

تشخیص مؤلف متون ادبی و قانونی، بحثی در زبان‌شناسی قانونی^۱

اعظم استاجی

دانشگاه فردوسی مشهد

چکیده

در حوزه ادبیات و سبک‌شناسی، تشخیص نویسنده یک متن (اعم از نثر و نظم) از اهمیت زیادی برخوردار است. گاهی با توجه به ویژگی‌های سبکی، محتوایی و معنایی متن، انتساب آن به یک نویسنده مشخص مورد تردید قرار می‌گیرد. مسئله تشخیص نویسنده متن در حوزه‌های حقوقی و قانونی نیز اهمیت زیادی دارد. مقاله حاضر پس از مروری بر پیشینه مطالعات انجام‌شده در زمینه تشخیص نویسنده متن به معرفی برخی رویکردهای آماری و غیرآماری در این زمینه می‌پردازد و با ارائه نمونه‌هایی آن‌ها را مورد ارزیابی قرار می‌دهد. **کلیدواژه‌ها:** تشخیص نویسنده متن^۲، سبک فردی، زبان‌شناسی قانونی، سبک‌شناسی قانونی.

۱. مقدمه

یکی از مسائل بحث برانگیز در حوزه ادبیات و سبک‌شناسی، تشخیص نویسنده متن (اعم از نثر و نظم) است. گاهی با توجه به ویژگی‌های سبکی، محتوایی و معنایی متن، انتساب آن به یک نویسنده مشخص مورد تردید قرار می‌گیرد. به عنوان مثال مباحث مربوط به رباعیات خیام یا آثار شکسپیر در این مقوله جای می‌گیرند. در این مباحث با تکیه بر ویژگی‌هایی چون افکار، مبانی فلسفی، زبان و ساختار اشعار خیام به قبول یا رد رباعیات منسوب به او می‌پردازند.

مسئله تشخیص نویسنده متن در حوزه‌های حقوقی و قانونی نیز اهمیت زیادی دارد. به عنوان مثال در نامه‌های تهدیدآمیزی که به‌طور ناشناس فرستاده می‌شوند، تشخیص نویسنده یا نویسندگان مظنون به حل دعاوی قانونی و حقوقی کمک زیادی می‌کند.

مقاله حاضر پس از مروری بر پیشینه مطالعات انجام شده در زمینه تشخیص نویسنده متن به معرفی برخی رویکردهای آماری و غیرآماری در این زمینه می‌پردازد و با ارائه نمونه‌هایی (ادبی و قانونی) آن‌ها را مورد ارزیابی و بررسی قرار می‌دهد.

۲. بررسی‌های اولیه

به نظر نمی‌رسد در دنیای غرب بحث مربوط به تشخیص نویسنده یک متن تا قبل از قرن هجدهم مطرح شده باشد، چرا که تا قبل از این دوره کتاب‌ها به‌طور گسترده‌ای توزیع نمی‌شدند و مفهوم دکارتی فردیت و تملک فردی افکار مورد قبول عام بود. نظر عموم بر این است که اولین مورد شناخته شده از بحث درباره نویسنده متن به کتاب مقدس مربوط می‌شود. کشیشی آلمانی به نام اچ. بی. ویتر^۳ در سال ۱۷۱۱ به این

۱ در این مقاله اصطلاح زبان‌شناسی قانونی به قیاس با اصطلاح "پزشکی قانونی" در مقابل عبارت Forensic Linguistics به کار رفته است

۲ authorship attribution

۳ H. B. Witter

سال چهارم - شماره دوم - پیاپی هشتم - پاییز و زمستان ۱۳۸۷ - ۱۸۰ صفحه - ۱۷۰۰۰ ریال

زبان و زبان‌شناسی

مجله انجمن زبان‌شناسی ایران

علمی - پژوهشی



انجمن زبان‌شناسی ایران

نکته اشاره کرد که نام‌های متفاوت رب‌النوع در کتب پنج‌گانه نشان می‌دهد که نویسندگان متعددی آن‌ها را نوشته‌اند (السن، ۲۰۰۴: ۱۱).

منازعات مربوط به آثار شکسپیر نیز به قرن هجدهم بر می‌گردد. در سال ۱۷۸۵ جیمز ویلموت^۱ بیان داشت که بیکن نویسنده واقعی نمایش‌نامه‌های شکسپیر است، اما از آنجا که هیچ علاقه‌ای به بحث و جدل نداشت عقاید خود را منتشر نکرد. این منازعات تا سال ۱۸۰۵ که جیمز کورتون کاول^۲ عقاید خود را در انجمن فلسفی ایپسویچ^۳ بیان داشت صورت آشکاری به خود نگرفت (همان: ۱۱). پس از این، حدود نیم قرن طول کشید تا علاقه به مسئله تشخیص نویسنده متن وجه علمی به خود بگیرد. از نیمه قرن نوزدهم به این سو رویکردهای آماری متفاوتی برای تشخیص نویسنده متن ارائه شده است. در این رویکردها تلاش شده است تا با تکیه بر برخی ویژگی‌ها و شاخص‌های آماری متون و مقایسه آن‌ها با یکدیگر، نویسنده متن مشخص گردد.

در واقع در این تحقیقات تلاش شده است تا به شیوه‌ای عینی راهی برای کشف سبک فردی نویسنده یا به عبارتی اثر انگشت زبانی نویسنده پیدا شود، چرا که برخی معتقدند همان‌طور که هرکسی را می‌توان بر اساس اثر انگشتش شناسایی کرد، هر متن هم اثر انگشت نویسنده خود را به همراه دارد و از طریق متن می‌توان نویسنده را شناسایی کرد. منظور از اثر انگشت زبانی، مجموعه‌ای از نشانگرهاست که یک سخنگو یا نویسنده را از دیگران متمایز می‌کند (همان: ۳۱). در همین زمینه شمیسا (۱۳۷۲: ۷۵) می‌نویسد: «هرکسی کم و بیش برای خود سبک شخصی دارد. این که می‌گوییم آثار کثیری از شاعران و نویسندگان در یک دوره شبیه به هم است سخنی است مبتنی بر مسامحه. آری به قول لیچ هر نویسنده اثر انگشت زبانی‌بی دارد.» به نظر شمیسا (همان: ۷۷) کشف مختصات سبک فردی و خصوصاً ویژگی بنیادی سبک شخصی غبطه هر سبک‌شناسی است و دشوارترین مرحله سبک‌شناسی محسوب می‌شود. به اعتقاد او فقط سبک‌شناسان باتجربه، که عمری دراز با آثار نویسندگان و شاعران خاصی محشور بوده‌اند، می‌توانند از عهده تبیین مختصات سبکی ویژه برآیند.

در زبان فارسی، اگر مشخصاً بحث خود را به خیام محدود کنیم، مبنای کار محققان برای تشخیص رباعیات خیام مراجعه به قدیم‌ترین آثار و نوشته‌هایی است که صریحاً رباعیاتی را از خیام ذکر کرده‌اند. پس از این هر محقق با مبنا قرار دادن تعدادی از این رباعیات و توجه به افکار، اندیشه‌ها و شخصیت خیام به بررسی دیگر رباعیات منسوب به او پرداخته است. به عنوان مثال فیتزجرالد^۴ از نخستین کسانی است که برای تشخیص رباعیات اصیل خیام از رباعیاتی که به او نسبت داده‌اند، از معیار هماهنگی و موافقت با اندیشه و شخصیت خیام استفاده کرده است. کسانی چون فردریش روزن^۵، صادق هدایت، فروغی، رشید یاسمی، حسین دانشفر و علی دشتی نیز بر همین اساس رباعیات اصیل خیام را انتخاب کرده‌اند (بدره‌ای، ۱۳۷۴: ۳۰). اما کریستین سن^۶، محقق دانمارکی، در سال ۱۹۲۷ با معیاری متفاوت گزینش را انجام

^۱ J. Olsson

^۲ J. Wilmot

^۳ J. C. Cowell

^۴ Ipswich

^۵ E. Fitzgerald

^۶ F. Rosen

^۷ A. Christiansen

داده است و از مجموع ۱۲۳۱ رباعی منسوب به خیام ۱۲۱ رباعی را اصیل شمرده است. شیوه مورد استفاده او شیوه‌ای بسامدی و آماری است. در این شیوه کریستین سن براساس ترتیب و بسامد قرار گرفتن رباعیات در نسخه‌های مختلف دست به انتخاب زده است (همان.)، اما در این شیوه آماری به ویژگی‌های آماری نهفته در متن اشعار توجهی نشده است.

۳. رویکردهای آماری

یکی از نخستین کسانی که به مسئله تشخیص نویسنده متن ابراز علاقه کرده است، آگوستوس دمورگان^۱ اولین استاد ریاضیات در دانشگاه لندن است. در سال ۱۸۵۱ دمورگان به دوستش، که به مسئله تشخیص نویسنده نامه‌هایی که به سنت پل^۲ منسوب بود ابراز علاقه کرده بود، نامه‌ای نوشت و پیشنهاد کرد که:

یکی از متون هرودوت، مثلاً کتاب اولش، را انتخاب کن و تمام واژه‌ها و حروف سازنده واژه‌ها را بشمار. تعداد حروف را بر تعداد واژه‌ها تقسیم کن تا میانگین تعداد حروف واژه‌ها را در آن کتاب به دست آوری. همین کار را برای کتاب دوم انجام بده. پیش بینی می‌کنم که نتایج بسیار به هم نزدیک باشند. اگر در کتاب اول میانگین تعداد حروف ۵/۶۲۴ باشد، تعجب‌آور نیست اگر در کتاب دوم این رقم برابر ۵/۶۱۵ باشد (نقل از السن، ۲۰۰۴: ۱۲).

پس از دمورگان کسان دیگری مثل مندنهال^۳ در سال ۱۸۸۷، ادنی یول^۴ در سال ۱۹۴۴، ماسترل^۵ و والاس^۶ در سال ۱۹۶۴ و کنی^۷ در سال ۱۹۸۲ شاخص‌های آماری دیگری مثل میانگین طول واژه، میانگین طول جمله و ... را برای تشخیص نویسنده متن معرفی کردند. السن (همان.) پس از مروری بر این رویکردهای آماری به این نکته مهم توجه می‌کند که رویکردهای آماری محض این ایراد را دارند که هیچ نوع دانش زبانی‌ای در آن‌ها لحاظ نمی‌شود. به بیان السن: «اگر یول زبان‌شناس بود می‌دانست که تفاوت زیادی بین طول جمله در گفتار و نوشتار وجود دارد. به عبارت دقیق‌تر با افزایش میزان رسمی بودن، میانگین طول جمله کاهش می‌یابد. طول جمله مانند هر مقیاس دیگری بیشتر وابسته به سبک و سیاق است تا نویسنده‌ای خاص» (همان.). مهدی یازرگان (۱۳۵۵) از جمله نخستین کسانی است که با اصطلاح "جمله‌شناسی" به ویژگی‌های آماری متن در زبان فارسی اشاره می‌کند و بیان می‌دارد در حالی که در فن سبک‌شناسی کلیه جهات لفظی، آهنگ، معنی و تعبیرهای ادبی کلام یک شاعر یا صاحب اثر بررسی می‌شود، در جمله‌شناسی صرفاً به تعداد کلمات تشکیل دهنده جملات توجه می‌شود. در "جمله‌سنجی" پس از شمارش جملات و تعداد واژه‌های به کاررفته در هر جمله در دستگاه ترمیمی مختصات دکارتی، محور طول برحسب تعداد واژه‌های تشکیل دهنده جملات تقسیم‌بندی می‌شود. درصد تعداد جملات دارای آن مقدار واژه در محور

^۱ A. de Morgan

^۲ St. Paul

^۳ T. C. Mendenhall

^۴ U. Yule

^۵ F. Mosteller

^۶ D. Wallace

^۷ A. Kenny

این اعداد نهایی را می‌توان روی نموداری نمایش داد که محور افقی آن نشان‌گر تعداد جملات و محور عمودی آن نشان‌گر اعداد حاصل از مرحله سوم (جمع انباشتی طول جمله) است. نقطه صفر نشان‌گر میانگین است، اعداد بالاتر از میانگین، بالای نقطه صفر و اعداد کوچکتر از میانگین زیر نقطه صفر قرار می‌گیرند. اما این نمودار به تنهایی به تشخیص نویسنده متن کمک نمی‌کند. بنابراین باید آن را با عامل دیگری که نشان‌دهنده عادات زبانی ناخودآگاه نویسنده است مقایسه کرد. این عادات زبانی کلید شناسایی نویسنده‌اند. یکی از این عادات زبانی استفاده از واژه‌های نقشی دو، سه یا چهار حرفی است، که حروف اضافه، ضمائر، حروف ربط و افعال کمکی را شامل می‌شوند. در زبان انگلیسی این واژه‌ها عبارتند از: *the, that, of, it, is, in, I, he, and, a, to, was*. این دوازده واژه در واقع پربسامدترین واژه‌های زبان انگلیسی‌اند. دومین عادت زبانی استفاده از واژه‌هایی است که با واکه آغاز می‌شوند. تحلیل گران کیوسام ۹ نوع آزمون را مشخص کرده‌اند که می‌توان بر داده‌ها اعمال کرد تا عادات زبانی نویسنده شناسایی شوند. متداول‌ترین این آزمون‌ها استفاده از واژه‌های دو یا سه حرفی، واژه‌های واکه‌آغازین و ترکیب دو مورد پیشین (مجموع تعداد واژه‌های دو یا سه حرفی و تعداد واژه‌های واکه‌آغازین) است. مورد آخر بعنوان مفیدترین شاخص برای تشخیص یک دستی متن تعیین شده است.

با انتخاب یکی از این عادات زبانی و مقایسه آن با میانگین طول جمله می‌توان تحلیل کیوسام را ادامه داد. در این مورد نیز تحلیلی سه مرحله‌ای مورد نیاز است. در گام اول باید موارد وقوع عادت موردنظر را محاسبه کرد و میانگین آن را در جمله به‌دست آورد. در گام دوم میزان انحراف هر جمله را از میانگین به‌دست آمده محاسبه کرد و در گام سوم جمع انباشتی انجام داد و نتایج را روی نموداری نشان داد. در این مرحله با مقایسه دو نمودار و منطبق کردن آن‌ها بر یکدیگر^۱ می‌توان به این نتیجه رسید که متن توسط شخص واحدی نوشته شده است یا بیش از یک شخص آن را نوشته‌اند. اگر دو نمودار با یکدیگر مطابقت یابند شخص واحدی آن را نوشته است اما اگر با یکدیگر مطابقت نیابند توسط بیش از یک فرد نوشته شده است. به ادعای فارینگدان، برای اثبات این مطلب، اگر در وسط یک متن متنی را از نویسنده دیگری اضافه کنیم، در نمودار مربوط به این بخش نوعی ناهماهنگی به چشم می‌خورد که نشان می‌دهد کل متن توسط یک فرد نوشته نشده است. در نمونه‌هایی که فارینگدان از تحلیل نامه‌ای از دی اچ لارنس^۲ انجام می‌دهد این مطابقت به خوبی مشهود است. این مطابقت نشانگر عادت ثابتی است که در تمام طول متن به چشم می‌خورد.

^۱ برای تطبیق این دو نمودار بر یکدیگر می‌توان از نمودار واحدی استفاده کرد که دو محور عمودی *y* در سمت راست و چپ داشته باشد و محور افقی

x نشان‌گر شماره جملات باشد.

^۲ D. H. Lawrence

عرض نشان داده می‌شود. نمایش هندسی حاصل "جمله‌نگار" نامیده می‌شود (بازرگان، ۱۳۵۵: ۴۷-۴۶). بازرگان (همان.) به جمله‌نگاری دیباچه گیلستان می‌پردازد و نشان می‌دهد شاخص‌های اصلی جمله‌نگار در دیگر بخش‌های کتاب گیلستان کم‌وبیش مانند دیباچه است، اما جمله‌نگار مقدمه گیلستان، که نوشته محمدعلی فروغی است، چهره کاملاً متفاوتی را نشان می‌دهد.

البته بازرگان با اشاره به تنوع جمله‌سازی در زبان‌ها و در میان ملل مختلف به این نکته اشاره می‌کند که جمله‌سازی تابع موضوع نیز می‌شود: «مثلاً در بحث‌های فلسفی و حقوقی که بیشتر از الفاظ و عبارات استمداد می‌شود و با صغری و کبری چیدن استنتاج می‌نمایند، جمله‌ها طولانی در می‌آید... در داستانسرای مخصوصاً برای کودکان کلمات شمرده و جمله‌های کوتاه انتخاب می‌گردد» (همان: ۸۴-۸۳).

به این ترتیب ملاحظه می‌گردد که محققان زیادی سعی داشته‌اند با تکیه بر شیوه‌های آماری راهی برای تشخیص نویسنده متن بیابند. در قسمت بعد به تفصیل به بررسی و ارزیابی یکی از این شیوه‌های آماری یعنی تحلیل کیوسام می‌پردازیم.

۱.۳. تحلیل کیوسام^۱

این شیوه تحلیل که توسط اندرو کونین مورتون^۲ معرفی شده، از سال ۱۹۹۰ برای تشخیص نویسنده متن و تحقیقات حقوقی به کار رفته است. شیوه مذکور بر این فرض مبتنی است که هر سخنگو، هنگام استفاده از زبان، عادات منحصر به فردی دارد و این عادات به لحاظ آماری الگوهای قابل تشخیصی را در متن تشکیل می‌دهند (فارینگدان^۳، ۲۰۰۵). فارینگدان (همان.) در مقاله‌ای با عنوان «چگونه می‌توان یک کارآگاه ادبی بود؟» چگونگی انجام تحلیل کیوسام را شرح می‌دهد. به ادعای فارینگدان این تحلیل، که بر اساس اندازه‌گیری و مقایسه کار می‌کند، می‌تواند نشان دهد یک متن را یک شخص یا بیش از یک شخص نوشته است. فارینگدان مدعی است این شیوه تحلیل را می‌توان بر تمام انواع متون (گفتاری یا نوشتاری) بدون در نظر گرفتن زمان تالیف یا نوع آن‌ها اعمال کرد.

اولین گام در انجام این تحلیل انتخاب یک متن و محاسبه میانگین طول جملات آن است که از طریق شمارش تمام واژه‌های متن و تقسیم حاصل آن بر تعداد جملات متن به دست می‌آید. دومین گام محاسبه اختلاف طول هر جمله با میانگین طول جملات است. برخی از جملات کوتاه‌تر از میانگین و برخی بلندترند. نتیجه این محاسبات به صورت اعداد مثبت (بزرگتر از میانگین) و منفی (کوچکتر از میانگین) نشان داده می‌شود. سومین گام این است که ارقام به دست آمده از گام دوم را به ترتیب با یکدیگر جمع جبری نماییم (به این نوع جمع، جمع انباشتی^۴ گفته می‌شود) و نتیجه جمع هر دو عدد متوالی را در جدولی ثبت کنیم.

^۱ واژه کیوسام در برابر اصطلاح Cusum / Qsum به کار رفته است که خود مخفف اصطلاح Cumulative sum با "جمع انباشتی" است.

^۲ A. Q. Morton

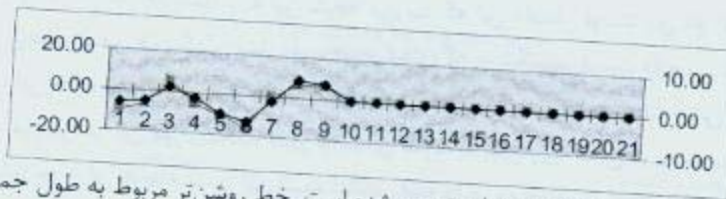
^۳ J. Farrington

^۴ cumulative sum

اعتقاد او برای اعمال این شیوه بر زبان فارسی باید پر بسامدترین واژه‌های نقشی زبان را مشخص کرد و بر اساس آن‌ها تحلیل را انجام داد.

۲.۳. تحلیل متون فارسی با استفاده از کیوسام

تحلیل اولیهٔ متنی، که دارای حدود ۳۵ هزار واژه است، پر بسامدترین واژه‌های زبان فارسی را به ترتیب زیر نشان می‌دهد: «و»، که، را، از، به، با، در، تا، این، آن». هر چند لازم است برای تعیین پر بسامدترین واژه‌های زبان به متون مختلف مراجعه کرد، اما به هر حال به نظر می‌رسد در زبان فارسی نیز واژه‌های نقشی پر بسامد عمدتاً دو حرفی و گاه سه حرفی‌اند. با در نظر گرفتن این واژه‌ها به عنوان کلید تشخیص نویسنده، این شیوهٔ تحلیل را بر متون فارسی اعمال می‌کنیم. البته باید توجه داشت که در زبان فارسی ضمائر به شکل متصل نیز به کار می‌روند و ممکن است این عناصر بسامد بالایی داشته باشند. اما از آنجا که این شیوهٔ تحلیل مبتنی بر واژه‌های نقشی است نه تکواژهای نقشی، فعلاً این عناصر را وارد تحلیل نمی‌کنیم. برای اطمینان از درستی برنامهٔ ساده‌ای که به همین منظور توسط نگارنده تهیه شده است، داده‌های مربوط به نمودار ۱ فارینگدان به این برنامه داده شد، که نتیجه آن به صورت زیر درآمد:



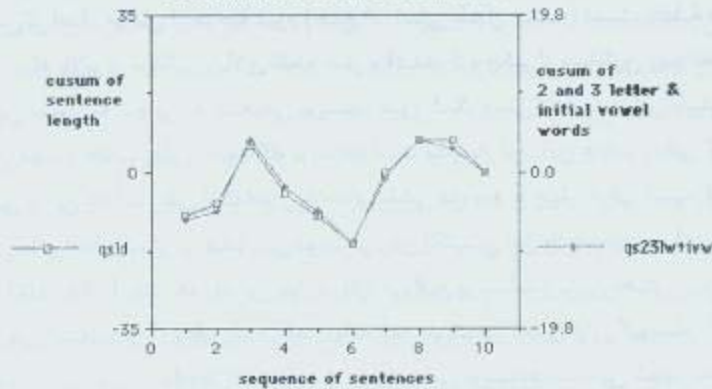
نمودار ۱ فارینگدان (۲۰۰۵) که توسط نگارنده دوباره ترسیم شده است. خط روشن‌تر مربوط به طول جمله و خط تیره‌تر مربوط به واژه‌های دو سه حرفی + واکه‌آغازین است.

در نمونهٔ زیر تحلیل کیوسام بر متنی کوتاه از همین مقاله اعمال شده است (متن شامل ۸ جمله است که با شماره مشخص شده‌اند)

۳.۳. تحلیل بخشی از مقاله حاضر

- ۱- این شیوهٔ تحلیل که توسط ا.ک. مورتون معرفی شده است از سال ۱۹۹۰ برای تشخیص نویسندهٔ متن و تحقیقات حقوقی به کار رفته است. ۲- شیوهٔ مذکور بر این فرض مبتنی است که هر سخنگو، هنگام استفاده از زبان، عادات منحصر به فردی دارد و این عادات به لحاظ آماری الگوهای قابل تشخیصی را در متن تشکیل می‌دهند. ۳- اولین گام در انجام این تحلیل انتخاب یک متن و محاسبهٔ میانگین طول جملات است، که از طریق شمارش تمام واژه‌های متن و تقسیم آن بر تعداد جملات متن به دست می‌آید. ۴- دومین گام محاسبهٔ اختلاف طول هر جمله با میانگین طول جملات است، برخی از جملات کوتاه‌تر از میانگین و برخی بلندترند. ۵- نتیجهٔ این محاسبات به صورت اعداد مثبت و منفی نشان داده می‌شود. ۶- سومین گام این است که ارقام

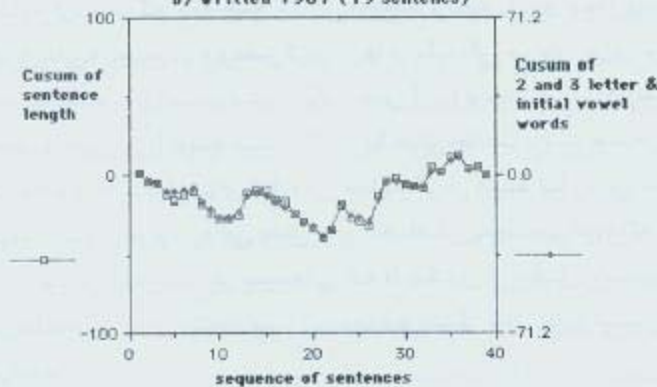
The sentence length and habit graphs superimposed to form a cusum chart



نمودار ۱- تحلیل نامه‌ای از دی ایچ لارنس (فارینگدان، ۲۰۰۵).

در نمونه‌های دیگر دو نامه از دی ایچ لارنس (۱۹۱۳ و ۱۹۰۷) به دنبال یکدیگر تحلیل می‌شوند و بازهم الگوی ثابتی را نشان می‌دهند (فارینگدان، ۲۰۰۵).

Two letters by D.H. Lawrence
a) written 1913 (20 sentences)
b) written 1907 (19 sentences)



نمودار ۲- تحلیل دو نامه از دی ایچ لارنس (همان).

اما در نمودار دیگری که متنی از آلدوس هاکسلی^۱ (بعد از جمله ۲۰) به نامه‌های لارنس اضافه می‌شود این تطابق به هم می‌خورد (همان).

^۱ A. Huxley

09-11-01

THIS IS NEXT
TAKE PENACILIN NOW
DEATH TO AMERICA
DEATH TO ISRAEL
ALLAH IS GREAT

ترجمه: (۰۱-۱۱-۰۹ این مورد بعدی است. همین الآن پنی سیلین مصرف کنید. مرگ بر امریکا. مرگ بر اسرائیل. الله اکبر.)

۴.۴. گزارش اف. بی. آی^۱ در مورد ویژگی‌های زبانی نامه‌ها

اف.بی. آی در تاریخ ۹ نوامبر ۲۰۰۱ با ارائه گزارشی در مورد ویژگی‌های زبانی نامه‌ها از مردم خواست تا چنانچه فردی را با این مشخصات می‌شناسند، معرفی نمایند. ویژگی‌های مذکور عبارت‌اند از:

۱. نویسنده هر سه نامه یک نفر است.
 ۲. در نوشتن تاریخ به جای خط مایل از خط تیره استفاده شده است.
 ۳. عدد یک به جای خط راست که مرسوم‌تر است به صورت 1 نوشته شده است.
 ۴. can not به جای cannot که بیش‌تر استفاده می‌شود به کاررفته است.
 ۵. نویسنده همه نامه‌ها را با حروف بزرگ نوشته است، اما حرف اول اولین واژه هر جمله و اسم‌های خاص اندکی بزرگ‌تر نوشته شده است.
 ۶. نام و آدرس روی نامه‌ها از چپ به راست شیب دارد.
- گفتنی است که اف.بی. آی علاوه بر ارائه و بررسی مشخصات زبانی این نامه‌ها به ارائه دیگر مشخصات مربوط به این پرونده، مثل ویژگی‌های روانشناختی و رفتاری مظنون، نوع میکروپ به کاررفته در نامه‌ها و نحوه آماده‌سازی آن‌ها نیز پرداخته است. بنابراین بررسی زبان‌شناختی این نامه‌ها تنها یکی از وجوه تحقیقات گسترده آنهاست. بر همین اساس نتایج حاصل از این تحلیل‌ها را باید مکمل و مؤید سایر نتایج در نظر گرفت.

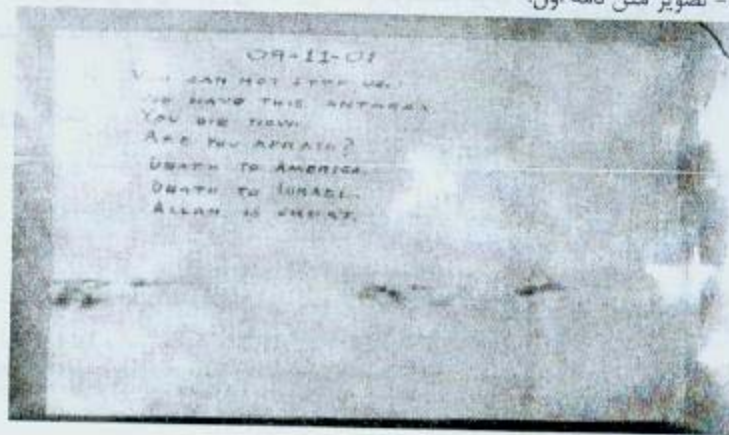
۵.۴. تحلیل السن، زبان‌شناس قانونی

السن، زبان‌شناس قانونی، نیز با بررسی نامه‌های مذکور به موارد زیر اشاره می‌کند:

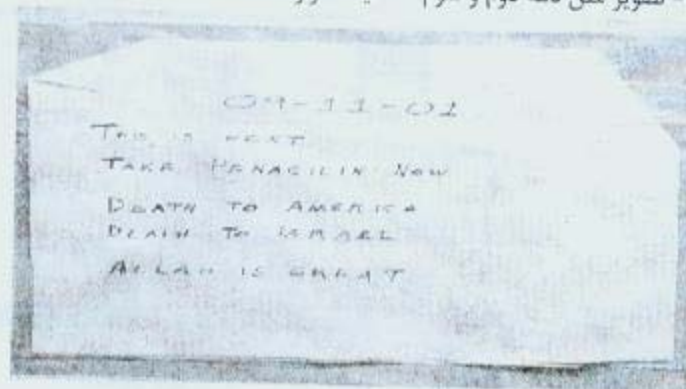
۱. شیوه نوشتن تاریخ به صورت سال+روز+ماه (۰۱-۱۱-۰۹) است. این شیوه معمول نوشتن تاریخ در آمریکا است. در انگلیس صورت روز+ماه+سال و در چین صورت سال+ماه+روز متداول است.
 ۲. در کشورهای عربی شیوه آمریکایی نوشتن تاریخ مرسوم است.
 ۳. بین اعداد تاریخ از خط تیره استفاده شده است درحالی‌که معمولاً از / استفاده می‌شود.
- او در تحلیل متن نامه‌ها به نکات زیر اشاره می‌کند:
۱. نامه‌ها زبانی موجز و دقیق دارند.

¹ available at: <http://www.fbi.gov/anthrax/amerithrax.htm>

۱- تصویر متن نامه اول:



۲- تصویر متن نامه دوم و سوم که عیناً تکرار شده است:



متن نامه ها:

09-11-01
YOU CAN NOT STOP US.
WE HAVE THIS ANTHRAX.
YOU DIE NOW.
ARE YOU AFRAID?
DEATH TO AMERICA.
DEATH TO ISRAEL.
ALLAH IS GREAT.

ترجمه: (۰۹-۱۱-۰۱ نمی‌توانید ما را متوقف کنید. ما این سیاه زخم را داریم. شما الآن می‌میرید. ترسیده‌اید؟ مرگ بر امریکا. مرگ بر اسرائیل. الله اکبر.)

تشخیص مؤلف متون ادبی و قانونی....

این نامه آیا عرب است یا سخنگوی بومی زبان انگلیسی؟ یا این که آیا نویسنده عرب نیست اما فردی است که از الفبای عربی برای نوشتن استفاده می کند (فارسی، اردو، پاکستانی یا پنجابی)؟

۲.۴. مشخصات آدرس روی پاکت

۱- شیوه نوشتن آدرس روی پاکت و استفاده از حروف بزرگ این تصور را ایجاد می کند که نویسنده نمی تواند به راحتی انگلیسی را بنویسد یا یک دانش آموز کلاس چهارم (طبق آدرس روی پاکت) این نامه را نوشته است.

۲- شیب آدرس، که به سمت راست متمایل شده است، این تصور را ایجاد می کند که نویسنده این متن عادت به نوشتن از چپ به راست را ندارد

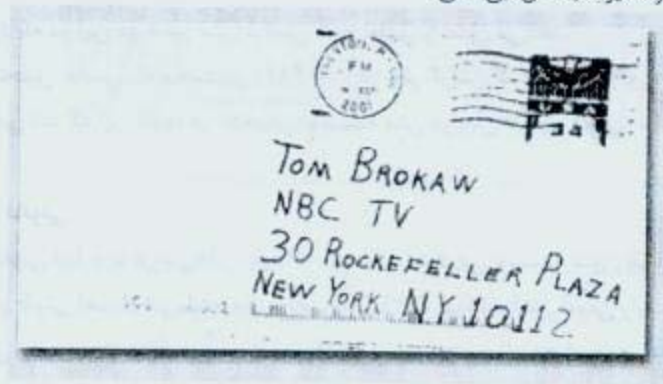
۳- علائم نقطه گذاری به ندرت به کار رفته اند، معمولاً اعراب هنگام استفاده از زبان انگلیسی از علائم کمتری استفاده می کنند.

۴- آدرس دبستان و کدپستی جعلی است. جعلی بودن آدرس می تواند نشان گر این نکته باشد که همه چیز خلاف واقع است (السن، ۲۰۰۴: ۱۰۳).

۳.۴. متن نامه ها



۱- نامه به تام براکو مدیر شبکه ان.بی.سی.



۲- نامه به سناتور داشل.



۳- نامه به سناتور لیهای.



السن (۲۰۰۴: ۱۰۱) در تحلیل خود با در نظر گرفتن حوادث پس از ۱۱ سپتامبر و فضای ضد عرب حاکم و احتمال دست داشتن شبکه القاعده در این تهدیدها، اولین سؤال خود را این گونه مطرح می‌کند که نویسنده



نمودار ۴- تحلیل هشت جمله از مقاله حاضر.

همان‌گونه که ملاحظه می‌شود جز در مورد جمله شماره ۶ بقیه نقاط با یکدیگر مطابقت می‌یابند. اما در نمودار زیر که ۴ جمله از "کرازی" (۱۳۷۳) به وسط متن فوق اضافه شده است این تطابق به هم می‌خورد:



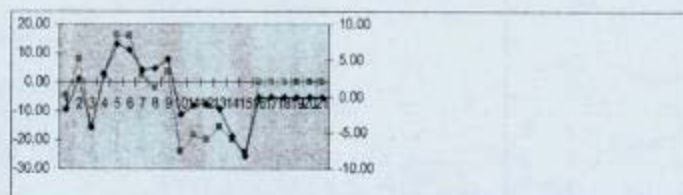
نمودار ۵- تحلیل متن ترکیبی

همان‌گونه که ملاحظه می‌شود، نمودار دوم نسبت به نمودار اول آشفتگی بیشتری را نشان می‌دهد. اما در مواردی تبیین این که دو خط یکدیگر را دنبال می‌کنند یا نه، تا حد زیادی فردی و سلیقه‌ای به نظر می‌رسد. مشخص نیست که دو خط تا کجا یکدیگر را دنبال می‌کنند یا در کجا با یکدیگر مطابقت نمی‌یابند. نمودار زیر تحلیل متنی از کرازی (۱۳۷۳: ۳۶۸-۳۶۷) شامل ۱۰ جمله را نشان می‌دهد:



نمودار ۶- تحلیل متنی از کرازی.

در نمودار زیر به انتهای ۱۲ جمله از کرازی، ۴ جمله از مشکوة الدینی (۱۳۷۳: ۴۰۳) اضافه شده است:



نمودار ۷- متن ترکیبی کرازی و مشکوة الدینی.

تشخیص مؤلف متون ادبی و قانونی...

به دست آمده از گام دوم را به ترتیب یا یکدیگر جمع جبری نماییم و نتیجه جمع هر دو عدد متوالی را در جدولی ثبت کنیم. ۷- این اعداد نهایی را می توان روی نموداری نمایش داد که محور افقی آن نشان گر تعداد جملات و محور عمودی آن نشان گر اعداد حاصل از مرحله سوم (جمع انباشتی طول جمله) است. ۸- نقطه صفر نشانگر میانگین است. اعداد بالاتر از میانگین، بالای صفر و اعداد کوچک تر از میانگین زیر نقطه صفر قرار می گیرند.

جدول ۱- محاسبه کیوسام طول جملات

شماره جمله	طول جمله	اختلاف طول جمله با میانگین طول جملات	Qsum
۱	۲۳/۰۰	-۲/۶۳	-۲/۶۳
۲	۳۳/۰۰	۷/۳۸	۴/۷۵
۳	۳۲/۰۰	۶/۳۸	۱۱/۱۳
۴	۲۱/۰۰	-۴/۶۳	۶/۵۰
۵	۱۲/۰۰	-۱۳/۶۳	-۷/۱۳
۶	۳۲/۰۰	۶/۳۸	-۰/۷۵
۷	۳۱/۰۰	۵/۳۸	۴/۶۳
۸	۲۱/۰۰	-۴/۶۳	۰

میانگین: ۲۵/۶۳

کیوسام فوق به این صورت محاسبه می شود که $۲/۶۳$ با $۷/۳۸$ جمع می شود و عدد حاصل $۴/۷۵$ ثبت می شود. عدد مذکور با $۶/۳۸$ جمع می شود و نتیجه آن $۱۱/۱۳$ ثبت می شود و الخ.

جدول ۲- محاسبه کیوسام واژه های نقشی

شماره جمله	واژه های نقشی	واژه های نقشی - میانگین	Qsum
۱	۶/۰۰	-۰/۳۸	-۰/۳۸
۲	۹/۰۰	۲/۶۳	۲/۲۵
۳	۹/۰۰	۲/۶۳	۴/۸۸
۴	۴/۰۰	-۲/۳۸	۲/۵۰
۵	۳/۰۰	-۲/۳۸	-۰/۸۸
۶	۱۰/۰۰	۳/۶۳	۲/۷۵
۷	۷/۰۰	۰/۶۳	۳/۳۸
۸	۳/۰۰	-۲/۳۸	۰

میانگین: ۶/۳۸

تشخیص مؤلف متون ادبی و قانونی....

محمدیان، امیر (۱۳۸۶). بررسی کارایی تحلیل کیوسام در زبان فارسی. پایان نامه کارشناسی ارشد. زبانشناسی. دانشگاه فردوسی مشهد.

مشکوة الدینی، مهدی (۱۳۷۳). "قاعده سازی زایا و یادگیری مهارت". مجموعه مقالات دومین کنفرانس زبانشناسی نظری و کاربردی. تهران: دانشگاه علامه طباطبائی. صص: ۴۲۱-۴۰۳.

Farrington, Jill (2004). *Funeral Elegy. Not written by W. S. After all*. Studies in Bibliography. Virginia University publications.

_____ (2005). *How to be a literary detective: Authorship attribution*.

<http://members.aol.com/qsums/QsumIntroduction.html>

Lindskoog, Kathryn (1996). *Analysing for Authorship with the Cusum Technique*. The Lewis Legacy-Issue 69.

McMenamin, Gerald R. (1993). *Forensic Stylistics*, Amsterdam: Elsevier. Available online at: Google ebooks.

Olsson, John (2004). *Forensic linguistics, an introduction to language, crim and law*. London, New York: Continuum.

۲. زبان نامه‌ها زبان شخصی تحصیل کرده است که عادت به نوشتن دارد.
۳. غلط املائی در PENACILIN و عبارت پیچین^۱ YOU DIE NOW احتمالاً نکات انحرافی‌اند. السن از بررسی و تحلیل نامه‌ها به نتایج زیر می‌رسد:
۱. نویسنده نامه می‌خواهد این تصور را ایجاد کند که یک مسلمان افراطی جوان تحصیل‌نکرده است، که به سختی قادر به نوشتن انگلیسی است.
 ۲. اما همه چیز برخلاف این است. نویسنده احتمالاً فردی است آمریکایی، تحصیل‌کرده، با احساسات ضدمسلمانی. هیچ شاهی وجود ندارد که نویسنده مسلمان است، هرچند او می‌خواهد این تصور را ایجاد کند (السن، ۲۰۰۴: ۱۰۵-۱۰۴).
- با توجه به این که نویسنده این نامه‌ها تاکنون شناخته نشده است، در مورد درستی این نتیجه‌گیری نمی‌توان قضاوت کرد. باین حال، موارد متعدد دیگری در سبک‌شناسی قانونی گزارش شده‌اند که به کمک تحلیل‌های زبان‌شناسی به تشخیص نویسنده متن منجر شده‌اند (مک منامین، ۱۹۹۳).

۵. بحث و نتیجه‌گیری

مقاله حاضر نشان می‌دهد چگونه از تحلیل‌های آماری و زبانی برای تعیین و تشخیص نویسنده متن استفاده می‌شود. لازم به توضیح است که نمونه‌های ارائه شده در این مقاله به صورت نمونه‌هایی اولیه و فقط به منظور آشنایی خوانندگان با این شیوه‌های تحلیل ارائه شده‌اند و نتیجه‌گیری قطعی در مورد کارایی یا عدم کارایی آن‌ها نیاز به بررسی‌های گسترده‌تری دارد.

البته باید توجه داشت که حتی در صورتی هم که این شیوه‌های تحلیل با داده‌های زبان فارسی مطابقت یابند، قطعیتی ندارند و تنها می‌توانند به عنوان راهنما برای تعیین سمت و سوی تحقیقات به کار روند. در هر حال، هدف اصلی ارائه این مقاله تأیید و اعمال شیوه‌های تحلیل مذکور بر زبان فارسی نبوده است، بلکه هدف نشان دادن نگرشی متفاوت بوده است به مسئله تشخیص مؤلف متون ادبی و قانونی و تأکید بر لزوم فراهم آوردن بستری که زبان‌شناسان بتوانند با تکیه بر دانش زبان‌شناسی خود به حل مشکلاتی که در سایر حوزه‌های علمی دیده می‌شود کمک کنند.

منابع

- بازرگان، مهدی (۱۳۵۵). *حمله شناسی قرآن*. تهران: بعثت.
- بدره ای، فریدون (۱۳۷۴). مقدمه ای بر بررسی انتقادی رباعیات خیام. در: کریستین سن، آرتور. ۱۳۷۴. *بررسی انتقادی رباعیات خیام*. ترجمه فریدون بدره‌ای. تهران: توس.
- شمیسا، سیروس (۱۳۷۲). *کلیات سبک شناسی*. تهران: فردوس.
- کریستین سن، آرتور (۱۳۷۴). *بررسی انتقادی رباعیات خیام*. ترجمه فریدون بدره‌ای. تهران: توس.
- کزازی، میرجلال‌الدین (۱۳۷۳). "پویه پایای پارسی". *مجموعه مقالات دومین کنفرانس زبان‌شناسی نظری و کاربردی*. تهران: دانشگاه علامه طباطبائی. صص: ۲۶۷-۲۶۷.

۱ گونه‌ای از انگلیسی که ساختارهای دستوری و واژگان محدودی دارد.

ملاحظه می‌شود که در این مورد نیز متن ترکیبی عدم تطابق را نشان می‌دهد. با این حال، در تحلیل جامعی که محمدیان (۱۳۸۶)، با در نظر گرفتن کلیدهای گوناگون، از این شیوه ارائه می‌دهد، مشخص می‌کند که این شیوه در تشخیص نویسنده متن در زبان فارسی کارایی ندارد.

۴. سبک‌شناسی قانونی^۱

سبک‌شناسی قانونی اشاره به بررسی‌هایی دارد که هدفشان تشخیص نویسنده متون ناشناسی است که در حوزه‌های حقوقی و قانونی اهمیت می‌یابند. مک منامین^۲ (۱۹۹۳) سبک‌شناسی قانونی را به صورت اعمال اصول سبک‌شناسی زبانی بر متون قانونی^۳ تعریف می‌کند. متونی چون نامه‌های تهدیدآمیز، پیام‌های خودکشی، وصیت‌نامه‌ها و... می‌توانند وارد این تحلیل‌ها شوند. در این نوع تحلیل‌ها معمولاً صورت و محتوای متن از جنبه‌های گوناگون مورد بررسی قرار می‌گیرند. در این قسمت به ارائه نمونه‌ای در این زمینه می‌پردازیم.

۴.۱. تحلیل متن قانونی (نامه‌های تهدیدآمیز)

پس از حادثه یازده سپتامبر در آمریکا نامه‌های مرگ‌باری که حاوی میکروب سیاه‌زخم بودند برای تنی چند از افراد برجسته آمریکایی ارسال شدند که باعث مرگ چند نفر و ایجاد فضایی رعب‌آور برای دیگران شد. پس از دریافت نامه‌ها تلاش برای کشف نویسنده آن‌ها آغاز شد. سازمان اف. بی. آی با انتشار تصاویر و گزارشی از ویژگی‌های نامه‌ها از مردم خواست چنان‌چه کسی را که دارای این مشخصات زبانی است می‌شناسد، معرفی کنند. السن (۲۰۰۴: ۱۰۵-۱۰۱) که زبانشناسی قانونی است نیز به تحلیل این نامه‌ها پرداخت و نتایج تحلیل خود را ارائه کرد. براین اساس چند نفر مورد اتهام قرار گرفتند. با این حال نویسنده این نامه‌ها هرگز شناخته نشد و پرونده مذکور هنوز باز مانده است. در این قسمت به بررسی ویژگی‌های این نامه‌ها می‌پردازیم و نشان می‌دهیم کدام مؤلفه‌ها مورد توجه تحلیل‌گران قرار گرفته‌اند. تصویر پاکت^۴ سه نامه متفاوت که دارای میکروب سیاه‌زخم بودند در زیر دیده می‌شود:

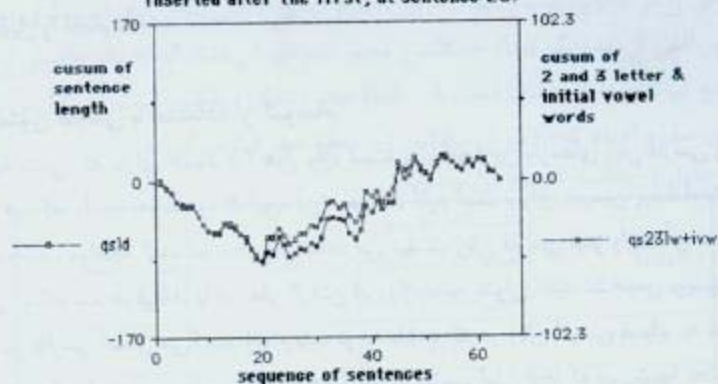
^۱ forensic stylistics

^۲ G. McMenamin

^۳ forensic texts

^۴ تصویر نامه‌ها از آدرس زیر گرفته شده است:

Combined analysis: 2 letters by Lawrence
(39 sents.) with 25 sents. by Aldous Huxley
inserted after the first, at sentence 20.



نمودار ۳- تحلیل متنی ترکیبی از هاگسلی و لارنس (همان).

به ادعای فارینگدان (۲۰۰۴) از این شیوه می‌توان برای تأیید یا رد انتساب نوشته‌ای به یک نویسنده استفاده کرد. به عنوان مثال با تحلیل داستان کوتاهی، که به دی اچ لارنس نسبت داده می‌شود، و افزودن آن به دیگر آثار لارنس، فارینگدان به این نتیجه می‌رسد که این داستان نوشته دی. اچ. لارنس نیست. در جایی دیگر بر همین اساس فارینگدان نشان می‌دهد مرثیه‌ای که به شکسپیر نسبت داده می‌شود، نوشته وی نیست. با این حال، مشخص نیست چگونه انتخاب این عادات زبانی می‌تواند به تشخیص نویسنده متن کمک کنند. مورتون در این مورد شخصا هیچ توضیحی ندارد. به بیان وی: «این شیوه درست کار می‌کند اما نمی‌دانم چرا» او که به درستی این شیوه اعتقاد راسخ دارد، معتقد است تحلیل مذکور همیشه موفق عمل می‌کند، حتی در مورد متون کوتاه (السن، ۲۰۰۴: ۱۶).

فارینگدان (۲۰۰۵: مکاتبات شخصی) نیز معتقد است این شیوه موفق بوده و موفق خواهد بود مگر این که استثنایی بر آن پیدا شود. وی در توضیح این مطلب اشاره می‌کند که در انگلیسی در ۲۵ درصد مواقع به هنگام صحبت کردن تنها ۱۲ واژه پربسامد را تکرار می‌کنیم، به همین دلیل چگونگی توزیع این واژه‌های پربسامد می‌تواند کلید تشخیص نویسنده باشد. اما به نظر نمی‌رسد این توضیح قانع‌کننده باشد و منطبق نهفته در این شیوه تحلیل را مشخص کند. از سوی دیگر، در مورد واژه‌های واکه‌آغازین چه توضیحی می‌توان داد؟ چرا استفاده از این نوع واژه‌ها باید به تشخیص نویسنده کمک کند؟ در زبان‌هایی که واژه‌ها با واکه آغاز نمی‌شوند این کلید تشخیص چه وضعیتی می‌یابد؟ در هر حال با وجود سؤالات و انتقاداتی که در این مورد مطرح است، لیندزکوک^۱ (۱۹۹۶) بیان می‌کند که از سال ۱۹۹۰ به این سو از این تکنیک در دادگاه‌ها برای تحلیل نامه‌های خودکشی، اعترافات، شهادت شهود، تلفن‌های ناشناس، مکالمات ضبط شده و سرقت‌های ادبی استفاده شده است.

تحلیل‌های فارینگدان بر متون انگلیسی انجام گرفته است، اما فارینگدان (۲۰۰۵: مکاتبات شخصی) بیان می‌دارد که این تحلیل بر زبان‌های فرانسه، آلمانی، لاتین، یونانی باستان و... نیز اعمال شده است. به

^۱ K. Lindskoog

ملاحظه می‌شود که در این مورد نیز متن ترکیبی عدم تطابق را نشان می‌دهد. با این حال، در تحلیل جامعی که محمدیان (۱۳۸۶)، با در نظر گرفتن کلیدهای گوناگون، از این شیوه ارائه می‌دهد، مشخص می‌کند که این شیوه در تشخیص نویسنده متن در زبان فارسی کارایی ندارد.

۴. سبک‌شناسی قانونی^۱

سبک‌شناسی قانونی اشاره به بررسی‌هایی دارد که هدفشان تشخیص نویسنده متون ناشناسی است که در حوزه‌های حقوقی و قانونی اهمیت می‌یابند. مک منامین^۲ (۱۹۹۳) سبک‌شناسی قانونی را به صورت اعمال اصول سبک‌شناسی زبانی بر متون قانونی^۳ تعریف می‌کند. متونی چون نامه‌های تهدیدآمیز، پیام‌های خودکشی، وصیت‌نامه‌ها و... می‌توانند وارد این تحلیل‌ها شوند. در این نوع تحلیل‌ها معمولاً صورت و محتوای متن از جنبه‌های گوناگون مورد بررسی قرار می‌گیرند. در این قسمت به ارائه نمونه‌ای در این زمینه می‌پردازیم.

۴.۱. تحلیل متن قانونی (نامه‌های تهدیدآمیز)

پس از حادثه یازده سپتامبر در آمریکا نامه‌های مرگ‌باری که حاوی میکروب سیاه‌زخم بودند برای تنی چند از افراد برجسته آمریکایی ارسال شدند که باعث مرگ چند نفر و ایجاد فضایی رعب‌آور برای دیگران شد. پس از دریافت نامه‌ها تلاش برای کشف نویسنده آن‌ها آغاز شد. سازمان اف. بی. آی با انتشار تصاویر و گزارشی از ویژگی‌های نامه‌ها از مردم خواست چنان‌چه کسی را که دارای این مشخصات زبانی است می‌شناسد، معرفی کنند. السن (۲۰۰۴: ۱۰۵-۱۰۱) که زبانشناسی قانونی است نیز به تحلیل این نامه‌ها پرداخت و نتایج تحلیل خود را ارائه کرد. براین اساس چند نفر مورد اتهام قرار گرفتند. با این حال نویسنده این نامه‌ها هرگز شناخته نشد و پرونده مذکور هنوز باز مانده است. در این قسمت به بررسی ویژگی‌های این نامه‌ها می‌پردازیم و نشان می‌دهیم کدام مؤلفه‌ها مورد توجه تحلیل‌گران قرار گرفته‌اند. تصویر پاکت^۴ سه نامه متفاوت که دارای میکروب سیاه‌زخم بودند در زیر دیده می‌شود:

^۱ forensic stylistics

^۲ G. McMenamin

^۳ forensic texts

^۴ تصویر نامه‌ها از آدرس زیر گرفته شده است:

<http://www.fbi.gov/pressrel/pressrel01/102301.htm>; 2001 Press Release.