



کامپیوتر و رسم الخط فارسی*

دکتر عباس خجری
عضو هیأت علمی دانشگاه تهران



• حروف فارسی دارای ویژگیهای متفاوتی است که سرجمع آنها به پیچیدگی کار کامپیوتری کردن اطلاعات فارسی می افزاید.

الف. حروفی در فارسی وجود دارد که محل قرار گرفتن آنها در کلمه، در صورت ظاهر آنها تأثیر دارد؛ نظیر «عین» (به صورت «ع»، «ع»، «ع»، «ع»). این حروف در نظام دستی مشکلی پدید نمی آورد و در خواندن و نوشتن نیز ذهن چنان با آنها آشناست که احساس چندگونگی نمی کند، اما در کامپیوتر، هریک هویتی مستقل و متفاوت می یابد.

ب. حروفی هستند که به حروف پیش از خود می چسبند، ولی به حروف پس از خود نمی چسبند، مانند ا، د، ذ، ر، ز، و هاء غیرملفوظ. در کلماتی که دارای این حروف هستند ممکن است کمی بی دقتی در فاصله گذاری سبب اتصال یا انفصال کاذب در کلمه شود و تغییر معنی ایجاد کند؛ مانند «برنامه» و «برنامه» یا «دوازده» و «دوازده».

ج. حروفی هستند که به دلیل موضع خود بزرگ نوشته می شوند، ولی با قرار گرفتن یا نگرافتن پس از حروف بند «ب» به آنها می چسبند یا جدا نوشته می شوند؛ مانند «ل» در «فضل» و «سال» یا «ن» در «من» و «زن». متصل یا منفصل نوشته شدن آنها ممکن است از نظر کامپیوتر سبب یگانگی یا چندگانگی آنها شود.

د. حروفی هستند که به مناسبت و به قیاس حرف آخر کلمه بزرگ نوشته می شوند، ولی درواقع در میان کلمه قرار دارند؛ نظیر «ش» در «دانش آموز» و «ت» در «زیست شناس». در این نوع

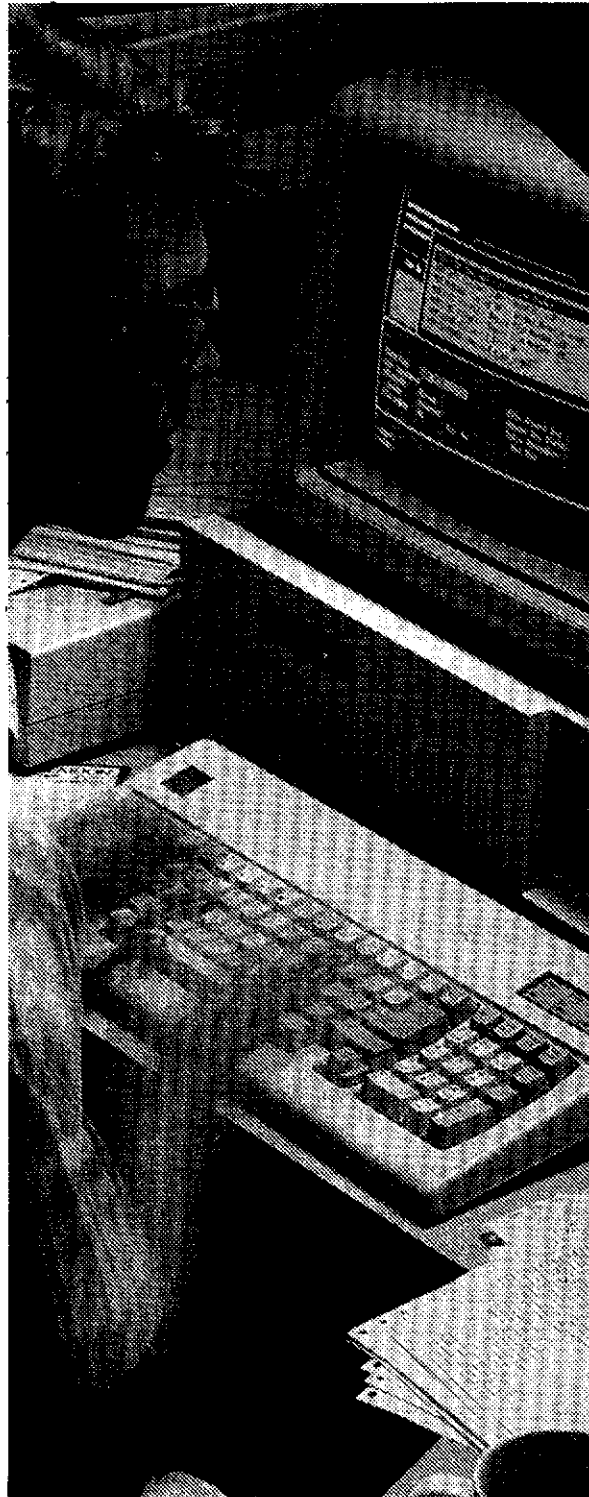
یکی از متغیرهای عمده در ذخیره و بازیابی اطلاعات فارسی، رسم الخط یا شیوه خط فارسی است. گرچه حروف و کلمات به عنوان ورودی و خروجی هر سیستم کامپیوتری در هر زبان از اهمیت برخوردار است، لیکن خط فارسی به دلیل ویژگی آن، در رویارویی با کامپیوتر دارای مسائل پیچیده تری است. این پیچیدگی حضوری تازه ندارد. اما چون در نظام دستی چندان دشواری آفرین نبوده، به صورت مسأله ای جدی عنوان نشده است، گرچه در مقاله هایی گاه به نکاتی در باب آن اشاره رفته است.

ورود کامپیوتر به عرصه های مختلف فعالیت های اجتماعی، فرهنگی، اقتصادی، و فنی سبب شده است که اطلاعات فارسی نیز خواسته یا ناخواسته با کامپیوتر درآمیزد. حضور ناگهانی کامپیوتر در گستره ای وسیع مجال آن را به صاحب نظران نداده است که راه حلی بنیانی برای مقابله با گرفتاریهای رسم الخط طراحی و عرضه کنند. اما امروز که تلاطم اولیه ورود کامپیوتر فروکش کرده و بدیع و تازه بودن خود را از دست داده است، آرامش و محالی پدید آمده و متخصصان حوزه هایی چون زبان شناسی، اطلاع رسانی، و کامپیوتر به این فکر افتاده اند که به مسأله رسم الخط فارسی به گونه ای جدی تر بپردازند.

راه حل هایی که تاکنون پیشنهاد شده بسیار متفاوت است: برخی بنیانی و بعضی مقطعی و موقتی است؛ اما هریک از آنها مسائلی خاص دارد که اشاره به آنها خالی از فایده نیست. پنج مورد یا تجربه مشخص را می توان نام برد که اختصاصاً به مسأله پیوند میان زبان فارسی و کامپیوتر مربوط می شود. این موارد بدین قرار است:

۱. هماهنگ کردن حروف

حروف فارسی دارای ویژگیهای متفاوتی است که سرجمع آنها به پیچیدگی کار کامپیوتری کردن اطلاعات فارسی می افزاید. این پیچیدگیها عبارتند از:



کلمات احتمال اینکه «دانش» و «زیست» از یک سو و «آموز» و «شناس» از سوی دیگر مستقل تلقی شوند بسیار است.

یکی از پیشنهاد‌های ضمنی برای پرهیز از عوارض این چندگانگی‌های حروف، به کار بردن شیوه‌ای یکسان در نوشتن حروف است، یعنی استفاده از حروف مستقل و بزرگ و کنار هم نهادن آنها برای تشکیل کلمه^۱. مثلاً به جای «ماست» بنویسیم «م است» و به جای «اندیشه» بنویسیم «ان دی شه». در این پیشنهاد دو نوع فاصله بیشتر نخواهیم داشت: یکی فاصله درونی — میان حروف یک کلمه — و دیگری فاصله بیرونی — میان هر کلمه با کلمه دیگر. در این شیوه احتمال خطا بسیار کاهش می‌یابد، اما دو نکته در آن قابل تأمل است:

یک. یکی اینکه «این پیشنهاد مورد قبول فارسی‌زبانان قرار نخواهد گرفت زیرا ما را از شیوه سنتی رسم الخط فارسی خیلی دور می‌کند و نگرانی‌هایی از این گونه را دامن می‌زند که 'پیوند ما با گذشته گسیخته می‌شود' یا 'خواندن متون کهن فارسی دشوار می‌شود' و مانند آن»^۲.

دو. دیگر اینکه مسأله کلمات مرکب کماکان به قوت خود باقی خواهد ماند. زیرا بر سر مرکب بودن یا نبودن بسیاری از کلمات هنوز میان زبان‌شناسان و ادیبان اختلاف رأی است. «دست به کار شدن» کلمه‌ای مرکب است یا چند کلمه مستقل، «به انجام رساندن» چه؟ پاسخ مثبت یا منفی به این پرسش در نوشتن شیوه پیشنهادی تأثیر خواهد داشت، یعنی ممکن است مثلاً «دست به کار شدن» یا «دست به کار شدن» نوشته شود.

۲. استفاده از تکواژها

«تکواژ اصطلاحی زبان‌شناسی است که به کوچک‌ترین جزء معنایی دار واژه اطلاق می‌شود. مثلاً واژه «زبان‌شناسی» از سه تکواژ «زبان+شناس+ی» و واژه «می‌توانم» از سه تکواژ «می+توان+م» تشکیل شده است»^۳. در این پیشنهاد تکواژها مستقل تلقی و نوشته می‌شوند و مستقلاً نیز توسط کامپیوتر خوانده می‌شوند. پیشنهاد تکمیلی برای بهبود این کار آن است که



● استفاده کنندگان از کامپیوتر عمدتاً افرادی با تخصص‌های گوناگون هستند که قصد ارتباط کلامی با یکدیگر دارند.

درازدت ناهماهنگی در کار پدید آید.
ب. به دلیل آنکه هر فقره اطلاعات هنگام بازیابی ناگزیر باید ابتدا از غربال سیاهه مورد نظر بگذرد، زمان کاوش اطلاعات افزایش یافته و کار بازیابی کند می‌شود.
اما این شیوه را می‌توان در نبود مرجعی واحد و موثق برای یکسان‌سازی رسم الخط جایگزینی مناسب تلقی کرد.

۴. پیوند ساختگی میان کلمات

در این شیوه، کلماتی که احتمال جدا یا پیوسته نوشتن اجزای آن می‌رود با تمهیدی با یکدیگر پیوند می‌یابند. به عنوان مثال، در کلمه «باستان‌شناس» میان «باستان» و «شناس» نشانه‌ای به کار می‌برند که در ذخیره آنها را به هم پیوند می‌دهد و در بازیابی نیز به هر ترتیب و با هر فاصله‌ای که وارد شود به صورت کلمه‌ای واحد خوانده خواهد شد. بنابراین تفاوتی میان «باستان‌شناس» منفصل و «باستان‌شناس» متصل نخواهد بود و جوینده با هریک از صورتها که وارد شود به جواب خواهد رسید. برنامه‌هایی فارسی (نظیر کوربیس^۶، نوسا، و جز آن) هم اکنون

تکواژها با فاصله‌ای تعریف شده نسبت به یکدیگر، و متفاوت با فاصله معمول میان کلمات، نوشته شوند. مثلاً به جای «من زبان‌شناسی نمی‌دانم» بنویسیم «من زبان‌شناسی نه می‌دانم»^۴. این پیشنهاد در مجموع مستلزم دو حرکت است: یکی تقطیع یا جدا کردن تکواژها از یکدیگر، دیگری بی‌فاصله نویسی^۵ تکواژهایی که یک کلمه را تشکیل می‌دهند.

علاوه بر وارد بودن دو اشکال پیشین که در پیشنهاد قبلی عنوان شد، در پیشنهاد اخیر اشکال دیگری نیز وجود دارد و آن اینکه تعیین تکواژها کار همه کس نیست، و تنها متخصصان یا پژوهندگان زبان‌شناسی بر این امر واقف اند، و این خود از جمله گرفتاریهای عظیم کار است.

۳. استفاده از سیاهه آماده

یکی از شیوه‌های هماهنگ‌سازی که بیش از همه در برخورد خط فارسی با کامپیوتر پذیرفته شده استفاده از سیاهه‌های از قبل تمبیه شده است. در این شیوه به جای هرگونه تغییری در رسم الخط فارسی، تمهیدی اندیشیده می‌شود که کامپیوتر بتواند رسم الخط‌های گوناگون را به شیوه‌ای واحد بخواند. یعنی خواه «می‌خواهم» بنویسیم یا «می‌خواهم»، «باستان‌شناس» بنویسیم یا «باستان‌شناس»؛ «اتاق» بنویسیم یا «اتاق»، پیوسته به یک شکل خوانده شود. دلیل اتخاذ این روش آن بوده است که طراحان چنین راه‌حلهایی خود را مجاز و صالح ندانسته‌اند که در خط فارسی دخل و تصرف نمایند یا پیشنهادی برای تغییر آن عنوان کنند، بلکه کوشیده‌اند از دل ناهماهنگی موجود نوعی هماهنگی برای مسائل خاص خود بیافرینند.

در این شیوه، برخلاف دو شیوه پیشین واحد سنجش و مطالعه کلمه است نه حرف یا تکواژ، و اساساً کلیه سیستم‌هایی که اصطلاح ویرایش کامپیوتری، غلط‌یاب و امثال آن را به کار می‌برند غرضشان استفاده از چنین روشی است.

تردیدی نیست که در این شیوه، ناگزیر باید سیاهه‌ای از قبل برای کلمات تعیین گردد و احتمالات گوناگون رسم الخط از طریق ارجاعات با یکدیگر مرتبط شود. این بدان معناست که برای هر مورد جدید باید سیاهه پیشین را مورد تجدید نظر قرار داد و این کار باید به صورت مستمر ادامه یابد. بنابراین دو دشواری در مورد این شیوه وجود دارد:

الف. سیاهه مورد اشاره هیچ‌گاه کامل نیست، بلکه صورتی جاری از کلمات است که با هر ورودی می‌بایست تکمیل گردد، و این خطر وجود دارد که برخی بی‌دقتی‌ها سبب شود که در



● پیشنهاد می‌شود که فرهنگستان زبان کمیته‌ای را اختصاصاً مأمور تدوین دستورالعملی برای شیوه خط فارسی کند.

موجود است که مبنای کار آنها این گونه پیوند میان اجزاء کلمه است.

اصل در این روش بر تعریف فاصله‌های درونی اجزاء کلمه است. اما اشکال عمده در این شیوه آن است که قبل از ورود اطلاعات به کامپیوتر می‌بایست متخصصی این‌گونه کلمات را شناسایی و با کد مربوط مجهز کند.

در این روش، امکان استفاده از اسکتر نیست، زیرا عملیاتی مقدماتی باید قبل از ورود صورت گیرد و از طریق صفحه کلید به سیستم تغذیه شود. اما به هر تقدیر، در حال حاضر، یکی از شیوه‌های مطلوب برای حل مسائل مقطعی برنامه‌های فارسی موجود است.

۵. هماهنگی رسم الخط

این پیشنهاد نیز ناظر بر کلمه است نه حرف یا تکواژ. در این پیشنهاد تأکید بر آن است که مرجعی قابل اطمینان رسم الخط واحدی را تصویب و عرضه کند، و اجرای آن نیز الزام‌آور باشد تا بتوان مرز و شیوه نگارش کلمات را تابع قاعده واحدی کرد. در حال حاضر مراکز مختلف نشر و پژوهش و تألیف، هریک بنا به تشخیص خود از رسم الخط خاصی پیروی می‌کند. گرچه وجوه اشتراک در آنها بسیار است، اما وجوه افتراق نیز اندک نیست. مراکزی نیز از هیچ اصل و قاعده‌ای پیروی نمی‌کنند. افرادی نیز اصرار بر حفظ رسم الخط شخصی خود دارند. بدیهی است کامپیوتر با چنین ملفه‌ای سازگار نیست. کامپیوتر با «اگر... پس» منطقی سروکار دارد. باید بتوان همه چیز را تعریف کرد، حتی استثناءها نیز باید تعریف پذیر باشند. بنابراین لازم است شیوه‌ای واحد به کلیه دست‌اندرکاران تألیف و نشر ابلاغ شود. امر هماهنگ‌سازی می‌بایست به موارد زیر توجهی خاص روا دارد:

الف. چسبیده یا جدا نوشتن کلمات را تعریف کند و تکلیف چگونگی نگارش پیشوندها، میانوندها، و پسوندها را روشن سازد (مثلاً می‌کنم یا میکنم، خودبه‌خود یا خودبخود، واقف‌اند یا واقفند؟)

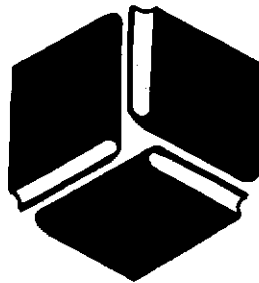
ب. جمع‌های فارسی یا رایج در فارسی تعریف پذیر یا یکدست شود (مثلاً مردها یا مردان، کارخانه‌ها یا کارخانجات، جواهرها یا جواهرات، مدرسه‌ها یا مدارس؟)
ج. اختلاف املاءها (مانند اتاق یا اطاق، اثریش یا اطیش، دائریا دایر؟) از میان برود.

نتیجه و پیشنهاد

از میان راه‌های مختلفی که تاکنون عرضه یا تجربه شده است، شاید مناسبترین آنها راه پنجم یعنی سیاست‌گذاری واحد برای یکسان‌سازی رسم الخط باشد. دلیل برتری آن نسبت به راه‌ها و شیوه‌های دیگر این است که در عین پایه‌ای بودن کار، مسائل مورد ابتلای خط فارسی عالمانه‌تر حل می‌شود. با توجه به پنج شیوه‌ای که بحث مربوط به آنها به اجمال گذشت، می‌توان دریافت که تقطیع عناصر زبانشناختی در آنها کم و بیش متفاوت است. در نوع اول - یعنی هماهنگ کردن حروف - حرف به عنوان عنصر اصلی مبنای قرار گرفته است و در نوع دوم تکواژ. اما در نوع سوم و چهارم و پنجم، مبنای تقطیع

صورت پذیرد، و هرچه دیرتر آغاز شود، متونی که بر مجموعه موجود افزوده خواهند شد بیشتر بوده و کار هماهنگ کردن بعدی آنها مستلزم هزینه بیشتری خواهد بود.

بدیهی است حوزه‌های مختلفی از نتایج این امر بهره‌مند خواهند شد که از جمله مراکز اطلاع‌رسانی، سازمانهای نشر، نظامهای مدیریت، و مراکز فنی و صنعتی را می‌توان نام برد. تا چنین حرکتی صورت نگیرد، ایجاد پایگاههای زبان فارسی یا تهیه واژه‌نامه‌های کامپیوتری فارسی، و طراحی و اجرای نظامهای هماهنگ اطلاع‌رسانی از توفیق چندانی برخوردار نخواهند بود، زیرا کلیه داده‌های نظام دستی با تمام ناهماهنگی‌هایش وارد کامپیوتر خواهد شد و طبقاً اطلاعات نیز به گونه‌ای هماهنگ بازیابی نخواهد گردید.



نوشته‌ها:

۵. این مقاله در سمینار «اطلاع‌رسانی برای تحقیق و توسعه» (۱۸-۱۹ اردیبهشت ۷۲) ارائه شده است.
۱. محمدرضا باطنی، «نگاهی تازه به شیوه خط فارسی»، آدینه (آبان ۷۱)، ص ۴۴-۴۵.
۲. همان، ص ۴۴.
۳. ایرج کابلی، آدینه، شماره ۷۲، به نقل از محمدرضا باطنی، پیشین، ص ۴۴.
۴. همان.
۵. کریم امامی، «لغزوم بازیگری در شیوه نگارش خط فارسی»، آدینه (شهریور ۱۳۷۱)، ص ۱۸-۱۹.
۶. کوریس ۱، سیستم فهرست‌نویسی کامپیوتری کتاب است که در دفتر پژوهشهای فرهنگی زیر نظر نگارنده و با همکاری تعدادی از متخصصان کتابدلی و اطلاع‌رسانی، و نیز کامپیوتر طراحی و اجرا گردید. پیوند ساختگی میان کلمات در سال ۱۳۶۵ در این سیستم به کار گرفته شد و در ۱۳۶۹ عرضه عمومی یافت.

کلمه است. اگر بپذیریم که این تدابیر برای یافتن راهی روشن به منظور برقراری ارتباط انسان - ماشین - انسان است، استفاده کنندگان از کامپیوتر عمدتاً افرادی با تخصص‌های گوناگون هستند که قصد ارتباط کلامی با یکدیگر دارند و لذا کوچکترین واحد معنی‌دار برای آنها کلمه است نه حرف و نه تکرار. این امر خصوصاً در حوزه اطلاع‌رسانی بسیار حائز اهمیت است. به همین دلیل، انتخاب یکی از شیوه‌های سوم، چهارم، یا پنجم از لحاظ عملی مفید فایده بیشتری است. ورود اطلاعات به سیستم نیز چه با استفاده از اسکنر (تقطیع گر) باشد یا از طریق اپراتور، باید به چنان شیوه‌ای صورت گیرد که هنگام بازیابی نیز همان عناصر به دست آید یا با یکدیگر ترکیب منطقی معنی‌دار بیابند.

از میان سه شیوه آخر، شیوه سوم یعنی تهیه سیاهه آماده نه تنها هیچ‌گاه پایان‌یافته نیست، بلکه پیش از هرگونه ورودی ابتدا باید توسط نیروی انسانی مطالعه شود تا موارد جدید استخراج و به سیاهه افزوده گردد؛ و این کار نقض غرض کامپیوتری کردن است که از جمله ویژگیهای آن تقطیع دلخواه اجزاء نوشته با استفاده از کامپیوتر است. همین گرفتاری در مورد شیوه چهارم، یعنی پیوند ساختگی میان کلمات نیز وجود دارد.

شیوه پنجم، یعنی هماهنگ کردن رسم الخط فارسی، از سایر شیوه‌ها معقولتر به نظر می‌رسد، زیرا نه تنها به دلیل ضرورت برخورد با کامپیوتر، خط فارسی شیوه‌ای واحد خواهد یافت، بلکه آشنایی و چندگونگی فعلی رسم الخط نیز از میان خواهد رفت. برای تحقق چنین امری پیشنهاد می‌شود که فرهنگستان زبان کمیته‌ای را اختصاصاً مأمور تدوین دستورالعملی برای شیوه خط فارسی کند. تخصص‌های چنین کمیته‌ای منطقاً عبارت خواهد بود از زبان‌شناسی، ادبیات فارسی، ویراستاری، اطلاع‌رسانی، و کامپیوتر. این کمیته نخست به گردآوری و مطالعه کلیه سوابق مربوط به رسم الخط فارسی می‌پردازد و سپس با توجه به مسائلی از قبیل تداول، منطق خط فارسی، و مانند آن، شیوه‌ای را برمی‌گزیند و آن را مدون می‌کند؛ و پس از کسب نظر تعدادی از متخصصان و محققان در خارج از بافت کمیته و آزمونی مقدماتی با کامپیوتر، آن را به عنوان دستورالعملی لازم‌الاجرا به کلیه مراکز و سازمانهای آموزشی، پژوهشی، انتشاراتی، و جز آن ابلاغ می‌کند.

اجرای چنین طرحی گرچه ممکن است هزینه و زمانی درخور بطلبد، لیکن این هماهنگی یک بار و برای همیشه باید