرف برسانند. (برجیان، ۱۳۷۸: ۱۰۵)

همچنین شکل هر علامت در الفبای سرلیک ثابت است و برحسب جای آن در کلمه تغییر نمی‌کند لذا چاپ خط سرلیک قدری آسانتر از چاپ خط فارسی است. بعلاوه در خط سرلیک رایج در تاجیکستان - برخلاف خط فارسی که تعدادی از حرفها نقطه دارد- تنها یک علامت نقطه دار هست. (برجیان، ۱۳۷۲: ۱۷۱)

برای ارزیابی نویسه‌گردانی خودکار بین دو خط فارسی و سرلیک از برنامه رایانه‌ای تحت وب به نشانی اینترنتی <http://www.persian-tajik.ir> استفاده شده است. عملکرد این برنامه در نسخه اولیه بر اساس الگوهای زبانی ترسیم شده بود. در روشهای مبتنی بر الگو، ورودی (معمولاً متن) به دنبال الگو و یا کلمه کلیدی خاص که نشانگر رابطه مفهومی خاصی است میگردد. این الگوها انواع مختلفی اعم از نحوی یا معنایی و عمومی یا خاص دارند و برای استخراج عناصر مختلف هستان شناسی^۱ مثل روابط طبقه‌ای یا غیر طبقه‌ای و یا اصول بدیهی بکار میروند (شمس-فرد، ۱۳۸۱: ۵۲) در نسخه اولیه این سامانه، الگوهای زبانی تنها در مورد برخی واژه‌ها و لغات به صورت صحیح نتیجه دلخواه را فراهم کرده بود. پس از مواجه شدن با دشواریهای نویسه‌گردانی بین

^۱ Ontology

دو خط فارسی و تاجیکی و عدم توفیق در نویسه‌گردانی همه لغات از طریق الگوهای زبانی، در نسخه دوم که نسخه فعلی سامانه نیز میباشد از الگوریتم اتوماتای سلولی^۱ برای درج اطلاعات در بانک داده‌های طراحی شده استفاده شده است. الگوریتم اتوماتای سلولی برآمده از نظریه اتوماتا (Automata theory) یا نظریه ماشینها در علوم رایانه است که عبارت است از بررسی ریاضی-گونه ماشینهای محاسبه‌گر انتزاعی و تواناییهای آنها برای حل مسایل. به این ماشینهای انتزاعی اتوماتا گفته میشود. اما اتوماتای سلولی به شبکه‌ای از فضاها اطلاق میشود که به هر خانه آن یک سلول گفته شده و قوانین اتوماتای سلولی نحوه تاثیر پذیرفتن از سلولهای همسایه را مشخص میکنند. بر اساس این الگوریتم، برای طراحی بانک اطلاعات مورد نظر ابتدا جدولی با ۵ فیلد اصلی طراحی گردیده است. این فیلدها شامل id یا شناسه اصلی، واژه فارسی، معادل سرلیک، برجسب و یک فیلد اضافه است که برای آوانویسی واژه‌ها در نظر گرفته شده و البته در حال حاضر فاقد این نوع محتواست. تعداد رکوردهای ثبت شده در این بانک اطلاعات بیش از بیست هزار سطر میباشد. بانک اطلاعات مورد استفاده در این سامانه، حکم همان پیکره واژگانی را دارد که در فیلد چهارم بانک اطلاعات با برجسبهای همچون اسم، فعل، حرف اضافه و امثال آن برجسب‌دهی شده است. همچنین برای تشخیص اسامی خاص، این سامانه از الگوریتم شبکه عصبی واژگانی استفاده میکند. این سامانه با زبان برنامه نویسی C sharp طراحی شده و از بانک اطلاعاتی MS SQL استفاده میکند. این سامانه از دو بخش فرهنگ لغات و نویسه‌گردان متن تشکیل شده است و برای ارائه نتایج مطلوبتر از پیش‌پردازش بعنوان مرحله نخست نویسه‌گردانی استفاده میکند. در این مرحله اختلاف شیوه نگارش میشوند و پسوندها در دو خط فارسی و سرلیک بصورت دقیق مورد ارزیابی قرار میگیرد. بین ۶۰ تا ۷۰ پیشوند و پسوند از جمله نشانه‌های استمرار- نشانه‌های جمع ان، ها، ات و امثال آن در مرحله پیش پردازش شناسایی میشوند. لازم بذکر است کلیه این مراحل در چند ثانیه و همزمان با تبدیل نویسه‌ها انجام میشود.

در ارزیابی تبدیل ماشینی متون بین دو زبان مختلف سه سطح آوایی، معنایی و واژگانی میتواند مورد توجه قرار میگیرد. اما در این پژوهش سطح معنایی در نظر گرفته نمیشود چون معنای واحدی در یک زبان با دو رسم‌الخط وجود دارد. همچنین دستور زبان تاجیکی با اندکی اختلاف همان دستور زبان فارسی است. بنابراین از سطح واژگان و ارزیابی آن فراتر نخواهیم رفت. برای ارزیابی عملکرد این سامانه در مورد تبدیل واژگان نوشته شده در دو مرحله از خط سرلیک به

^۱ Cellular automation

فارسی و بالعکس نویسه‌گردانی نمونه واژگان ساده و مرکب و جملات متنی در این سامانه مورد آزمایش قرار گرفت.

نتیجه بدست آمده حاکی از آن بود که در مرحله تبدیل فارسی به سرلیک کلمات مفرد و اسامی، بدون درج هرگونه اعراب به درستی به نویسه‌های سرلیک تبدیل میشود و البته در مورد لغات مشابه (مانند مَلک، مَلک و مِلک) نویسه ارائه شده در زبان مقصد تنها یک حالت از زبان مبدا را نمایش میدهد. در صورت استفاده از واژگان مرکب اعم از افعال یا ترکیبات وصفی و اضافی و یا نشانه‌های استمرار و جمع که شیوه نگارش آن بیش از یک حالت در رسم‌الخط فارسی نمود مییابد این سامانه دچار خطا در نمایش معادل صحیح میشود. در مرحله تبدیل سرلیک به فارسی در سطح واژه‌های ساده یا مرکب، پیشوندها یا پسوندها و نشانه‌ها همانطور که خواننده میشوند نوشته میشود. عبارت دیگر در اصطلاح زبان‌شناسی تابع قوانین فونتیک است. (برجیان، ۱۳۷۲: ۱۷۱)

برای ارزیابی بهتر مبدل نویسه‌ها در متن، سطوح مختلفی از سادگی و پیچیدگی جملات در نظر گرفته شده است. طول جملات به فراخور تعداد افعال در یک جمله تغییر مییابد و در هنگام تغییر جهت متن قابل نمایش از سرلیک چپ چین به فارسی راست چین، چالش‌های جدیدی پیش روی هوش مصنوعی این سامانه قرار میگیرد. در سطح اول از یک جمله ساده برای برگردان از فارسی به تاجیکی استفاده شده است. همان متن را به بخش تاجیکی به فارسی سپرده و در ادامه برای آوانویسه‌های لاتینی آن، جمله تاجیکی به لاتین آوانویسی شده است.

جمله نخست: «من ایرانی هستم.» که معادل سرلیک آن برابر با عبارت زیر بود:

мани йронии ҳастам

افزایش مصوت بلند И در انتهای واژه نخست ناشی از اشتباه در تحلیل پسوندها توسط این سامانه است. آوا نویسه لاتینی عبارت ساده تاجیکی به شکل زیر نمایش داده میشود:

man īronii hastam

در صورتیکه فردی قادر به درک زبان فارسی و ناتوان از خواندن خط سرلیک باشد، از طریق این روش میتواند متن سرلیک را با الفبای لاتینی مشاهده نموده و یا از طریق سامانه‌های تبدیل نوشتار به گفتار یا Text To Speech که معمولاً از خط لاتینی استفاده میکنند تلفظ تاجیکی آن را نیز بشنود. این ابزارها که از آنها با عنوان کوچک شده TTS یاد میشود معمولاً برای نابینایان و موارد آموزشی کاربردهای بسیار زیادی دارد. لازم بذکر است برای تبدیل آوانوخته سرلیک به لاتین از نویسه‌گردان معتبر لاتین به سرلیک در نشانی lexilogos.com استفاده شده است.

در عبارت دوم که با سطح دشواری پیچیده تری از نظر تشخیص ساختار جمله مطرح شده، نتیجه عملکرد اندکی متفاوت بود. در جمله فارسی ارائه شده به این نویسه‌گردان عبارت زیر به سامانه عرضه شده و نتیجه زیر مشاهده گردید:

ترجمه ماشینی روشی نوین برای فن ترجمه است.

тарҷумаи мошинии روشи навин барои фан тарҷума аст

در تحلیل این عبارت که از سطح دشواری و پیچیدگی نسبتاً بیشتری برخوردار است تنها واژه برگردان نشده «روشی» و تنها مصوت از قلم افتاده И معادل کسره اضافه برای واژه فن میباشد. همچنین تشدید در این کلمه نیز قابل نمایش نیست. علت عدم برگردان واژه روشی مربوط به نقص پیکره دادگانی ارائه شده به این سیستم است، زیرا با حذف یای نکره از انتهای این کلمه معادل صحیح آن نمایش داده میشود. قطعاً تعداد بسیار زیادی از اینگونه واژگان در سطح پیکره دادگان مبداء و مقصد وجود دارد که با تکمیل آن میتوان از بروز نقایصی اینچنین جلوگیری کرد. این امر در بیشتر پیکره‌های دادگان زبان فارسی قابل مشاهده است. عدم نمایش برگردان روشی ناشی از عدم تشخیص «ی» نکره در این واژه است. صحت این ادعا در حذف این حرف و برگردان مجدد قابل مشاهده است: ترجمه ماشینی روش نوین برای فن ترجمه است.

тарҷумаи мошинии raviši navin baroi fan tarçuma аст

اما عدم تشخیص کسره اضافه در واژه فن به عدم اتکای این نویسه‌گردان به حوزه نحوی و استناد آن به واژگان قابل نمایش در هر جمله است. این نویسه‌گردان بصورت عمومی برای بیشتر واژگان از علامت И برای نمایش کسره اضافه استفاده میکند. اما تشخیص کاربرد صحیح آن امری دشوار است. نکته دیگر در ترجمه این عبارت عدم اعمال تشدید در حرف نون واژه فن است. این تشدید نوای گفتار را در زبان فارسی و از روی تاکید بر این حرف نشان میدهند. در حالیکه این تاکید در برگردان سرلیک آن قابل مشاهده نیست. در سطح سوم از برگردان برای آوانویسی عبارت سرلیک به حروف لاتین، مجدد آنرا به ماشین ترجمه سپرده‌ایم. نتیجه برگردان به لاتین بصورت زیر قابل نمایش بوده است.

tarçumai mošiniî raviši navin baroi fan tarçuma аст.

شیوه نمایش چپ به راست، صحت عملکرد برگردان آوانویسه و درست بودن قدرت تشخیص همگی نشان دهنده توان این نویسه‌گردان و کاربرد موثر آن در استفاده و انتقال بین زبانی است. این عملکرد تا سطح عبارتهای ساده سطح یک و دو به خوبی قابل مشاهده است. اما در سطح سوم رابطه واژگان در سطح عبارت و علائم نقطه گذاری ماهیت نویسه‌گردانی را به محک واقعی میسجد. در عبارت سوم چند جمله فارسی دارای چند فعل بصورت یک متن واحد مورد ارزیابی قرار گرفت.

جمله مرجع فارسی:

چرا تصمیم به تدوین کتابی جدید در زمینه زبان‌شناسی رایانه‌ای گرفته‌ایم؟
اهداف اصلی، خوانندگان مورد نظر، ویژگی‌های اصلی و رابطه این کتاب با
شاخه‌های مختلف علم کامپیوتر، چه چیزهایی هستند؟

برگردان سرلیک:

чиро тасмим ба тадвини китобии чадид дар заминаи
забони шиносии роёнаи аии гирифтаи ими ? аҳдофи
аслӣ , хонандагони мавриди назар , вижагӣҳои аслии
робитаи ин китоб бо шоха ҳоии мухталифи илми
компютер , чӣ чӣзҳои ҳастанди?

نتیجه بررسی عملکرد این نویسه‌گردان در سطح ترجمه عبارتهای فارسی به سرلیک با درجه
دشواری ۳ اندکی متفاوت از دو مرحله قبل می‌باشد. در این مرحله مصوت بلند И در برخی موارد
مازاد بر ضرورت در برخی واژه‌ها نمایش می‌یابد. در صورت رعایت نکردن رسم‌الخط فارسی و عدم
رعایت فاصله‌ها در این زبان امکان انتقال آن از طریق خط سرلیک وجود ندارد. این نویسه‌گردان
برای همزه کوچک یا جایگزین امروزی آن که همان «ی» در واژه‌هایی مانند «کوچه‌ی بن بست» و
«بسته بزرگ» که هر دو صورت آن در زبان فارسی استفاده می‌شود از همان نشانه И استفاده می‌کند
و در برخی موارد که خواننده فارسی زبان در برابر متنی که به خط فارسی نوشته شده، از طریق
هوش ناخودآگاه خود قادر به تشخیص شیوه درست عبارت می‌باشد، این سیستم قادر به تشخیص
صحیح جمله نیست. رعایت چینش چپ به راست، رعایت نشانه گذاری صحیح مانند علالت سوال و
ویرگول از نقاط قوت این مرحله و جداسازی بی مورد، عدم درک نیم فاصله در واژه‌های بیش از
یک بخش مانند «شاخه‌های» و امثال آن از نقاط ضعف سیستم هستند. در مجموع این نویسه-
گردان در بخش فارسی به سرلیک بیش از ۹۰٪ صحت عملکرد از خود نشان می‌دهد.
آوا نوشته این عبارتها به زبان انگلیسی نشان دهنده میزان دقت آن در تبدیل آوایی واژگان مورد
اشاره است:

ciro tasmim ba tadvini kitobii çadid dar zaminai
zabonşinosii royonai giriftaiimi ? ahdofi aslī ,
xonandagoni mavridi nazar , vizagīhoi aslii robitai in
kitob bo şoxa hoii muxtalifi ilmi kompiuter , cī cīzhoi
hastandi ?

در مرحله تبدیل واژگان بین دو خط فارسی و سرلیک تفاوت‌هایی بین دو رسم‌الخط وجود دارد که
در انتقال بین خطوط چالش‌های اساسی برای مبدل میانی ایجاد خواهد کرد. از آن جمله می‌توان به
موارد زیر در قالب جمع‌بندی نکات اصلی در این بخش اشاره کرد:

- یکسان بودن نویسه C سرلیک برای همخوانهای ث/ص/س فارسی و Z سرلیک بجای ز/ض/ظ/ذ

- یکسان بودن نویسه /و/ فارسی برای همخوانهای y, в, ā در واژه‌های تو/Ту/ ولی /вали/ و روشن /рӯшан/
- عدم نگارش مصوت‌های کوتاه برای تفکیک معنایی واژه‌هایی که از همخوانهای مشابه استفاده میکنند مانند مَلک، مُلک و مِلک در خط فارسی
- ناتوانی در مرزبندی واژه‌ها و وجود فاصله زائد و عدم تشخیص نیم فاصله تایی در تبدیل متون فارسی به سرلیک
- عدم درج و نمایش کسره اضافه در ترکیب‌های وصفی و اضافی
- ناتوانی خط فارسی برای ضبط و ثبت وام‌واژه‌های لاتینی و عربی که معمولاً معادل‌هایی بیش از یک حالت دارند.
- عدم نمایش تشدید در همه واژه‌ها در خط فارسی
- ناتوانی خط فارسی در نمایش واژه‌های مربوط به ایران باستان و میانه به دلیل حذف شدن برخی همخوانها
- افزایش ضریب امکان اشتباهات چاپی در نگارش
- پیوند پذیر بودن برخی حروف و پیوند ناپذیر بودن برخی دیگر
- چپ‌چین بودن نگارش اعداد و فرمول‌های ریاضی و راست‌چین بودن همخوانها در خط فارسی
- ناتوانی خط فارسی برای کلمات مخفف مشابه آنچه در لاتین وجود دارد مانند: NASA نویسنده معتقد است کارکرد اصلی این سامانه و توان اصلی انتقال بین دو خط مختلف از سمت سرلیک به فارسی با اشکالات کمتری همراه است. این موضوع ناشی از تسلط طراحان این سامانه به دستور زبان فارسی و روانتر بودن ساختار معادل‌یابی حروف در سرلیک است. همچنین، سامانه مورد نظر دارای قابلیت تبدیل حروف سرلیک به فارسی است که نقطه قوت آن نیز محسوب میشود. کارکرد مربوطه به انتقال بین زبانی از سمت سرلیک به فارسی با برگرفتن چند مطلب از سایتهای اینترنتی و سپردن آنها به این سامانه مورد ارزیابی قرار گرفت. در این مرحله بدلیل مثبت بودن نتایج آزمایشات صورت گرفته در سطح واژگان و جملات ساده، مستقیماً به تبدیل متن پرداخته شده است.
- نتایج بدست آمده، هنگام ارائه متن از خط سرلیک به فارسی از ضریب اطمینان بالایی برخوردار بود. برای این موضوع نوع مطلب و یا واژگان خاص یا عمومی بودن آن چالشی برای نویسه‌گردان محسوب نشده و با الگوریتم بکار گرفته شده در آن، متون مربوطه با درجه دقت بالایی به خط فارسی نمایش داده شده است.
- یک متن خبری به خط سرلیک از آرشیو پایگاه اطلاع‌رسانی ملت به تاریخ 01/09/2012 گرفته شده و معادل نویسه‌گردانی شده آن به فارسی نیز بوسیله سامانه مورد نظر استخراج شد. این متن

بعنوان یک گزینه منتخب جهت ارزیابی عملکرد مبدل سرلیک به فارسی انتخاب گردیده است. لازم بذکر است هیچگونه دخل و تصرفی در فرآیند خودکار نویسه‌گردانی متن مورد نظر به فارسی صورت نگرفته و اصل دو منبع اخذ شده و حروف گردانی شده در پایان مقاله آمده است. از میان ۴۸۵ واژه متن سرلیک که ۲۶۵۴ نویسه را در خود جای داده‌اند، تمامی واژگان بنحوی نویسه‌گردانی شده‌اند که مخاطب فارسی‌زبان ایرانی به خوبی قادر به درک تمامی آن باشد و تنها ۴,۶٪ لغات (معادل ۲۲ مورد) اشکال در این فرآیند مشاهده شده که دلایل بروز این اشکالات در پی آمده است. ویژگیهای محل اشکال در این نویسه‌گردانی در بخشهای زیر آمده است:

اختلاف گنجواژه‌ها

بخش اول ناشی از عدم وجود برخی واژه‌ها در فرهنگها و پیکره زبان فارسی ایرانی است. نویسه‌گردان برای تبدیل واژه‌های سرلیک به فارسی در صورت مواجهه با لغاتی خارج از پیکره زبانی خود از همان الگوریتم تبدیل نویسه‌ها به یکدیگر استفاده کرده و در نتیجه، هیچ واژه برگردان نشده‌ای وجود ندارد. اما واژه‌هایی چون «قصدن» که از این نوع هستند و معادل آن در بین ایرانیان فارسی زبان بصورت عمداً یا از روی عمد استفاده میشود، در پیکره زبانی مقصد وجود نداشته، نویسه‌گردان از روی حروف موجود در دو خط آنرا تبدیل کرده است. هرچند این ابزار واژه مورد نظر را بصورت **قسدن** نمایش داده اما، بروز اختلاف در اینگونه واژه‌ها ناشی از شیوه رسم‌الخط فارسی است که به دلیل استفاده از حروف عربی باعث پدید آمدن یک نویسه برای صامتهای هم آوا شده است. طبق جدولی که پیشتر نمایش داده شد، سه نویسه ص، ث، س در فارسی دارای یک شکل نویسه C هستند. از اینرو واژه‌هایی مثل سبا (نام سرزمین) و صبا به معنی نام یکی از بادهای معروف، نام دختر و نام خانوادگی و... در سرلیک دارای تفاوت در نویسه نیستند. هرچند شکل ظاهری آنها در خط فارسی متفاوت بوده و نویسه‌گردانی یکسان منجر به تولید غلط املائی برای مخاطبان در خط فارسی خواهد شد. وجود حروف اضافه و ربط خاص زبان تاجیکی مانند qati (در معنی با و به همراهی) و یا واژه‌های خاص تاجیکی مانند کار چلانی (سوداگری) (باربندی - بسته بندی) که در سوی دیگر پیکره زبانی فارسی زبانان ایرانی منسوخ شده و یا اصلاً وجود نداشته میتواند مزید بر علت عدم موفقیت نویسه‌گردانی صحیح شود.

اختلاف گویش

بخش دوم اشکالات بوجود آمده ناشی از شیوه گفتار فارسی زبانان تاجیکستان است که در گویش خود شیوه‌ای متفاوت در ادای برخی آواها در واژگان را بکار میگیرند. سبک نوشتار نیز تحت تاثیر گویش منجر به بروز اختلافاتی در نویسه‌گردانی واژگان شده است. در حقیقت یکی از نقایصی که همزمان با تغییر خط از فارسی به لاتین و سرلیک پدید آمد همین تفاوت در املائی کلمات بود. این

اختلاف بیش از هرچیز ناشی از کثرت لهجه‌های فارسی است و از آنجاکه خط سرلیک بر خلاف خط فارسی قابلیت انعکاس تلفظ را دارا بود {در همان ابتدای زمان تغییر خط} هر کس سهواً یا عمداً کلمات را مطابق تلفظ ولایت خودش مینوشت بنابراین بسیاری از کلمات چند صورت مکتوب پیدا کرد. (برجیان، ۱۳۷۲: ۱۷۴)

هرچند تلاش‌هایی برای یکسان سازی این رسم الخط انجام شده ولی اختلاف حاضر بین دو گویش فارسی و تاجیکی منجر به پیدایش تفاوت در نگارش نویسه‌ها بر اساس فونتیکی متفاوت شده است. از جمله کوتاه نام بی. بی. سی که در گویش تاجیکی بصورت ب. ب. سی تلفظ شده و بهمان صورت در متن اصلی به خط سرلیک نمود یافته، هنگام تبدیل آن به خط فارسی معادل نگارشی صحیح و مطابق با رسم الخط تاجیکی نمایش مییابد. نویسه گردانی واژه‌هایی مانند فیستوال/ فستیوال، سینما / سینما، رادیوی / رادیویی، دیداربینی/ دیداربینی، دگر / دیگر، چار / چهار، سأل / سوال از این جمله هستند.

نکته دیگر که البته مربوط به ساختار برنامه نویسی این نویسه‌گردان است، تکرار برخی از واژه‌هاست. در پاراگراف پنجم واژه ملت دو بار تکرار شده و این موضوع ارتباطی با موضوع نویسه‌گردانی ندارد.

صورت نوشتار که شکل گفتاری را در خود جای داده، باعث می‌گردد تا برخی از واجها در دو سوی فارسی ایرانی و تاجیکی متفاوت شوند. بنابراین بصری تغییر و جایگزین کردن یک واکه یا همخوان با معادل آن در آنسوی زبان فارسی نمیتوان به نتیجه درستی در تبدیل حروف دست یافت. به عنوان مثال همخوان (غ) در فارسی تاجیکی کار تبدیل آن به فارسی ایرانی را با چالش رو به رو میسازد. بنابراین در تبدیل واژه‌هایی از جنس زیر کار حروف گردانی بدون داشتن پیکره زبان و مبدل بین زبانی به نتیجه دلخواه نمیرسد.

تاجیکی

ایرانی

qimat

qeymat

[i] در مقابل [ey]

Neft

Naft

[e] در مقابل [a]

Zinda

Zende

[i] در مقابل [e]

ʔavōz

ʔâvâz

[a] در مقابل [o]

همچنین فارسی زبانان تاجیک برخلاف فارسی زبانان ایرانی و با پیروی از شیوه گفتار و نوشتار روسی، قادر به استفاده از زنجیره CCV بجای CVC زبان فارسی ایرانی هستند.

škâf

ʔeš kâf

Stalin

ʔestâlin

اختلاف در نویسه‌ها

یکسان بودن نویسه X بجای دو صامت ح و خ نیز باعث پدیدار شدن برخی اشکالات میشود. در انتقال واژه خوبی در عبارت «که این بهانه خوبی برای دیداربینی با مخاطبان بی بی سی در تاجیکستان خواهد بود» حرف خ در واژه خوبی بدلیل یکسان بودن با آوای ح و مصوت بلند او با مصوت کوتاه که در رسم‌الخط فارسی نوشته نمیشود جایگزین شده است این موضوع در واژه هم در پاراگراف پنجم که بصورت خم نمایش داده شده نیز مشاهده میشود.

در برخی واژه‌های دیگر نیز اشتباه در تبدیل مشاهده میشود. از جمله در واژه‌های تلویزان بجای تلویزیون. در حقیقت شیوه درست این واژه در سرلیک در پاراگراف هفتم بصورت **телевизиюн** نمایش داده شده است. همانطور که پیشتر ذکر شده نویسه **ю** بعنوان معادل آوای **IO** مورد استفاده قرار میگیرد و در ۳ بار استفاده از این واژه در متن مورد نظر، نویسنده متن اصلی سرلیک تنها یکبار آنرا به صورت صحیح تایپ کرده که در همان یکبار هم بصورت صحیح نویسه گردانی شده است. این نویسه که در کنار ۳ نویسه دیگر بعنوان حروف (یات بر سر) شناخته میشوند. در خط سرلیک تاجیکی سه نشانه **я, ё, ю** بجای خوشه‌های (یو) و (ی) و (یا) در خط لاتینی بکار میروند؛ به دو منظور: یکی کوتاه نویسی و دیگری مصرف غیر ضروری حروف روسی. (برجیان، ۱۳۷۸: ۱۱۲)

اختلاف در سبک نوشتار

در هنگام حروف گردانی دستی یا ماشینی، بیشترین تغییرات در شیوه نگارش براساس تفاوت‌های تصریفی واژگان بوجود می‌آید.

– در سرلیک نشانه‌های جمع و مفعول بیواسطه همیشه پیوسته نوشته میشوند، گویی که نقش پسوندهای تصریف نام (اسم و صفت و ضمیر) را دارند. (برجیان حبیب، ۱۳۷۸: ۱۱۲) ولی در فارسی نشانه‌های جمع در برخی موارد منفصل و در بیشتر موارد بصورت چسبیده با رعایت نیم فاصله نوشته میشوند. در سرلیک، نیم فاصله مفهومی ندارد. مانند: **Филмҳои** (فیلمهای) در فارسی بصورت فیلمهای و درختها

– نشانه‌های استمراری می و همی بصورت متصل به نگارش درمی‌آید: **Miravam** - **Meravam** وند نمود که برای افعال استمراری بکار می‌رود در زبان فارسی بصورت پیش وند + نشانه استمرار + فعل بکار می‌رود ولی در فارسی تاجیکی جای پیشوند و نشانه استمرار عوض شده و تغییری در واکه این نشانه نیز حادث میشود. مانند **بر میدارد** - (میبرداشت) **meibardost** همچنین افعال استمراری بدون وند نیز معنا می‌یابند مانند: **میشناسم** -

Šinosam شناسم

- (را) بعنوان نشانه مفعول بیواسطه در تمام موارد در سرلیک بصورت چسبیده می‌آید:
нишастро (نشست را) که بصورت (نشسترا) نوشته میشود.
- همچنین (واو عطف) در نوشتار بصورت چسبیده آمده و «واو معدوله» در کلماتی مانند *خواب* و *خواهر* نمایش نمییابد و بصورت *خاب* یا *خاهر* نوشته میشوند.
- اسامی که با عدد مبهم همراه هستند و در زبان فارسی ایرانی فاقد نشانه جمع تلقی میشوند و در فارسی تاجیکی با نشانه جمع همراه میشوند: مانند چهارکس **čork as-on** که بصورت چارکسان نوشته میشود.
- گونه‌ای از نشانه جمع بصورت *هان* وجود دارد در کلمات *ماهان* و *شماهان* (ما و شما) دیده میشود. این نشانه در فارسی ایرانی بکار نمیرود. (کلباسی، ۱۳۷۴: ۶۴)
- در سرلیک کلمات تنوین دار و مشدد عربی بسیار کم کاربرد هستند. کلمات تنوین دار مانند *قصدن* بصورت *آوانوشته* به نمایش درآمده و نشانه مستقلی برای آن وجود ندارد. در مورد تشدید نیز بجز تکرار حرف برای نمایش آنها نماد دیگری وجود ندارد ولی در نسخه فارسی این عناصر به خوبی رعایت شده‌اند.
- نشانه اضافه که در فارسی ایرانی بصورت کسره بعد از مضاف یا موصوف می‌آید و معمولاً مشاهده نمیشود در فارسی تاجیکی بصورت واکه‌ای مستقل و قابل مشاهده ذکر میشود. اینگونه واژگان بصورت متصل به نگارش درمی‌آیند. پدر من - **pedariman** پسر بزرگ - **pisarikalon**
- صفت و موصوفهایی که معمولاً بصورت مقلوب نیز کاربرد پیدا کرده و در شیوه نگارش مردم تاجیکستان به نمایش در می‌آیند معمولاً بصورت متصل و بهم پیوسته نوشته میشوند. هرچند این نمونه‌ها در فارسی ایرانی بصورت منفصل نوشته میشوند. دختر جوان - **javo Duxtar**
- همچنین صفت‌های برتر و عالی در تمامی موارد بصورت متصل به نگارش درمی‌آیند. سریعترین - **saritarin**
- صفت اشاره آن و این همیشه بصورت متصل به موصوف خود آورده می‌آید مانند آن مردان - **onmardon**
- صفت اشاره تاکیدی همین و همان در همین خانه متصل می‌آید: **haminxona**
- افعال مرکب در همه موارد بدون رعایت فاصله میانی و بصورت بهم چسبیده در سرلیک نمایش مییابد ولی در فارسی با رعایت نیم فاصله یا فاصله کامل و این موضوع در تمامی موارد نویسه‌گردانی شده رعایت شده است.
- در سرلیک همخوانهای همزه و عین از یک نویسه یکسان استفاده میکنند در حالیکه در خط فارسی کلماتی که با این دو نویسه شکل مییابند کاملاً متفاوت از یکدیگرند. در شناسایی این نویسه‌ها در برگردان فارسی بجز یک مورد در تمامی موارد صحیح عمل شده است.

- است در پایان جملات همیشه بصورت متصل آمده‌است. مانند: **Nišastast** بجای نشسته است.

- در حوزه فعل نیز اختلاف سبک نوشتار فارسی تاجیکی با ایرانی باعث پدید آمدن مشکلاتی میشود. بعنوان مثال فعل حال در حال انجام احتمالی شاید دارم میروم در زبان تاجیکی به صورت (رفته استادگی استم) **raftaistodagistam** نمایش داده میشود.

در زمان حروف‌گردانی از خط تاجیکی به فارسی در دو مرحله میتوان کیفیت متن برگردانده شده را اصلاح نمود. سطح اول بلافاصله بعد از تبدیل حروف و کلمات به خط مقصد حاصل میشود که البته بصورت نسبی برای گویشوران آن زبان قابل درک و فهم است و سطح دوم معمولاً پس از اعمال تغییراتی که از پی سالها دوری گویشوران دو زبان بوجود آمده و تا تبدیل شدن به متنی مشابه آنچه در متن مقصد وجود دارد راه درازی را باید بپیماید.

۵. نتیجه گیری

تفاوت در سبک نوشتار فارسی تاجیکی و ایرانی بصورت کلی در شاخه‌های زیر قابل تفکیک و طبقه‌بندی است.

تفاوت‌های سبک واجی

تفاوت‌های تصریفی واژگان

تفاوت‌های دستوری و ساخت جمله

در خط سرلیک واژگان مرکب اعم از افعال، اسامی و ترکیبات بصورت بدون فاصله و بهم پیوسته نمایش داده میشوند. نویسه گردان خط سرلیک به فارسی با نمایش ۹۵٫۴٪ صحت عملکرد در نویسه گردانی واژگان انتقالی و استفاده ترکیبی از الگوریتم اتوماتای سلولی و پیکره واژگان، نتیجه قابل قبولی از خود ارائه داده است اما نسخه فارسی به سرلیک آن هنوز دارای ایرادهای ساختاری بسیاری است که باید اصلاح شود.

برای اصلاح ساختار این نویسه گردان و تبدیل آن به ابزاری که سطح بالایی از انتقال صحیح واژگان را در دو زبان نشان دهد، ابتدا لازم است مدخلهای پیکره زبانی فارسی از پیکره ۲۰,۰۰۰ عددی حاضر به پیکره ۶۰,۰۰۰ عددی فرهنگستان زبان فارسی ارتقا یابد و پیکره سرلیک مورد استفاده نیز ایجاد و بعنوان برابر نهاده‌های فارسی به سرلیک بصورت عمومی و تخصصی تولید و مورد استفاده قرار گیرد.

در خصوص آثاری که بصورت تصویری اعم از میکرو فیلم و دستنویس در منابع هر دو خط وجود دارند، استفاده از نویسه خوان نوری^۱ ضروری است. درحالیکه هیچگونه روشی بجز متن تایپ شده بعنوان ورودی در این سامانه پیش بینی نشده است. ضروری است با استفاده از فناوریهای موجود ابزارهای تبدیل تصویر به متن امکان گسترش شیوه های ورود متن در هر دو زبان گسترش یابد. در نسخه فارسی به سرلیک این نویسه‌گردان، برای کشف ارتباط برخی واژه‌های بهم پیوسته و عدم تفکیک آنها ضروری است تمامی واژه‌ها برچسب گذاری گردند. در اینصورت به فرض جدانویسی واژه‌های بهم چسبیده و یا بالعکس، نمایش نویسه‌گردانی آن به خط سرلیک دچار اشکال نمیگردد. سامانه‌های پیشرفته تبدیل متون همچون گوگل از حیث شیوه انتخاب بهترین برابر نهاده، برپایه تحلیل‌های آماری و شبیه مغز انسان عمل میکند. بعنوان یک انسان در هنگام مواجهه با یک واژه بیگانه در مغز خود در پی یافتن بهترین برابر نهاد ممکن بر می‌آییم و این امر ممکن است در کسری از ثانیه صورت گیرد. با آموختن بیشتر، برابر نهاده‌های مناسبتر جایگزین موارد قبلی میشود. موتورهای مترجم متن گوگل نیز از همین شیوه استفاده میکنند و در صورت اصلاح یک عبارت یا واژه، واژه منتخب در نتایج بعدی مورد استفاده قرار میگیرد. اگر این سامانه نیز توانایی آموختن از موارد اصلاح شده کاربران و اصلاح خود را داشته باشد شاهد کاهش اشتباهات و افزایش سطح اطمینان در نویسه‌گردانها تا رسیدن به بهره‌وری کامل خواهیم بود. امید است با اصلاح و گسترش ابزارهایی اینچنین، شاهد استفاده بهینه آن در کاربردهای فرهنگی و ادبی و تولید و تبدیل آثار ارزشمندی از هر دو منبع فارسی و سرلیک باشیم.

منابع:

- ۱- -----۱۳۷۲. تجربه تاجیکستان در تغییر خط فارسی. ایران شناسی. شماره ۱۷ صص ۱۷۰-۱۸۳
- ۲- استخراج دانش مفهومی از متن با استفاده از الگوهای زبانی و معنایی، شمس فرد مهرنوش، عبدالله زاده بارفروش احمد. ۱۳۸۱. تازه‌های علوم شناختی. سال ۴ شماره ۲.
- ۳- آشنایی با نظامهای نوشتاری صفوی کوروش. ۱۳۸۶. تهران: نشر پژواک کیوان
- ۴- بررسی علمی شیوه خط فارسی خلخال نازیلا. ۱۳۷۵. چاپ اول. تهران: انتشارات ققنوس
- ۵- تاریخ زبان فارسی، ابوالقاسمی محسن، ۱۳۸۵، چاپ هفتم. تهران: انتشارات سمت

¹ Optical character recognition

- ۶- درآمدی برچگونگی شیوه خط فارسی ،ادیب سلطانی میر شمس الدین، ۱۳۷۸. چاپ سوم. تهران: انتشارات امیر کبیر
- ۷- زبانهای ایرانی. ترجمه دکتر علی اشرف صادقی، ارانسکی ای ام. ۱۳۷۸. تهران: انتشارات سخن
- ۸- ساختمان خط تاجیکی، برجیان حبیب. ۱۳۷۸. ، نشریه نامه فرهنگستان، شماره ۱۳، پاییز صص ۱۰۳-۱۱۶
- ۹- فارسی ایران و تاجیکستان (یک بررسی مقابله‌ای)، کلباسی ایران. ۱۳۷۴. تهران: موسسه چاپ و انتشارات وزارت امور خارجه
- پیوست ۱: متن و خروجی مستقیم نویسه گردان سرلیک به فارسی**

<p>Содиқи Сабо: «Би-би-сӣ қасдан аз пӯшиши ҳаводиси Бадахшон чашмпӯшӣ накардааст.»</p> <p>Содиқи Сабо, аз хабарнигорони маъруф ва раиси бахши форсии Би-би-сӣ, ки ба хотири ширкат дарчашнвораи филмҳои мустанади Би-би-сӣ ва муаррифии Би-би-сӣ ба Тоҷикистон омадааст, дар як нишаст бо хузури хабарнигорони дохилу хориҷи кишвари мо ҳадафи сафараш ба Тоҷикистонро дар миён гузошт. ӯ гуфт, қарор аст аз 31 август то 2 сентябр дар Душанбе фестивали филмҳои мустанади Би-би-сӣ баргузор шавад, ки ин баҳонаи хубе барои дидорбинӣ бо мухотабони Би-би-сӣ дар Тоҷикистон хоҳад буд.</p> <p>Дар ин нишаст матбуотӣ, ки рӯзи 27 август бо мусоидати Иттифоқи журналистон дар шаҳри Душанбе баргузор шуд, аз Би-би-сӣ ба унвони яке аз муътабартарин расонаҳои хабарии ҷаҳон ном бурда шуд, ки ҷомеаи Тоҷикистон аз он таваккуъоти зиёде дорад, вале рӯйдодҳо ва таҳаввулоте, ки дар кишвар иттифоқ меуфтад, дар Би-би-</p>	<p>صادق سبا: «ب-ب-سی» قسدن از پوشش حوادث بدخشان چشم‌پوشی نکرده است.</p> <p>صادق سبا، از خبرنگاران معروف و رئیس بخش فارسی ب-ب-سی، که به خاطر شرکت در جشنواره فیلمهای مستند ب-ب-سی و معرفی ب-ب-سی به تاجیکستان آمده‌است، در یک نشست با حضور خبرنگاران داخل و خارج کشور ما هدف سفرش به تاجیکستان را در میان گذاشت. او گفت، قرار است از ۳۱ آگوست تا ۲ سپتامبر در دوشنبه فیستوال فیلمهای مستند ب-ب-سی برگزار شود، که این بهانه خوبی برای ددارینی با مخاطبان ب-ب-سی در تاجیکستان خواهد بود</p> <p>در این نشست مطبوعاتی، که روز ۲۷ آگوست با مساعدت اتفاق ژورنالیستان در شهر دوشنبه برگزار شد، از ب-ب-سی بعنوان یکی از معتبرترین رسانه‌های خبری جهان نام برده شد، که جامعه تاجیکستان از آن توقعات زیادی دارد، ولی رویدادها و تحولاتی، که در کشور اتفاق میفتد، در ب-ب-سی بگونه‌ای، که انتظار می‌رود، بازتاب نمی‌یابد.</p>
---	--

<p>сй ба гунае, ки интизор меравад, бозтоб намеёбад.</p> <p>Ба пурсиши хабарнигорон Содиқи Сабо, раиси бахши форсии Би-би-сй, Ценнй Нортон, сардабири Сервиси Чаҳонии Би-би-сй, Дено Таҳмосбй, масъули равобити умумии Би-би-сии форсй ва Сухроби Зиё гузоришгари маҳаллии телевизиони форсии Би-би-сй посух гуфтанд.</p> <p>Хабарнигори «Миллат» аз Содиқи Сабо, раиси бахши форсии Би-би-сй пурсид, ки чаро телевизиони форсии Би-би-сй ба пӯшиши ҳаводиси Бадахшон, ки зиёда аз як моҳ боз чараён дорад ва теъдоди зиёде аз афрод низ кушта шудаанд, чандон тавачҷух нашоштааст, дар ҳоле ки ба кушта шудани як нафар дар Ироку дигар ҷойҳо зиёд мепардозад. Агарчи сайти форсии Би-би-сй ин ҳаводисро хеле пӯшиш дод ва бахши русй ҳам ғаёб буд. Ва ин чараёни содае низ набуд, ки Би-би-сии форсй ба он тавачҷух накунад.</p> <p>Содиқи Сабо, дар посух ба ин суол гуфт, ки Ирон ба хотири ин ки ҳар рӯз бо ҷомеаи чаҳонӣ даъво дорад ва дар Афғонистон як ҷанги дохилӣ ҳаст, таҳаввулоти ин кишварҳо дар Би-би-сй бештар мунъакис мешаванд, аммо бахши форсии Би-би-сй амдан ё қасдан аз масоили Тоҷикистон, ба вижа ҳаводиси Бадахшон дар як моҳи ахир, ҷампӯшӣ накардааст: «Дар мавриди Бадахшон, ман воқеан мутаассифам, ки мо натавонистем он тур, ки диламон мехост, пӯшиш бидиҳем. Бадахшон масъалаи муҳиммест. Ба ҳар ҳол, баъд аз солҳо дар бахше аз Тоҷикистон даргирии</p>	<p>бе پرسش خبرنگاران صادق سبا، رئیس بخش فارسی ب-ب-سی، ب - ب-سی، جینی نارتان، سردبیر سرویس جهانی ب-ب-سی، ب - ب-سی، دینا طهماسبی، مسئول روابط عمومی ب - ب-سی فارسی و سهراب ضیاء گزارش گر محلی تیلیوزان فارسی ب - ب-سی پاسخ گفتند.</p> <p>خبرنگار «ملت» «ملت» از صادق سبا، رئیس بخش فارسی ب - ب-سی پرسید، که چرا تیلیوزان فارسی ب - ب-سی به پوشش حوادث بدخشان، که زیاده از یک ماه باز جریان دارد و تعداد زیادی از افراد نیز کشته شده‌اند، چندان توجه نداشته‌است، در حالیکه به کشته شدن یک نفر در عراق و دگر جاهای زیاد میپردازد. اگرچ سایت فارسی ب - ب-سی این حوادث را خیلی پوشش داد و بخش روسی خم فعال بود؛ و این جریان ساده‌ای نیز نبود، که ب - ب-سی فارسی به آن توجه نکند.</p> <p>صادق سبا، در پاسخ به این سال گفت، که ایران بخاطر این که هر روز با جامعه جهانی دعوا دارد و در افغانستان یک جنگ داخلی هست، تحولات این کشورها در ب - ب-سی بیش تر منعکس میشوند، اما بخش فارسی ب - ب-سی آمدن یا قسدن از مسائل تاجیکستان، بویژه حوادث بدخشان در یک ماه اخیر، چشم‌پوشی نکرده‌است: «در مورد بدخشان، من واقعاً متأسفم، که ما نتوانستیم آن تور، که دلمان میخواست، پوشش بدهیم. بدخشان مسأله مهمیست. بهر حال، بعد از سالها در بخشی از تاجیکستان درگیری نظامی بود و عده‌ای کشته شدند. ولی مشکل بزرگ ما این بود، که دولت</p>
--	--

<p>низомӣ буд ва иддае кушта шуданд. Вале мушкили бузурги мо ин буд, ки давлати Тоҷикистон вуруд ба Бадахшонро мамнӯъ карда буд.» Раиси бахши форсии Би-би-сӣ афзуд, ки сайти форсии Би-би-сӣ дар пӯшиши ҳаводиси Бадахшон бехтар амал карда, чун сайт ба тасвир ниёз надорад, вале барои телевизион агар тасвире вучуд надошта бошад, дасти хабарнигорон комилан баста аст. «» Содиқи Сабо ҳамчунин гуфт, эродхо ва дархостҳои ироашуда дар ин нишастро ба барасӣ хоҳад гирифт ва то ҷое имкон дорад, камбудхо рафъ хоҳад шуд. Ё дар поёни нишаст ҳозиринро ба чашнвораи филми Би-би-сӣ, ки дар Кохи синамои пойтахти Тоҷикистон баргузор мегардад, дар рӯзи 31 август даъват кард. Зимнан, фаъолияти Бахши форсии Би-би-сӣ дар Тоҷикистон аз ҳудуди 20 сол пеш, дар замони ҷанги дохилии кишвар шуруъ шудааст. Барномаи роудиюии «Маҷаллаи Осиёи Миёна», ки Би-би-сӣ солиёни зиёде пахш мекард, дар соли 2008 мутаваққиф шуд. Аз ҷор сол ба ин сӯ телевизиёни форсии Би-би-сӣ фаъолият дорад.</p>	<p>таҷикистон ورود به بدخشان را ممنوع کرده بود.» رئیس بخش فارسی ب - ب-سی افزود، که سایت فارسی ب - ب-سی در پوشش حوادث بدخشان بهتر عمل کرده، چون سایت به تصویر نیاز ندارد، ولی برای تلویزیون اگر تصویری وجود نداشته باشد، دست خبرنگاران کاملاً بسته است. «» صادق سبا همچنین گفت، ایرادها و درخواستهای ارائه شده در این نشست را به بررسی خواهد گرفت و تا جای امکان دارد، کمبودها رفع خواهد شد. او در پایان نشست حاضرین را به جشنواره فیلم ب-ب-سی، ب - ب-سی، که در کاخ سنمای پایتخت تاجیکستان برگزار میگردد، در روز ۳۱ آگوست دعوت کرد. زمنن، فعالیت بخش فارسی ب - ب-سی در تاجیکستان از حدود ۲۰ سال پیش، در زمان جنگ داخلی کشور شروع شده است. برنامه رادیوی و «مجله آسیای میان»، که ب - ب-سی سالیان زیادی پخش میکرد، در سال ۲۰۰۸ متوقف شد. از چار سال به این سو تلویزیان فارسی ب - ب-سی فعالیت دارد.</p>
---	---