

فصلنامه علمی پژوهشی کاوش‌نامه

سال یازدهم (۱۳۸۹)، شماره ۲۱

تشخیص سبک فردی نویسندگان بر اساس شیوه تحلیلی کیوسام و بررسی کارآیی آن در زبان فارسی*

امیر محمدیان**

دانشجوی دکتری زبان شناسی، دانشگاه فردوسی مشهد

چکیده

تشخیص مؤلف، یکی از شاخه‌های زبان شناسی حقوقی است که در آن سعی می‌شود تا نویسنده متن (از متون ادبی گرفته تا حقوقی)، با توجه به ملاک‌هایی که هر شیوه ارائه می‌کند مشخص شود. از آنجا که در زبان فارسی بیشتر تلاش‌های صورت پذیرفته در این زمینه، در حوزه متون ادبی و بر پایه ملاک‌های سبکی و سلیقه‌ای بوده‌اند و از روش‌های مطمئن تر برای این منظور استفاده نشده است، عدم قطعیت و اختلاف نظر در مورد نتایج حاصل از آن دور از انتظار نیست. بر این اساس، مقاله حاضر پس از مرور اجمالی پیشینه مطالعات انجام گرفته در زمینه تشخیص مؤلف، به معرفی یکی از رویکردهای مطرح آماری موسوم به تحلیل کیوسام پرداخته و با اعمال آن بر پاره‌ای از متون نثر فارسی معاصر، کارآیی شیوه مذکور را در این زبان بررسی می‌کند.

کلیدواژه‌ها: زبان شناسی حقوقی، تشخیص مؤلف، تحلیل کیوسام، جمع تصاعدی.

تاریخ پذیرش نهایی: ۸۹/۴/۲۴

* تاریخ دریافت مقاله: ۸۸/۷/۲۰

نشانی پست الکترونیکی نویسنده: am_mohammadian@yahoo.com

مقدمه

زبان، به عنوان مهمترین و پیشرفته‌ترین ابزار ارتباطی بشر، از دیر باز مورد توجه اندیشمندان بوده است. بررسی و مطالعه علمی زبان، زبان‌شناسی (Linguistics) نامیده می‌شود که خود، جنبه‌های گوناگونی دارد و علاقه زبان‌شناس ممکن است به مباحث مختلفی نظیر توصیف دستور زبان، نقش زبان در اجتماع، چگونگی فراگیری آن و ... معطوف شود. خوشبختانه در دهه‌های اخیر، مطالعات زبان‌شناختی به طرز چشمگیری پیشرفت کرده و به تدریج از سمت مباحث نظری و بیان مفاهیم عام و انتزاعی، به سوی مسایل کاربردی و میان رشته‌ای هدایت شده است. مطالعات مربوط به تعامل زبان و مسایل حقوقی نیز زاینده این تغییر رویکرد در بررسی‌های زبان‌شناختی است که امروزه تحت عنوان زبان‌شناسی حقوقی (Forensic Linguistics) شناخته می‌شوند.

اصطلاح زبان‌شناسی حقوقی نخستین بار توسط یان سوارتویک (Jan Svartvik) در سال ۱۹۶۸، هنگامی که به تحلیل اعتراف نامه‌های شخصی به نام تیموتی جان ایوانز (Timothy John Evanz)، که به خاطر قتل همسر و فرزندش به دار مجازات آویخته شده بود می‌پرداخت، پدید آمد (Olsson, 2004:3). بعد ها، رفته رفته این عقیده در میان برخی از صاحب نظران قوت گرفت که می‌توان از علم زبان‌شناسی در حلّ پاره‌ای از مسایل حقوقی که به نوعی با زبان سر و کار دارند، بهره جست. از آن زمان، به تدریج، زبان‌شناسی حقوقی چه از نظر تعداد افرادی که در این حوزه فعالیت دارند و چه از لحاظ شمار زیر شاخه‌هایی که در بطن آن شکل گرفته و بالیده‌اند، رشد چشمگیری داشته است. تا جایی که در این مورد بیان شده است که: "... امروزه مواردی نظیر تشخیص صدا، عدم توازن قدرت در دادگاه‌ها، ارتباط وکیل و موکل، ماهیت شهادت دروغ، تقلب در نام‌های تجاری، ترجمه مکالمات در دادگاهها (در مورد اتباع خارجی)، تحلیل مکالمات ضبط شده، ... و نیز تشخیص نویسنده مدارک مکتوب و بررسی درستی یا نادرستی متون منسوب (مثلاً نامه‌های تهدید آمیز)، از جمله مسایل مورد توجه زبان‌شناسان حقوقی است ... " (شای، ۱۳۸۳، ص ۳۶۵).

مبحث تشخیص مؤلف، یکی از زیر شاخه‌های زبان شناسی حقوقی به شمار می‌رود و هدف اصلی آن، تشخیص و تعیین هویت نگارنده متن، و نیز سنجش درستی یا نادرستی متون منسوب به نویسنده‌ای خاص است. اساس کار در حوزه تشخیص مؤلف بر این فرض استوار است که هر نویسنده به هنگام کاربرد زبان، عاداتی منحصر به فردی دارد که این عادات، به شکل الگوهای قابل تشخیصی در متن آشکار می‌شوند. در این حوزه سعی بر آن است تا با توجه به ویژگی‌های متن و نیز با استفاده از ملاک‌های خاصی که هر شیوه ارائه می‌کند، ضمن تطبیق نتایج حاصل از متن مورد تحلیل با متون تحلیل شده قبلی، انتساب متن را به شخصی خاص مورد ارزیابی قرار داده و یا هویت نویسنده متنی نامعلوم را معین کرد. اگرچه، به نظر می‌رسد پیشینه تلاش‌های صورت گرفته در این رشته، بسیار پیش تر از پیدایش زبان شناسی حقوقی باشد، با این حال، این حوزه هنوز موفق به تثبیت جایگاه خود به عنوان یکی از زیر شاخه‌های زبان شناسی نشده است. چرا که شیوه‌های طرح شده در زمینه تشخیص مؤلف، یا بر پایه آرا و سلیقه فردی و بدون در نظر گرفتن ویژگی‌های عینی و زبان شناختی موجود در متن قرار گرفته‌اند و یا در مورد میزان کارایی شیوه‌های دیگری که عوامل زبانی - آماری متن را در نظر می‌گیرند، مبالغه شده و طراحان این شیوه‌ها مدعی هستند که رویکرد آنها بر هر نوع متن و در هر زبان قابل اجراست که البته اثبات این ادعا، نیازمند بررسی و تحلیل عملی است.

در زبان فارسی، عمده تلاش‌های انجام شده در زمینه تشخیص مؤلف، در حوزه متون ادبی و غالباً بر پایه رده بندی‌های سبکی و جهان بینی نویسنده بوده است. از آنجا که دو معیار مزبور، کلیدهای زبانی را نادیده گرفته و بیشتر بر پایه آرا و احتمالات شخصی محقق استوارند، احتمال وجود اختلاف نظر در مورد نتایج حاصل شده و متون منسوب دور از انتظار نیست و همان گونه که در عمل مشاهده می‌شود، انتخاب گران متون ادبی فارسی، نسخه‌های متفاوتی از متون منتسب به نویسنده‌ای خاص را ارائه می‌کنند. بدیهی است در صورتی که کارایی شیوه‌هایی که از متغیرهای عینی و

زبانی- آماری درون متنی بهره می‌جویند و عادت‌ها و رفتارهای زبانی نویسنده را در نظر می‌گیرند در زبان فارسی اثبات شود، می‌توان به راهکاری قابل اعتمادتر در این زمینه دست یافت، از شیوه‌های مطمئن تری برای تشخیص مؤلف متون ادبی استفاده کرد و حتی می‌توان در تحلیل متون حقوقی نیز از آنها کمک گرفت.

مقاله حاضر با در نظر گرفتن مسایل ذکر شده، سعی دارد پس از مروری کوتاه بر پیشینه تشخیص مؤلف، با انتخاب نمونه‌هایی از متون فارسی، کارآیی یکی از شیوه‌های مطرح در این زمینه را که بر پایه محاسبات آماری است در زبان فارسی بررسی کند. بر این اساس، شیوه تحلیل کیوسام (abbreviated form of cumulative Q-sum (Cusum))، به این دلیل که طراحان آن معتقد به کارآیی شیوه مذکور در هر زبان و بر هر نوع متن هستند، به عنوان شیوه انتخابی مقاله، در ادامه مورد سنجش قرار خواهد گرفت.

پیشینه

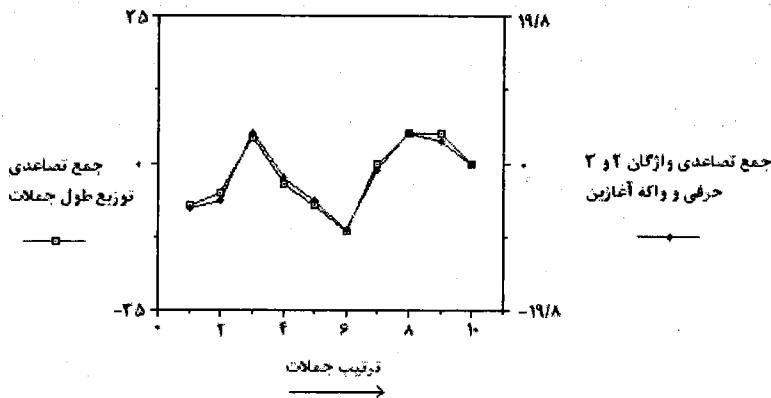
به نظر می‌رسد نخستین تلاش‌های مربوط به بحث تشخیص مؤلف نهایتاً به قرن هجدهم باز گردد. چون تا قبل از این تاریخ، به دلیل محدودیت صنعت چاپ، آثار مکتوب در دسترس همگان نبود و دیگر این که نظریه فردگرایی دکارت (Descartes) و این نگرش که هر فرد می‌تواند دارای عقاید شخصی و افکار منحصر به فرد باشد، هنوز مورد پذیرش عموم واقع نشده بود. تردید درباره گردآورنده متون کتاب مقدس، یکی از نخستین موارد مرتبط با حوزه تشخیص مؤلف است که در سال ۱۷۱۱ توسط کشیشی آلمانی به نام ویتتر (H. B. Witter) مطرح شده است. او معتقد بود که نام‌های متفاوتی که در نسخه‌های مختلف تورات برای خدا منظور گردیده است، احتمالاً نشان‌دهنده این مطلب است که افراد مختلفی این نسخه‌ها را گردآوری کرده‌اند (Olsson, 2004:11). در سال ۱۷۸۵، کشیشی به نام جیمز ویلموت (James Wilmot) مدعی شد که نویسنده واقعی نمایشنامه‌های منسوب به شکسپیر، در واقع، بیکن بوده است. بیست سال بعد، جیمز کورتون کوول (James Corton Cowell)، مجدداً همین نظر را اعلام کرد. با این

حال به نظر می‌رسد که آن دو برای ادعای خود شاهد علمی نداشته و بر پایه حدسیات، چنین نظری را مطرح کرده‌اند. در مجموع، می‌توان گفت که تا آن زمان، هنوز مسایل و ادعا های مربوط به بحث تشخیص مؤلف، رنگ و بوی علمی به خود نگرفته بوده است. از نیمه دوم قرن نوزدهم به بعد، شیوه‌های مختلفی برای تشخیص نویسنده متن ارائه شده که می‌توان آنها را در دو گروه جای داد: شیوه‌های مبتنی بر ملاحظات سبکی و سلیقه‌ای و نیز شیوه‌هایی که بر اساس معیارهایی عینی و قابل اندازه‌گیری به تشخیص مؤلف متن و ارزیابی متون منسوب می‌پردازند. مک منامین (Mc Menamin) و هنلاین (Hänlein) از جمله افرادی هستند که ملاحظات سبکی را در تشخیص مؤلف دخیل دانسته‌اند. به علاوه افرادی چون بهار و دیگر گزینشگران متون ادبی فارسی، تحقیقات خود را بر اساس ملاک‌های سبکی پیش برده‌اند. از طرف دیگر، روش‌هایی که بر پایه کلیدهای عینی به تشخیص مؤلف متن می‌پردازند، از معیارهایی درون متنی مانند میانگین طول واژه و جمله و... استفاده می‌کنند. شیوه تحلیل کیوسام یکی از روش‌هایی است که با استفاده از چند معیار آماری، امروزه به عنوان یکی از شیوه‌های مطرح در زمینه تشخیص مؤلف قلمداد شده و طراحان آن معتقد به کارایی آن در هر زبان و بر هر نوع متن هستند. به همین دلیل در ادامه به معرفی روال کار در شیوه تحلیلی کیوسام پرداخته و سپس با اعمال آن بر تعدادی از متون نثر ادبی فارسی معاصر به عنوان نمونه، کارایی آن را در زبان فارسی خواهیم سنجید.

تحلیل کیوسام

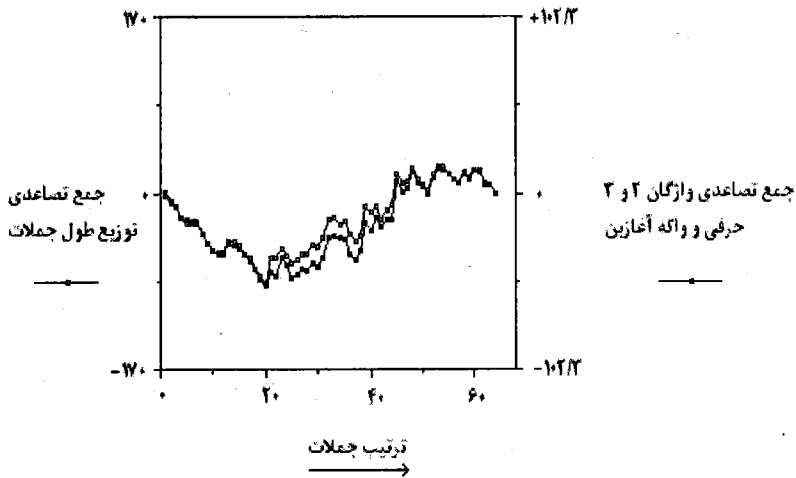
تحلیل کیوسام از دو مرحله تشکیل شده است که هر مرحله، به سه قسمت تقسیم می‌شود. در مرحله نخست به منظور تعیین توزیع طول جملات، ابتدا میانگین طول جملات متن را از طریق تقسیم تعداد واژگان موجود بر تعداد جملات به دست می‌آوریم، سپس تفاوت میانگین هر جمله را محاسبه کرده، و در گام سوم از مرحله

نخست، به جمع تصاعدی اعداد مربوط به تفاوت با میانگین جملات با یکدیگر می‌پردازیم و آنها را به صورت منحنی توزیع طول جملات در نمودار رسم می‌کنیم. در مرحله دوم، پس از تعیین تعداد واژگان پربسامد نقشی و واکه آغازین موجود در متن و محاسبه میانگین حضور این واژگان در هر جمله از طریق تقسیم کردن تعداد آنها بر تعداد جملات، تفاوت با میانگین هر جمله را از نظر دارا بودن تعداد واژگان پربسامد نقشی^۱ و واکه آغازین به دست آورده و در گام بعد، به جمع تصاعدی این اعداد می‌پردازیم و ارقام حاصل را به عنوان منحنی عادهای زبانی ناخود آگاه نویسنده بر روی نمودار می‌کشیم. در تحلیل کیوسام، چنانچه دو منحنی مذکور در نمودار (منحنی توزیع طول جملات و منحنی عادهای زبانی ناخود آگاه نویسنده) بر یکدیگر منطبق بوده و با یکدیگر همخوانی داشته باشند، متن مورد تحلیل یکدست تلقی می‌شود و یا انتساب متن مورد تحلیل به نویسنده‌ای خاص مورد تأیید قرار می‌گیرد. برای مثال منحنی حاصل از تحلیل ده جمله متعلق به لورنس (Lawrence) به شکل زیر خواهد بود:



نمودار ۱: تلفیق منحنی‌های توزیع طول جملات و عادهای زبانی نویسنده، به دست آمده از تحلیل ده جمله از نامه لورنس، در یک نمودار.

اما چنانچه متنی از نویسنده‌ای دیگر به نمودار بالا افزوده شود، تطابق میان دو منحنی از بین خواهد رفت:



نمودار ۲: تحلیل سه متن از دو نویسنده مختلف:

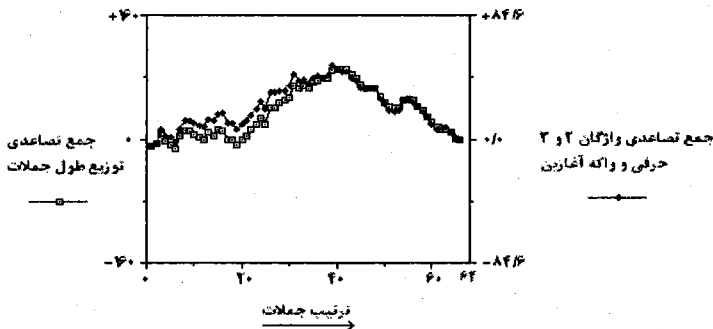
۲۰-۱ = نامه سال ۱۹۱۳ نوشته شده توسط لورنس.

۴۵-۲۱ = نامه هاکسلی.

۶۴-۴۶ = نامه سال ۱۹۰۷ نوشته شده توسط لورنس.

از این شیوه می‌توان برای بررسی متون منسوب به یک مؤلف نیز استفاده کرد. به این ترتیب که متن منسوب مورد تردید، میان متونی از نویسنده مورد نظر قرار داده می‌شود و در صورت همخوانی دو منحنی نمودار، انتساب متن به نویسنده مورد نظر تأیید می‌شود.

برای مثال در نمودار زیر، ادعای انتساب داستانی کوتاه با نام "جاده پستی" به لورنس مورد بررسی قرار می‌گیرد:



نمودار ۳: بررسی درستی یا نادرستی انتساب داستان کوتاه "جاده پستی" به لورنس:

۲۰-۱ = نامه لورنس.

۴۰-۲۱ = جملاتی از داستان "جاده پستی".

۶۴-۴۱ = نامه لورنس.

همان گونه که می‌بینیم، در دو سوی نمودار، منحنی مربوط به جمع تصاعدی طول جملات و نیز منحنی جمع تصاعدی واژگان پرسیامد و واژه آغازین (عادهای زیبایی نویسنده) بریکدیگر انطباق داشته، حال آن که در قسمت میانی که جملاتی از داستان کوتاه "جاده پستی" منسوب به لورنس درج گردیده است، این همخوانی از بین می‌رود. عدم تطابق منحنی‌های مزبور در قسمت میانی نمودار (جملات ۴۰-۲۱) بیان‌کننده این مطلب است که نویسنده جملات این بخش (جاده پستی)، متفاوت از نویسنده دو متن مورد تحلیل در دو سوی نمودار است. در نتیجه، بنا بر نتایج حاصل از تحلیل کیوسام، داستان کوتاه مذکور را نمی‌توان از نوشته‌های لورنس دانست. (Farrington, 2005-b)

امروزه تحلیل کیوسام در بسیاری از زمینه‌های مربوط به حوزه ادبیات از جمله تحلیل متون، بررسی نسخه‌های ویرایش شده، ترجمه‌ها، گفتار کودکان، بررسی و تشخیص نویسنده متون ناشناس و غیره مورد استفاده قرار می‌گیرد. به علاوه در تحلیل

متون حقوقی نیز می‌توان از یافته‌های حاصل از تحلیل کیوسام سود جست که مهمترین موارد کاربرد آن عبارتند از:

- اظهارات محتاطانه‌ای که تمام یا بخشی از آن توسط خواننده مورد تردید و اعتراض واقع شده است.

- اظهارات شهود، بخصوص شاهدانی که تحت پیگرد بوده و دادگاه معتقد است که تمام یا بخشی از این اظهارات توسط شخص یا اشخاص دیگری به جز شاهد مورد نظر نوشته شده‌اند.

- نامه‌های مربوط به خودکشی، بخصوص نامه‌هایی که با دستخط متوفی نگاشته شده‌اند، اما پرسش‌هایی مبنی بر این که آیا نگارش آن نامه به اختیار فرد بوده یا شخص دیگری فرد را مجبور به نگارش آن کرده است، مطرح باشند. به عبارت دیگر آیا مورد حاضر، خودکشی یا قتل است؟

- تماس‌های تلفنی ناشناس که در آن متهم از ارتکاب به آن امتناع می‌ورزد.
- مکالمات تلفنی یا ضبط شده‌ای که براساس شواهد، می‌توان آنها را به متهم منسوب دانست اما متهم آنها را انکار می‌کند.

- سرقت‌های ادبی، مشروط بر این که متهم، بخش‌هایی از نوشته‌های شخص دیگری را عیناً رو نوشت کرده باشد. بدیهی است، در مواردی که سارق ادبی، اندیشه‌های سرقت شده را به زبان و سبک خود نوشته باشد، تحلیل کیوسام نمی‌تواند راهگشا باشد.

- جعل اسناد، در مواردی که مؤلف تمام یا بخشی از یک سند مانند وصیت نامه مورد تردید باشد.

- کپی رایت یا حق نسخه برداری، در مواقعی که تردیدهایی درباره مؤلف نوشته‌هایی که سود دهی بالایی دارند نظیر کتابها، نمایشنامه‌ها، ترانه‌ها، اشعار و غیره وجود داشته باشد.

همانگونه که گذشت تحلیل کیوسام بر اساس محاسبه و مقایسه عمل می‌کند و در زبان انگلیسی، در امر تشخیص مؤلف و سبک فردی نویسنده، تعیین یکدستی اثر و نیز بررسی متون منسوب به مؤلف موفق بوده است. *فارینگدن* مدعی است که این شیوه را می‌توان بر انواع متون گفتاری و نوشتاری، بدون توجه به زمان خلق و نوع آنها اعمال کرد (2005-b). مورتون نیز به عنوان طراح شیوه کیوسام، مدعی گردیده است که این شیوه همواره درست عمل می‌کند، اما در مورد کیفیت آن نظری ندارد. به علاوه او معتقد است که تحلیل مذکور همیشه و در مورد هر نوع متن، حتی متون کوتاه موفق خواهد بود (Olsson, 2004: 16). *فارینگدن* این نکته را بیان می‌کند که تحلیل کیوسام علاوه بر زبان انگلیسی، در تمامی زبان‌های دیگر نیز موفق است و نتایج مطلوبی را عرضه خواهد کرد مگر این که خلاف آن ثابت شود و یا استثنایی برای آن یافت گردد. (2005-a) وی در توضیح این مطلب اشاره می‌کند که به هنگام خلق متن (نوشتاری یا گفتاری) در ۲۵ درصد مواقع، تنها از ۱۲ واژه پربسامد زبان انگلیسی استفاده می‌شود و از این رو، محاسبه توزیع این واژگان پربسامد می‌تواند به عنوان کلیدی برای تشخیص نویسنده و عادهای زبانی منحصر به فرد او، مؤثر باشد. هر چند که عمده تحلیل‌های انجام شده توسط *فارینگدن* بر اساس متون انگلیسی بوده‌اند، اما طبق گفته او، این تحلیل در زبان‌های دیگری مانند فرانسه، آلمانی، لاتین، یونانی باستان و ... نیز انجام شده و نتایج مطلوبی را در برداشته است (همان).

با وجود این، به نظر می‌رسد که هیچ یک از طراحان و پیروان تحلیل کیوسام نتوانسته‌اند توجیه مناسبی برای این مطلب بیان کنند که چگونه و بر اساس چه منطقی تنها کلیدهای جمع تصاعدی تفاوت با میانگین طول جملات و عادهای زبانی نویسنده (جمع تصاعدی تفاوت با میانگین جملات از نظر دارا بودن واژگان پربسامد و آکه آغازین) می‌توانند سبک فردی مؤلف را مشخص سازند. به علاوه در زبان‌هایی که واژه آکه آغازین ندارند (مانند فارسی)، وضعیت به چه شکل خواهد بود؟ *فارینگدن* در این مورد معتقد است که برای اعمال تحلیل کیوسام بر زبان فارسی باید

پربسامد ترین واژگان زبان را به عنوان کلیدی برای تشخیص عادهای زبانی نویسنده مشخص نمود و براساس آن، تحلیل را ادامه داد (2005-a).

در هر صورت، با وجود پرسش‌ها و انتقادهای موجود، این شیوه از سال ۱۹۹۰ علاوه بر متون ادبی، به صورت گسترده‌ای در دادگاهها و به منظور تحلیل نامه‌های مربوط به خودکشی افراد، اعترافات، شهادت‌شهود، تلفن‌های ناشناس، مکالمات ضبط شده، سرقت‌های ادبی و... مورد استفاده قرار گرفته است (Lindskoog, 1996).

چنین به نظر می‌رسد که استفاده از تحلیل کیوسام در امور خطیر و حساس دادگاهی، گواهی بر این نکته باشد که نتایج حاصل از اعمال این شیوه بر زبان انگلیسی و زبان‌های آزمایش شده دیگر، قابل اعتماد بوده و ضریب خطای پایینی در این زبان‌ها دارد.

با توجه به نکات مزبور، در ادامه با انتخاب شماری از متون ادبی نثر فارسی معاصر و اعمال شیوه تحلیلی کیوسام بر آنها، قابلیت و میزان اعتماد پذیری نتایج حاصل از این تحلیل در زبان فارسی مورد بررسی قرار خواهد گرفت.

اعمال تحلیل کیوسام بر پاره‌ای از متون نثر ادبی فارسی معاصر:

نگارنده با تحلیل متنی که دارای حدود ده هزار واژه بوده است، پربسامدترین واژه‌های نقشی زبان فارسی را به صورت زیر به دست آورده است:

و که را از به با در تا این آن است بود

همانطور که مشاهده می‌شود، در زبان فارسی واژه‌های پربسامد، یک، دو و سه حرفی هستند. در تحلیل‌های آینده، واژگان بالا به عنوان واژگان پربسامد متن محاسبه شده و تحلیل کیوسام در زبان فارسی بر اساس این واژگان انجام خواهد شد. با این حال در شمارش واژگان بالا، موارد زیر در نظر گرفته شده‌اند:

- "و" در اکثر موارد به عنوان واژه پر بسامد شمرده می‌شود، مگر این که در صورت های مربوط به اسامی خاص، اعداد، سالها، اصطلاحات و یا ضرب المثل ها باشد.

- "در" فقط هنگامی به عنوان واژه ای پر بسامد مورد شمارش قرار می‌گیرد که اولاً در اصطلاحات و ضرب المثل ها نباشد و ثانیاً به معنای "داخل/ درون" و به عبارتی واژه ای نقشی باشد، چرا که گاهی واژه "در" با معانی دیگری همچون "در/ باب" نیز مورد استفاده قرار می‌گیرد و بدیهی است در این گونه موارد دیگر به عنوان واژه نقشی پر بسامد مورد شمارش قرار نخواهد گرفت.

- فعل "است" به صورت های مختلفی نظیر "است" مثلاً "حوب است"، "-ست" مانند "حویست" و یا "ه" در "حویه" ظاهر می‌شود. در شمارش واژگان پر بسامد در تحلیل های این مقاله، فقط صورتی از "است" مورد شمارش قرار می‌گیرد که به شکل کامل خود، یعنی "است" ظاهر شده باشد (البته "است" در ضرب المثل ها و صورت های ثابت زبانی مانند اصطلاحات شمرده نخواهد شد).

- "تا" اگر به عنوان واحد شمارش مثلاً "دو تا زن" به کار رفته باشد، به عنوان واژه پر بسامد تلقی نمی‌شود. اما در مواردی که به عنوان حرف ربط استفاده شده است، شمرده می‌شود.

- "را" هنگامی که به عنوان نشانه مفعول و به صورت مستقل ظاهر شود، مورد شمارش قرار می‌گیرد و در صورت هایی نظیر "مر/ا"، "تور/ا" و ... محاسبه نخواهد شد.

توضیح این نکته ضروری است که هیچ یک از این موارد به این دلیل در ضرب المثل ها، اعداد و اصطلاحات مورد شمارش قرار نمی‌گیرند که این صورت های زبانی برای تمامی کاربران زبان ثابتند و بر اساس عاداتهای زبانی فردی نویسنده تغییر نمی‌کنند. مشکل دیگر، مربوط به شمارش واژگان واکه آغازین موجود در متن است که به عنوان یکی از متغیر های نشان دهنده عاداتهای زبانی نویسنده در تحلیل کیوسام معرفی

شده است. زیرا بیشتر دستور نویسان و نیز زبان شناسان فارسی مانند ثمره (۱۳۶۴) معتقدند که در زبان فارسی هیچ هجا و در نتیجه هیچ واژه ای با واکه آغاز نمی‌شود و حتی در مورد واژگانی نظیر "اسب"، "آسمان"، "امروز"، "امید"، "او" و غیره که ظاهراً با واکه آغاز شده‌اند، واج /۲/ در ابتدای این واژگان قرار گرفته است.

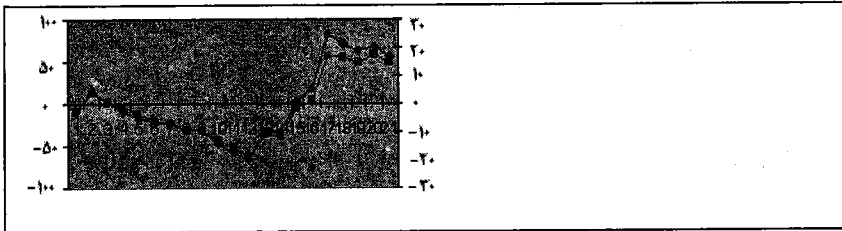
از آن جا که بر پایه این نظر، زبان فارسی جزو زبان های فاقد واژه واکه آغازین است و به دلیل این که در این مقاله بر این اساس عمل می‌شود، به هنگام اعمال تحلیل کیوسام بر متون فارسی، از این متغیر صرف نظر شده و به منظور تعیین رفتار زبانی نویسنده متن فارسی تنها به شمارش واژگان پر بسامدی که پیش از این ذکر گردیده است، بسنده خواهد شد. این تصمیم‌گیری پس از انجام مکاتبات الکترونیکی با *فارینگدن (2005-a)*، یکی از صاحب نظران و طراحان شیوه تحلیلی کیوسام، و بیان ویژگی‌های زبان فارسی مبنی بر عدم وجود واژه واکه آغازین در آن، مطابق نظر وی انجام شده است.

با در نظر گرفتن این موارد، در ادامه به منظور بررسی کارایی شیوه تحلیلی کیوسام در زبان فارسی، ابتدا چند مورد از مواردی که این تحلیل در تشخیص یکدستی متن در زبان فارسی موفق بوده است به عنوان نمونه ذکر و به دنبال آن نمونه‌هایی از نا کارآمدی و عدم موفقیت شیوه تحلیلی کیوسام در تشخیص یکدستی متن بیان شده است. سپس تأثیر تغییر در ترتیب متون مورد تحلیل، بررسی خواهد شد. بدین معنی که آیا در تحلیل متون فارسی بر اساس شیوه کیوسام، جابه‌جایی و تغییر در توالی متون مورد نظر، تأثیری در نتیجه نهایی خواهد داشت؟ پس از آن نیز چند جمله از یک متن که نویسنده آن برای نگارنده نامعلوم بوده است همراه با جمله‌هایی از یک نویسنده شناخته شده تحلیل شده است تا بدین ترتیب کارایی تحلیل کیوسام در انتساب متنی ناشناس یا مورد تردید به نویسنده‌ای خاص مورد ارزیابی قرار گیرد. در نمودارهای پیش رو، محور افقی نشان دهنده ترتیب جملات تحلیل شده از چپ به راست، محور

عمودی سمت راست بیانگر میزان جمع تصاعدی واژگان پربسامد موجود در متن و محور عمودی سمت چپ مربوط به جمع تصاعدی تفاوت با میانگین طول جملات متن خواهد بود.

الف- موارد تأیید کننده:

در این بخش دو نمونه از موارد نشان دهنده موفقیت شیوه تحلیل‌سی کیوسام در تشخیص یکدستی متون فارسی ذکر می‌شود. بدین ترتیب که ابتدا سه متن ده جمله‌ای از سه داستان مختلف از آل احمد تحلیل شده که نمودار حاصل از آن را می‌توان در زیر مشاهده کرد:



نمودار ۴: تحلیل سه متن ده جمله‌ای (سی جمله) از آثار آل احمد.

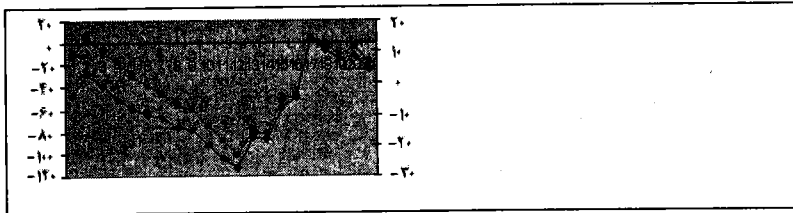
۱-۱۰ = از داستان "لاک صورتی" اثر آل احمد.

۲۰-۱۱ = از داستان "دزد زده" اثر آل احمد.

۳۰-۲۱ = از داستان "الگمارک و الگموس" اثر آل احمد.

همانگونه که مشاهده می‌شود در نمودار بالا، از آن جا که نویسنده هر سه متن شخص واحدی بوده است، دو منحنی مربوط به جمع تصاعدی طول جملات و جمع تصاعدی واژگان پربسامد (عادهای زبانی نویسنده) تا حد زیادی بریکدیگر انطباق داشته‌است.

این در حالی است که افزوده شدن ده جمله دیگر از نویسنده‌ای متفاوت به تحلیل بالا، باعث بر هم خوردن این انطباق شده است:



نمودار ۵: تحلیل سه متن ده جمله‌ای (سی جمله) از آثار آل احمد به همراه یک متن ده جمله‌ای از نوشته‌های علوی.

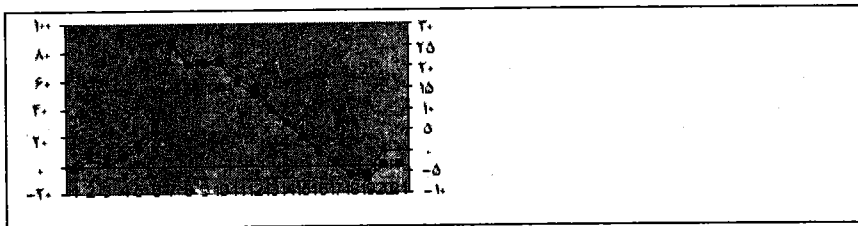
۱۰-۱ = از داستان "لاک صورتی" اثر آل احمد.

۲۰-۱۱ = از داستان "دزد زده" اثر آل احمد.

۳۰-۲۱ = از داستان "الگمارک و الکموس" اثر آل احمد.

۴۰-۳۱ = از داستان "سرباز سربی" اثر علوی.

به عنوان نمونه‌ای دیگر، تحلیل بیست جمله از دو متن ده جمله‌ای از دو داستان مختلف علوی نموداری به شکل زیر به دست داده است که در آن دو منحنی بر یکدیگر منطبق هستند و این نشان دهنده یکسان بودن نویسنده دو متن است:

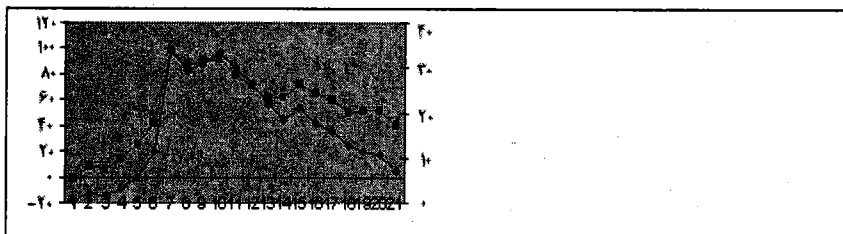


نمودار ۶: تحلیل بیست جمله از دو متن از آثار علوی.

۱۰-۱ = از داستان "دز آشوب" اثر علوی.

۲۰-۱۱ = از داستان "گیله مرد" اثر علوی.

حال آن که با اضافه شدن ده جمله از نوشته‌های گلی ترقی به تحلیل بالا، همپوشی موجود میان دو منحنی، دیگر مشاهده نمی‌شود:



نمودار ۷: تحلیل بیست جمله از دو متن از آثار علوی به همراه یک متن ده جمله‌ای از نوشته‌های گلی ترقی.

۱-۱۰ = از داستان "دزآشوب" اثر علوی.

۱۱-۲۰ = از داستان "بازی ناتمام" اثر گلی ترقی.

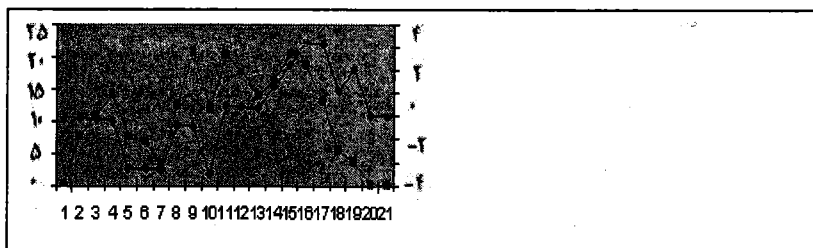
۲۱-۳۰ = از داستان "گیله مرد" اثر علوی.

ب- موارد نقض کننده:

بر خلاف نمونه‌های مشاهده شده در بخش پیشین، که نشان دهنده موفقیت تحلیل کیوسام در تعیین یکدست بودن متون فارسی بودند، نمونه‌های فراوان دیگری را نیز می‌توان یافت که حاکی از نا کارآمدی تحلیل مزبور در زبان فارسی هستند که در ادامه دو نمونه ذکر می‌گردد.

برای مثال نمودار زیر نتیجه تحلیل اعمال شده بر بیست جمله از یک متن از آثار

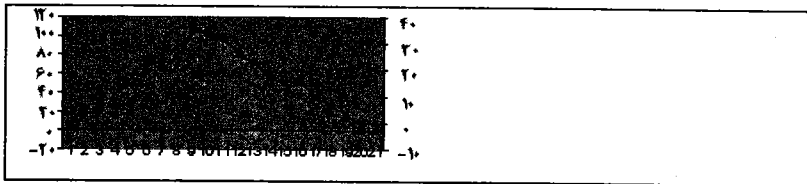
زویا پیرزاد است:



نمودار ۸: تحلیل بیست جمله از داستان "عادت می‌کنیم" اثر زویا پیرزاد.

همانطور که مشاهده می‌شود، با وجود این که تمامی جملات تحلیل شده بالا به وسیله یک نویسنده نوشته شده‌اند، اما دو منحنی در نمودار اختلاف چشمگیری با یکدیگر دارند.

نمونه‌ای دیگر از موارد نقض تحلیل کیوسام در زبان فارسی را می‌توان در مورد زیر مشاهده کرد. در نمودار زیر با این که دو متن ده جمله‌ای از دو نویسنده مختلف تحلیل شده‌اند. اما بر خلاف پیش‌بینی کیوسام، انطباق قابل توجهی میان دو منحنی نمودار برقرار است:

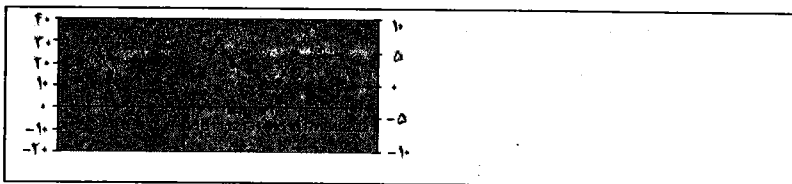


نمودار ۹: تحلیل بیست جمله از دو متن مختلف از دو نویسنده متفاوت.

۱-۱۰ = از داستان "بوف کور" اثر هدایت.

۱۱-۲۰ = از داستان "قربانی" اثر علوی.

در تحلیل زیر نیز بیست جمله از نوشته‌های آل احمد از دو داستان مختلف او تحلیل شده است، اما همانگونه که مشاهده می‌شود در نمودار مربوط به آن، با وجود یکی بودن نویسنده تمامی جملات، دو منحنی با یکدیگر همپوشی نداشته و میان آن دو فاصله محسوسی به چشم می‌خورد:



نمودار ۱۰: تحلیل بیست جمله از دو متن مختلف از آثار آل احمد.

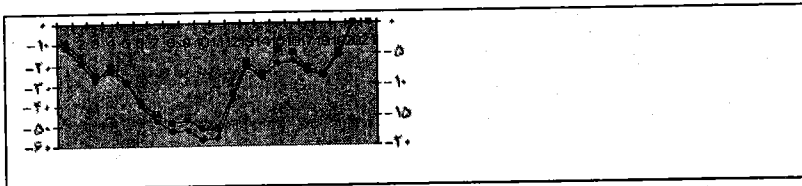
۱-۱۰ = از داستان "نوبر توت فرنگی".

۲۰-۱۱ = از داستان "عکاس با معرفت".

پ- بررسی تأثیر تغییر در ترتیب متون مورد تحلیل:

در این بخش، این نکته بررسی می‌شود که آیا در مواردی که انطباق منحنی‌های موجود در نمودار، نشان دهنده یکسان بودن نویسنده جملات مورد تحلیل بوده است، تغییر در ترتیب متون مورد تحلیل، تغییری در نتیجه نهایی یعنی منطبق بودن منحنی‌ها بر یکدیگر دارد؟ بر این اساس در ادامه ابتدا دو متن از آثار احمد و سپس دو متن از علوی تحلیل شده و تأثیر جابه‌جایی ترتیب متون در نتیجه تحلیل، مورد بررسی قرار می‌گیرد.

در نمودار زیر دو متن مختلف از آثار آل احمد تحلیل شده است:

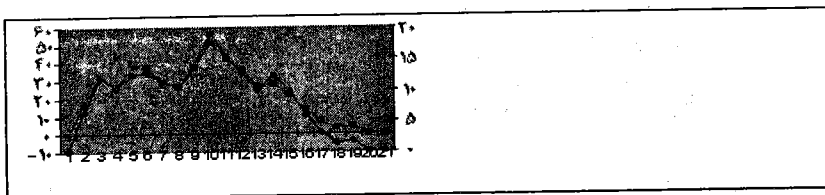


نمودار ۱۱: تحلیل دو متن ده جمله‌ای از آثار آل احمد.

۱۰-۱ = از داستان "بچه مردم".

۲۰-۱۱ = از داستان "سه تار".

در نمودار بالا، میان دو منحنی، تطابق کامل مشاهده می‌شود. چنین به نظر می‌رسد که در مورد این دو متن، تغییر در ترتیب، تأثیری ندارد. چرا که در نمودار زیر که در آن دو متن قبل جابجا شده‌اند، همچنان تطابق میان دو منحنی برقرار است:

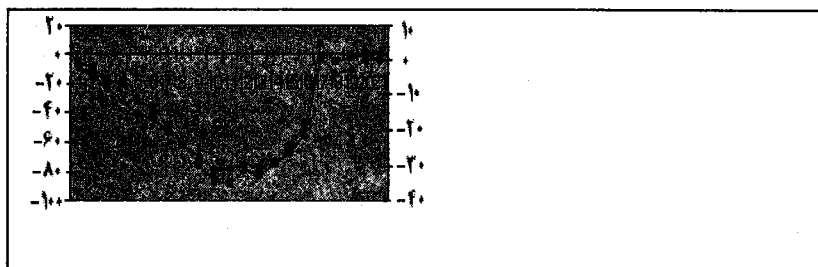


نمودار ۱۲: تغییر در ترتیب دو متن نمودار قبل از آثار آل احمد.

۱۰-۱ = از داستان "سه تار".

۱۱-۲۰ = از داستان "بیچۀ مردم".

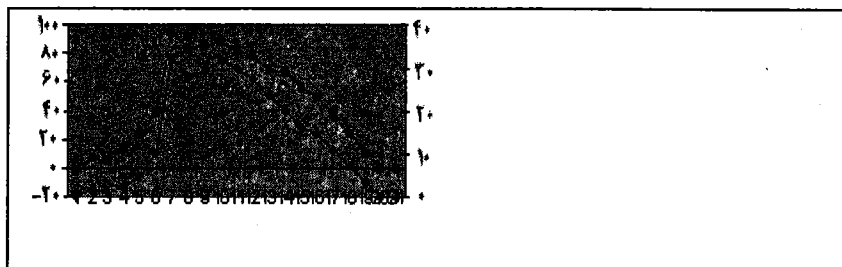
با وجود این، در بسیاری از موارد نیز مشاهده می‌شود که ترتیب متون مورد تحلیل در شکل نهایی نمودار و انطباق منحنی‌ها بر یکدیگر تأثیر دارد. برای مثال در تحلیل متون زیر (ده جمله از داستان "خائن" و ده جمله از داستان "دزآشوب" از آثار علوی) ترتیب قرار گرفتن آن دو در تحلیل، تأثیر قابل توجهی در انطباق یا عدم انطباق منحنی‌های نمودار با یکدیگر دارد:



نمودار ۱۳: تحلیل دو متن ده جمله‌ای از آثار علوی.

۱-۱۰ = از داستان "خائن".

۱۱-۲۰ = از داستان "دزآشوب".



نمودار ۱۴: تغییر ترتیب دو متن نمودار قبل از آثار علوی.

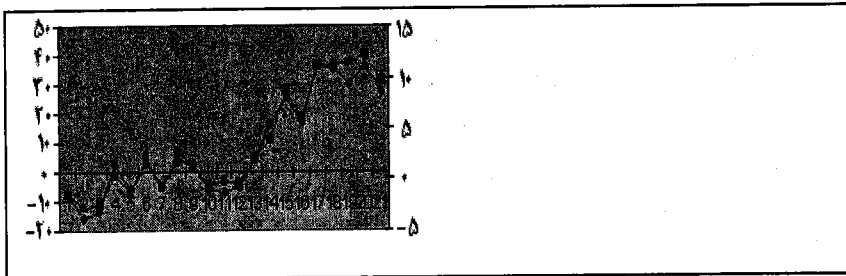
۱-۱۰ = از داستان "دزآشوب".

۱۱-۲۰ = از داستان "خائن".

بر این اساس، با توجه به این که ترتیب متون تحلیل شده می‌تواند منجر به نتایج متفاوتی شود، چنین به نظر می‌رسد که نتایج حاصل از شیوه تحلیلی کیوسام در زبان فارسی نمی‌تواند چندان قابل اعتماد باشد.

ت- بررسی کارآیی تحلیل کیوسام در انتساب متن به یک نویسنده:

در این قسمت، قابلیت شیوه تحلیلی کیوسام در تعیین نویسنده و انتساب متن، مورد بررسی قرار می‌گیرد. بدین منظور، متنی که نویسنده آن برای نگارنده نامعلوم بوده است همراه با متونی از یک نویسنده خاص (علوی) تحلیل شده است تا مشخص شود که آیا نویسنده متن ناشناس، همان نویسنده دو متن دیگر است یا نه. نکته جالب این است که بر اساس نتایج نمودارها، گاهی متن ناشناس را باید متعلق به علوی دانست و گاهی نیز این انتساب متن به علوی نادرست است:



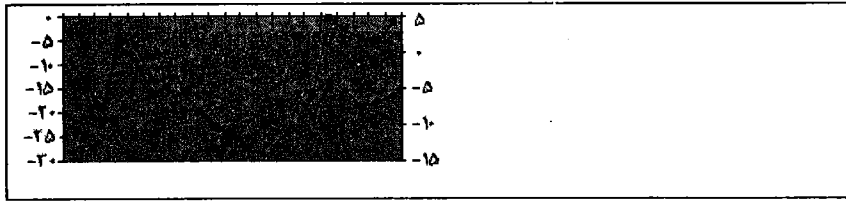
نمودار ۱۵: تحلیل دو متن از آثار علوی به همراه یک متن ناشناس.

۱-۱۰ = از داستان "عروس هزار داماد" اثر علوی.

۱۱-۲۰ = متن ناشناس.

۲۱-۳۰ = از داستان "رقص مرگ" اثر علوی.

همانگونه که مشاهده می‌شود با در نظر گرفتن نمودار بالا، متن ناشناس باید یکی از نوشته‌های علوی باشد چرا که در آن، دو منحنی با یکدیگر همپوشی داشته‌اند. اما اگر همین متن ناشناس با متن‌های دیگری از علوی مورد تحلیل قرار گیرد نتیجه‌ای کاملاً متفاوت به دست خواهد آمد:



نمودار ۱۶: تحلیل دو متن از آثار علوی به همراه متن ناشناس.

۱-۱۰ = از داستان "یره نچکا" اثر علوی.

۱۱-۲۰ = متن ناشناس.

۲۱-۲۹ = از داستان "یک زن خوشبخت" اثر علوی.

با توجه به نمودار بالا، این بار نویسنده متن ناشناس نمی‌تواند علوی باشد چرا که منطبق نبودن منحنی‌های نمودار، شاهدهی بر این مطلب هستند.

با در نظر گرفتن نکات بالا، به نظر می‌رسد که نتایج حاصل از تحلیل کیوسام برای تعیین انتساب متون فارسی نیز همیشه نمی‌تواند قابل اعتماد باشد.

نتیجه‌گیری

مقاله حاضر که با هدف سنجش کارایی شیوه تحلیلی کیوسام در تشخیص و تعیین نویسنده متون فارسی نوشته شده است، نشان می‌دهد که نتایج به دست آمده از تحلیل متون فارسی بر اساس این شیوه نمی‌تواند همواره درست و قابل استناد باشد. چرا که برخلاف موارد ذکر شده در بخش الف، در سایر بخش‌ها (ب، پ و ت) نمونه‌هایی از عدم موفقیت این شیوه تحلیلی در زبان فارسی قابل مشاهده است.

هر چند بهتر بود که به منظور ارزیابی تحلیلی کیوسام در تشخیص مؤلف متون فارسی، علاوه بر متون ادبی، متونی از دیگر حوزه‌ها مانند متون حقوقی نیز انتخاب می‌شدند تا بتوان این تحلیل را بر آنها اعمال کرد، اما به دلیل عدم دسترسی به متون حقوقی از قبیل نامه‌های کسانی که دست به خودکشی زده‌اند و امکان پذیر نبودن استفاده از آنها برای نگارنده، در مقاله حاضر تنها به بررسی کارایی تحلیلی کیوسام در

زمینه تشخیص مؤلف متون نثر فارسی معاصر بسنده شده است. با وجود این، همان گونه که مشاهده شد، تحلیل این دسته از متون به منظور ارزیابی شیوه مذکور کافی به نظر می‌رسد. چرا که عدم کارایی و غیر قابل اتکا بودن نتایج حاصل از تحلیل کیوسام در تشخیص مؤلف متون ادبی متثور فارسی معاصر، گواهی بر ناکارایی آن در این زبان و در سایر حوزه‌ها نیز می‌تواند باشد.

اگر چه مقاله حاضر نتوانسته است راهکاری قابل اعتماد برای تشخیص مؤلف متون فارسی ارائه نماید و در این باره شیوه تحلیلی کیوسام، به عنوان یک رویکرد مطرح‌زبانی-آماری، نتایج مطلوبی در زبان فارسی به دست نداده است، با این حال به نظر می‌رسد مسأله نبودن روشی مطمئن در زمینه تشخیص مؤلف و انتساب متون فارسی و ضعف شیوه‌هایی که ملاک‌های سبکی، موضوعی و نگرشی را در این زمینه به کار می‌برند، همچنان به قوت خود باقی است. از این رو، شایسته است تلاش‌های دیگری از سوی پژوهشگران در زمینه ارائه راهکاری قابل اعتماد و رفع این کاستی انجام شود. استفاده از کلیدهای دیگر، به جز آنچه که در این تحلیل برای اعمال تحلیل کیوسام بر متون فارسی مورد استفاده قرار گرفته است، از قبیل در نظر گرفتن میزان استفاده مؤلف از توصیف‌ها، تشبیه‌ها و نیز اضافه‌های ادبی برای تعیین عاداتهای زبانی ناخود آگاه نویسنده، از جمله پیشنهادهایی است که می‌تواند مورد بررسی قرار گیرد. به علاوه، از آن جا که به نظر می‌رسد چگونگی به کارگیری صنایع ادبی ذکر شده در مورد تمامی افراد یکسان نباشد، بررسی چگونگی استفاده نویسندگان مختلف از صنایع مذکور در موقعیت‌های مشابه، پیشنهاد می‌شود. بدیهی است لازمه انجام دقیق این‌گونه از پژوهش‌ها، دسترسی به پیکره زبانی مناسبی از متون متنوع فارسی خواهد بود.

ضمناً، از آن جا که تشخیص مؤلف در زبان‌هایی نظیر انگلیسی، بر اساس کلیدهای زبانی و آماری تا حدودی امکان پذیر شده است، امکان دستیابی به رویکردی برای تشخیص مؤلف متون فارسی که از متغیرهای مشابه و نه لزوماً یکسان درون متنی سود جوید، دور از انتظار نیست. با این حال، در نظر داشتن این نکته نیز ضروری است که

ممکن است بررسی‌های بیشتر در این زمینه، حاکی از امکان پذیر نبودن این مسأله باشد.

به امید آن که این مقاله، انگیزه کافی جهت پژوهش‌های بیشتر در این زمینه را در علاقه‌مندان ایجاد کرده باشد.

یادداشت‌ها

این مقاله برگرفته از پایان نامه آقای امیرمحمدیان است با عنوان «بررسی و ارزیابی کارایی تحلیل کیوسام در تشخیص مؤلف متون فارسی» که به راهنمایی دکتر اعظم استاجی در دانشگاه فردوسی مشهد انجام شده است.

۱- براساس (Farrington, 2005-b)، ۱۲ واژه نقشی پر بسامد در زبان انگلیسی عبارتند

از: of, the, was, to, a, and, he, I, in, is, it.

منابع و مأخذ

الف) کتابها

- ۱- آل احمد، جلال، (۱۳۷۴)، گزیده داستان‌ها، تهران، انتشارات فردوس.
- ۲- بهار، محمد تقی، (۱۳۷۰)، سبک‌شناسی یا تاریخ‌تطور نثر فارسی، تهران، انتشارات امیر کبیر.
- ۳- پیرزاد، زویا، (۱۳۸۳)، عادت می‌کنیم، تهران، نشر مرکز.
- ۴- ترقی، گلی، (۱۳۸۰)، جایی دیگر، تهران، انتشارات نیلوفر.
- ۵- ثمره، ید...، (۱۳۶۴)، آواشناسی زبان فارسی، تهران، مرکز نشر دانشگاهی.
- ۶- شای، راجر. و.، (۱۳۸۳)، زبان‌شناسی قانونی، ترجمه اعظم استاجی، مجله حقوقی دادگستری ۴۹-۴۸ : ۳۸۰-۳۶۵.
- ۷- علوی، بزرگ، (۱۳۸۱-الف)، چمدان، تهران، انتشارات نگاه.
- ۸- _____، (۱۳۸۱-ب)، گیله مرد، تهران، انتشارات نگاه.

۹- هدایت، صادق، (۱۳۵۱)، بوف کور، تهران، مؤسسه انتشارات امیرکبیر.

ب) مقالات

۱- گلپرور، فرشته، (بی تا)، همسفر، گاهنامه داستان، شماره ۴، تهران، انتشارات حوزه هنری سازمان تبلیغات اسلامی، صص ۹-۱۳.

منابع لاتین

- 1- Farrington, Jill (2004), Funeral Elegy, Not Written by W.S. after all, Studies in Bibliography, Virginia: Virginia University Publications.
- 2- Farrington, Jill (2005-a), E-mails sent to A'zam Estaji (Assistant professor of Linguistics at Ferdowsi University of Mashhad/IRAN), www.estajiz@yahoo.com.
- 3- Farrington, Jill (2005-b), How to be a literary detective: Authorship attribution, <http://www.members.aol.com/qusums/Qsumintroduction.html>.
- 4- Lindskoog, Kathryn (1996), Analysing for Authorship with the Cusum Technique, The Lewis Legacy-Issue 69.
- 5- Olsson, John (2004), Forensic Linguistics, an introduction to language, crime and law, London, New York: Continuum.

Authorship Attribution Based on Cusum Analysis

**Amir Mohammadian, PhD candidate
Ferdowsi University of Mashhad**

Abstract

Authorship Attribution, a branch of Forensic Linguistics, is concerned with the attribution of a text (from literary to forensic) to its author via criteria presented by each method. Since most of the work done on Persian authorship attribution rests primarily on stylistic & subjective considerations and objective criteria have not been considered, dispute over the attributed texts are observed. Hence, in the present study, after reviewing the literature on authorship attribution, one of the current and most prevalent methods based on linguistic-statistical keys, by the name of Cusum Analysis, is firstly introduced and then using some contemporary Persian prose texts, the efficiency of the mentioned method in attributing Persian texts will be evaluated.

Keywords: *Forensic Linguistics, Authorship attribution, Cusum analysis, Average deviation, Cumulative sum.*