**در دشواریهای رایانه ای زبان فارسی**

**نویسنده : طباطبایی، علاءالدین**

در شماره 102 فصلنامه نشر دانش مقاله‏ای به چاپ رسیده بود با عنوان «دشواریهای پردازش رایانه‏ای خط فارسی» به قلم دکتر محرم اسلامی. ایشان اشکالاتی را که خط فارسی در پردازش رایانه‏ای ایجاد می‏کند در نهایت سادگی و دقت برشمرده‏اند و برای رفع آنها پیشنهادهای ارزنده‏ای ارائه کرده‏اند. آقای اسلامی آورده‏اند: «دشواریهای پردازش نوشتار فارسی را به دو گروه عمده زیر تقسیم می‏کنیم ... : 1) دشواریهای برخاسته از ماهیت زبان فارسی؛ 2) دشواریهای برخاسته از کاستیهای نظام نوشتاری زبان فارسی».(1) ایشان به چند ویژگی زبان فارسی اشاره کرده‏اند و مدعی شده‏اند که این ویژگیها پردازش رایانه‏ای آن را دشوار می‏کند.

مقاله ایشان تا آنجا که به طرح و شرح این دشواریها می‏پردازد روشنگر و جالب‏توجه است، اما از فحوای کلام ایشان چنین برمی‏آید (یا دست کم بنده چنین برداشت می‏کنم) که چنین دشواریهایی برای زبانهای دیگر، مثلاً انگلیسی و آلمانی، وجود ندارد. ما در این مقاله متعرض آنچه ایشان درباره کاستیهای خط فارسی گفته‏اند نمی‏شویم و فقط به نکاتی می‏پردازیم که به ماهیت زبان فارسی مربوط می‏شوند.

آقای اسلامی گفته‏اند: «... اگر پایانه‏های تصریفیِ نماینده فاعل و مفعول و متمم‏ها در زبان فارسی (مانند برخی زبانهای دیگر) وجود می‏داشت تشخیص مرز و نقش گروهها آسانتر می‏بود».(2) ایشان در ادامه گفته‏اند که جمله «این کارخانه را خراب کرد» به صورتهای زیر قابل تقطیع است:

«الف) ـ ø ـ این کارخانه را ـ خراب کرد.

ب) این ـ کارخانه را ـ خراب کرد.

ج) این کار ـ خانه را ـ خراب کرد».(3)

منظور آقای اسلامی این است که فاعل جمله را به سه صورت می‏توان تعبیر کرد: 1) ضمیر او` که مستتر است؛ 2) ضمیر این`؛ 3) این‏کار`. به بیان دیگر، از این زنجیره کلمات، بسته به نوع تقطیع، سه معنی متفاوت برداشت می‏شود:

1) اواین کارخانه را خراب کرد.

2) این (مسئله، عامل، آدم ...) کارخانه را خراب کرد.

3) این‏کارخانه ر خراب کرد.

به نظر می‏رسد که بخشی از این اشکال به خط فارسی مربوط می‏شود و اگر این جمله را با پیروی از رسم‏الخطهای جدید بنویسیم و بین اجزای سازنده واژه‏های مرکب، مانند کارخانه` ، فاصله نگذاریم بخشی از اشکال حل می‏شود. فرض می‏کنیم جمله بالا را به صورت زیر بنویسیم:

این کارخانه رخراب کرد.

این جمله را فقط به دو صورت می‏توانیم تقطیع کنیم:

الف) ø ـ این کارخانه را ـ خراب کرد.

ب) این ـ کارخانه راـ خراب کرد.

در جمله (الف) ضمیر او` که مستتر است فاعل به شمار می‏آید و معنی جمله چنین است: «او این کارخانه را خراب کرد». این تعبیر به دلایل زیر تعبیر غالب محسوب می‏شود و بعید است که فارسی‏زبانان در یک نوشته واژه این` را فاعل به شمار آورند. علت آن هم این است که کاربرد این` و آن` به تنهایی در مقام فاعل محدودیت زیادی دارد و اساسا در گفتار به کار می‏رود، آنهم در وضعیتی که مخاطب باتوجه به موقعیت تشخیص می‏دهد که منظور از این` کیست یا چیست. مثلاً ممکن است مادری با اشاره به یکی از فرزندانش بگوید: «این سیب را خورد» و منظورش این باشد که «این بچه سیب را خورد». ضمیر این` در یکی دو مورد، هم در گفتار و هم در نوشتار، در مقام فاعل جمله به کار می‏رود. اما در این موارد نشانه‏های صوریی وجود دارد که نشان می‏دهد این` فاعل است. مثلاً وقتی بعد از این` واژه‏های که` یا هم` قرار گیرد این` فاعل است: «این که گفتن ندارد» یا «این که هیچ» یا «این هم علی». در چنین وضعیتی به توصیه ویراستاران این‏که` و این‏هم` باید جدا نوشته شوند. مورد دیگری که این` نقش فاعل جمله را ایفا می‏کند زمانی است که بلافاصله بعد از آن فعل می‏آید: «این باعث می‏شود که دانشجویان به درس علاقه‏مند شوند». در این صورت هم وجود فعل بلافاصله بعد از این` نشان می‏دهد که این` فاعل است.

چنانکه می‏بینیم کاربرد این` در مقام فاعل و بدون هیچ نشانه‏ای که فاعل‏بودن آن را نشان دهد با محدودیت بسیاری روبه‏روست. در تمام مقاله‏ای که آقای اسلامی نوشته‏اند، بجز در

حاشیه:

1) نشر دانش، پاییز 1381، ص 29.

2) همان، ص 29.

3) همان، ص 29.

مثالها که ساخته خود ایشان هستند، و بجز در یکی دو مورد دیگر ضمیر این` به تنهایی در مقام فاعل به کار نرفته است و در آن موارد نیز هیچ ابهامی در جمله پدید نیامده است، چون مفعول جمله با نقش‏نمای را` مشخص شده و معلوم است که فعل با ضمیر این `مطابقت یافته است: «و این به نوبه خود پردازش نوشتار را دچار مشکل می‏کند».(4)

بنابراین در متون واقعی فارسی، اگر براساس رسم‏الخطهای جدید نوشته شده باشند، از آن نوع ابهامهایی که آقای اسلامی مطرح کرده‏اند بسیار به ندرت پیش می‏آید.

آقای اسلامی همچنین آورده‏اند: «در زبان فارسی تشخیص مقوله واژگانی کلمه‏ها در درون جمله به دلیل فقدان ابزارهای صرفی لازم دشوار است». ایشان سپس گفته‏اند که در جمله «زنِ نویسنده گفت ...» ساخت زن نویسنده` مبهم است، «زیرا معلوم نیست که این عبارت ترکیب اضافی است یا وصفی ... اگر در زبان فارسی نیز ابزارهای تصریفی وجود می‏داشت می‏شد اسم یا صفت بودن نویسنده را در مثالهای بالا با استفاده از آن ابزارها دریافت».(5)

ابهامی که آقای اسلامی در عبارت زن نویسنده` دیده‏اند به معنی زن` مربوط می‏شود. زن در فارسی دو معنی دارد، یکی همسر` و دیگری جنس زن. ابهامی که در این عبارت وجود دارد همان است که زبان‏شناسان به ابهام واژگانی تعبیر می‏کنند و خاص زبان فارسی نیست و در همه زبانها وجود دارد. برای مثال در جمله زیر که از زبان انگلیسی برگرفته شده است از آنجا که واژه `list دو معنای متفاوت دارد، دو معنی (A و B) از جمله برداشت می‏شود:

The captain corrected the list.

A) The captain corrected the enventory.

B) The captain corrected the tilt.

آقای اسلامی همچنین آورده‏اند: «زبان فارسی فاقد صورت سازهای نحوی مناسب جهت بازنمایی عناصر موجود در یک گروه نحوی است. این امر باعث دشواری و ابهام در تشخیص آغاز و پایان گروه نحوی می‏شود ... اگر زبان فارسی مانند زبان انگلیسی صورت سازهایی مانند `of و `s داشت در آن صورت تعیین گروههای نحوی، مثلاً در عبارت «معلم مدرسه» به آسانی انجام می‏شد».(6)

بر این سخن آقای اسلامی دست کم دو ایراد وارد است: یکی اینکه زبان فارسی نشانه ترکیب نحوی دارد، اما این نشانه (کسره اضافه) در بسیاری موارد در نوشتار ظاهر نمی‏شود. بنابراین ایراد ایشان به خط فارسی مربوط است نه زبان فارسی. دوم اینکه درست است که در زبان انگلیسی "of" و "s" نشانه‏های گروه نحوی هستند اما این عناصر همیشه به کار نمی‏روند. برای مثال در بسیاری موارد به جای اینکه دو اسم با `of به هم اضافه شوند و به‏اصطلاح یک `of phrase تشکیل دهند، اسمی که باید بعد از `of قرار گیرد قبل از اسم دیگر قرار می‏گیرد و of حذف می‏شود:

encyclopedia of linguistics û linguistics encyclopedia

disorders of hearing û hearing disorders

diseases of heart û heart diseases

علاوه بر این، در بسیاری موارد یک اسم قبل از اسم دیگر می‏آید و معنیِ صفت می‏دهد:

oil paperکاغذ روغنی

war filmفیلم جنگی

paper aeroplaneهواپیمای کاغذی

چنانکه می‏بینیم اسمهای oil و war و paper بدون هیچ نشانه صوری نقش صفت را ایفا می‏کنند. بنابراین، زبان انگلیسی هم در بسیاری موارد فاقد، به اصطلاحِ آقای اسلامی، صورت‏ساز نحوی است.

اشکال دیگری که آقای اسلامی مطرح ساخته‏اند اختیاری بودن فاعل در زبان فارسی است: «جمله در زبان فارسی دارای دو فاعل است که یکی از آنها اختیاری و دیگری اجباری است. اولین فاعل که یک گروه اسمی کامل است فاعل اختیاری جمله به‏شمار می‏رود و دومین فاعل که با شناسه‏های فعلی نمایش داده می‏شود فاعل اجباری جمله محسوب می‏شود».(7)

درست است که در زبان فارسی اگر فاعل جمله ضمیر باشد اختیاری است، اما فقدان ضمیر فاعلی باعث ابهام نمی‏شود زیرا در این زبان فعل صرف می‏شود و از شناسه فعل می‏توانیم به شخص` پی ببریم. این ویژگی هم به هیچ وجه خاص زبان فارسی نیست و در زبانهای بسیاری، از جمله ایتالیایی و اسپانیایی و لاتین، نیز وجود دارد و نشان می‏دهد که اصل اقتصاد بر همه ابعاد زبان حاکم است: اگر جمله کوتاه رفتم` همان معنی من رفتم` را می‏دهد، زبان به ما اختیار می‏دهد که من` را نیاوریم مگر اینکه بخواهیم بر ضمیر من `تأکید کنیم.

از آنجا که بررسی همه نکاتی که آقای اسلامی در ذیل

حاشیه:

4) همان، ص 30. 5) همان، ص 29.

5) همان، ص 30.

6) همان، ص 30.

نشر دانش » شماره 106 (صفحه 38)

«دشواریهای برخاسته از ماهیت زبان فارسی» آورده‏اند در حوصله این مقاله نیست به همین مختصر اکتفا می‏کنیم و مسئله را از دیدگاهی کلی‏تر مطرح می‏سازیم.

بحث را با یک سؤال آغاز می‏کنیم: آیا می‏توان گفت که پردازش رایانه‏ای برخی زبانها دشوارتر یا آسانتر از زبانهای دیگر است؟ (در پاسخ به این سؤال باید توجه داشته باشیم که خط و زبان را با هم خلط نکنیم).

به این سؤال فقط زمانی می‏توان جواب قطعی و متکی بر شواهد تجربی داد که پردازش رایانه‏ای دهها زبان از ابعاد مختلف مورد مقایسه قرار گیرد. اما این سؤال را در سطحی نظری‏تر نیز می‏توان مطرح ساخت: آیا می‏توانیم بگوییم فلان زبان از زبان دیگر ساده‏تر یا دشوارتر است؟ در اینجا نیز باید توجه داشته باشیم که زبان به ذات خود مدّ نظر است و نباید مثلاً ادبیات زبانها را در این بحث دخالت داد.

چنانکه می‏دانیم غالب زبان‏شناسان، و بویژه پیروان مکتب گشتاری ـ زایشی، اعتقاد دارند که زبان ساده یا بدوی وجود ندارد و «دستور تمام زبانها در مجموع به یک اندازه پیچیده است».(8) به همین دلیل است که در همه جوامع زبانی، کودکان تا قبل از سن ده‏سالگی بر زبان مادری خود تسلط می‏یابند، به این معنی که ساختمان دستوری (شامل دستگاههای واجی و صرفی و نحوی) و واژگان پایه آن را می‏آموزند و بی‏هیچ مشکلی با اطرافیان خود ارتباط برقرار می‏کنند. علاوه بر این میزان رشد زبانی آنها در سنین مختلف تقریبا برابر است. البته دستور هر زبان را باید در مجموع در نظر گرفت و فریب سادگی یا پیچیدگی این یا آن جزء را نخورد. سخنگویان به‏هر زبان با استفاده از دانش زبانی خود، که بخش اعظم آن ناآگاهانه است، معنی جملات را درمی‏یابند. اگر قرار باشد ماشینی بسازند که از عهده پردازش زبان برآید طبیعتا باید این دانش را به نحوی به ماشین منتقل کنند و این دانش به اعتقاد غالب زبان‏شناسان در همه زبانها در مجموع به یک اندازه پیچیده است.

آنچه در بالا آوردیم به این معنی نیست که تفاوت میان زبانها اندک و بی‏اهمیت است. مجموع ویژگیهای هر زبان چنان است که به آن هویتی یگانه می‏بخشد. مثلاً برخی از ویژگیهای فارسی از این قرار است:

1) آرایش سازه‏های آن آزاد است.

2) فاعل اگر ضمیر باشد اختیاری است.

3) فاقد حرف تعریف معین است.

4) در ترکیبهای نحوی اول هسته و بعد وابسته قرار می‏گیرد.

5) تفاوت جنسی را با عناصر دستوری نشان نمی‏دهد.

زبان فارسی در ویژگی اول به زبان روسی، در ویژگی دوم به اسپانیایی و ایتالیایی و لاتین، در ویژگی سوم به روسی، در ویژگی چهارم به فرانسه و عربی، و در ویژگی پنجم به انگلیسی شباهت دارد. اما تلفیق این ویژگیها و دهها ویژگی دیگر چنان است که از یک سو به آن هویتی منحصر به فرد می‏بخشد و از سوی دیگر اجزای گوناگونش را در چنان توازنی قرار می‏دهد که اصل اقتصاد در همه سطوحش جاری باشد. و البته وضع همه زبانها چنین است.

به گمان این نگارنده مشکلاتی که ما در پردازش رایانه‏ای زبان فارسی داریم نه از ماهیت این زبان، که از فقر علمی و پژوهشی ما سرچشمه می‏گیرد. اگر ما دست کم یک دستور توصیفی جامع داشتیم و اگر بنیه علمی‏مان در زمینه‏های مختلف این چنین ضعیف نبود بی‏تردید وضعمان در این عرصه نیز فرق می‏کرد.

فراموش نکنیم که ما هنوز نتوانسته‏ایم دستگاه «املانویس فارسی» بسازیم، حال آنکه برای زبان انگلیسی املانویسهای نسبتا کارآمدی ساخته‏اند.(9) چنین دستگاهی باید از عهده تشخیص آواهای زبان برآید. با توجه به اینکه فارسی فقط شش مصوت دارد و ساختمان هجایی آن در قیاس با زبان انگلیسی ساده‏تر است آیا باز هم می‏توان عقب‏ماندگی در این زمینه را به ماهیت زبان فارسی نسبت داد؟

زبان‏شناسی رایانه‏ای در ایران، و نیز در جهان، هنوز در آغاز راه است. امید است متخصصان این حوزه گرفتار قضاوتهای عجولانه و مبتنی بر شواهد اندک نشوند. در تاریخ زبان‏شناسی چنین قضاوتهایی کم نبوده است. نمونه بسیار معروف آن نظر اغراق‏آمیز «ورف» درباره ویژگیهای افعال در زبان هویی است؛(10) و نمونه دیگر آن قضاوت پیروان مکتب گشتاری ـ زایشی درباره نظریه‏ای است که در کتاب جنبه‏های نظریه نحو طرح و شرح شده بود. پیروان این مکتب نظریه مزبور را شتابزده نظریه استانده(11)` نامیدند حال آنکه دیری نپایید که معلوم شد در این نظریه نارساییهای مهمی وجود دارد.

در پایان برای آقای محرم اسلامی و همکارانشان که برای پردازش رایانه‏ای زبان فارسی کوشش بسیار می‏کنند آرزوی موفقیت می‏کنیم.

حاشیه:

8) ویلیام اُگرادی و دیگران (1989)، درآمدی بر زبان‏شناسی معاصر، ترجمه دکتر علی درزی (1380)، تهران، سمت، جلد اول، ص 15.

9) نک. لوح فشرده Encyclopedia Britanica 2003 در ذیل speech recognition .

10) نک.

Linguistics: the Cambridge Survey, vol. IV, Cambridge University Press, 1990, p. 16-17.

11) Standard theory.