

تحلیل محتوای مقالات علوم پزشکی بر اساس مطالعه پیکره زبانی*

مریم سلامی^۱، زهرا سادات جلالی^۲، مریم پاکدامن نائینی^۳، محمد علائی آرانی^۴

مقاله پژوهشی

چکیده

مقدمه: تحلیل محتوا برای مقایسه یک متن با متن دیگر، آثار یک نویسنده، با نویسندگان دیگر و با همان نویسنده، مقایسه متن‌های نوشته شده به یک زبان طبیعی با متون زبان دیگر و بررسی متون یک حوزه تخصصی به کار می‌رود، هدف پژوهش حاضر تحلیل محتوای مقالات علوم پزشکی و استخراج ساختار جملات در آن به منظور استفاده پژوهشگران بوده است.

روش بررسی: این مطالعه از دسته مطالعات تحلیل محتوا بر اساس ساختار دستوری است که به بررسی میزان کاربرد زنجیره‌های واژگانی چهار لغتی در مقالات پژوهشی در زمینه پزشکی پرداخته است. به همین منظور در این مقاله مجموعه‌ای بیش از دو میلیون و چهارصد هزار لغت از مقاله‌های بازبایی شده در ۳۳ حوزه پزشکی گردآوری و معیار بسامد بیست بار تکرار یک زنجیره واژگانی در یک میلیون لغت و همچنین تکرار آن در پنج متن متفاوت در نظر گرفته شد. برای به دست آوردن زنجیره‌ها از نرم افزارهای Word smith tools 5, Antconc3.23, ABBYY Fine Reader to professional edition استفاده شد. زنجیره‌ها با استفاده از طبقه‌بندی‌های موجود و براساس ساختار و عملکرد در متن دسته‌بندی شده‌اند.

یافته‌ها: در نتیجه این تحقیق ۱۰۲ زنجیره واژگانی پر کاربرد با ساختارها و عملکردهای متفاوت در حوزه های پزشکی شناسایی شدند. بیشتر نویسندگان مقالات تحقیقاتی پزشکی از زنجیره‌های واژگانی از نوع متن محور استفاده می‌کنند.

نتیجه‌گیری: این مطالعه دیدگاه‌های جدیدی در زمینه فهم بهتر مقالات پژوهشی پزشکی مربوط به زمینه‌ای خاص و نیز تحقیقات بیشتر در زمینه مطالعات پیکره ای در نوشتار دانشگاهی و انگلیسی برای اهداف ویژه فراهم می‌آورد. همچنین افق جدیدی را در حوزه تحلیل محتوای متون و سنجش آماری متون می‌گشاید.

واژه‌های کلیدی: تحلیل، زبان‌شناسی کامپیوتری، مقالات مجلات، پیکره زبانی

پذیرش مقاله: ۹۴/۰۳/۲۰

اصلاح نهایی: ۹۴/۰۳/۱۹

دریافت مقاله: ۹۳/۰۷/۰۷

ارجاع: سلامی مریم، جلالی زهرا سادات، پاکدامن نائینی مریم، علائی آرانی محمد. تحلیل محتوای مقالات علوم پزشکی بر اساس مطالعه پیکره زبانی. مدیریت اطلاعات سلامت ۱۳۹۴؛ ۱۲(۵): ۵۹۵-۶۰۷

*-این مقاله حاصل تحقیق مستقل می‌باشد.

۱- استادیار علم اطلاعات و دانش‌شناسی، دانشگاه پیام نور، مشهد، ایران

۲- کارشناس ارشد آموزش زبان انگلیسی، مدرس دانشگاه کاشان، کاشان، ایران

۳- دانشجوی دکتری، علم اطلاعات و دانش‌شناسی، دانشگاه پیام نور و کارشناس ارشد، پژوهشگاه بین‌المللی زلزله‌شناسی و مهندسی زلزله، تهران، ایران

۴- دانشجوی دکتری، علم اطلاعات و دانش‌شناسی، دانشگاه پیام نور و کارشناس ارشد، دانشگاه کاشان، کاشان، ایران (نویسنده مسؤل)

Email: alae62@gmail.com

مقدمه

متن دارای واژگان است و به عبارت بهتر متن دارای الگوهای واژگانی است. فراوانی ظهور واژه‌ها و توزیع آنها نیز مطرح است. از این الگوها می‌توان برای رتبه‌بندی متون بر اساس مشابهت ساختاری استفاده کرد، از طرفی کسب توانایی ارتباطی و تولید گفتمان فصیح و منسجم همیشه جزئی از اهداف زبان آموزان بوده است (۱). در سال‌های اخیر مطالعات زیادی در زمینه زبان شناسی کاربردی انجام شده و زبان شناسان از داده‌های مجموعه پیکره زبانی (corpus) استفاده کرده‌اند تا بر اهمیت دانش شناسایی الگوهای ترکیبات لغتی در زبان تاکید کنند، چرا که این نوع دانش به زبان آموزان کمک می‌کند تا به لحاظ ارتباطی توانا بوده و نوشتاری فصیح و گفتاری روان تولید نمایند.

اگر یک فرد آگاه متنی متشکل از مثلاً ۲۰۰۰ کلمه، برای انتشار در مجلات علمی بنویسد، باید از واژگان خاص یعنی ترکیبات واژگانی متداول در آن زمینه موضوعی، استفاده کند. ویراستاران و داوران عموماً بر این امر تاکید دارند و به همین علت نویسندگان با وجود آزادی زیادی که در انتخاب ترکیبات واژگانی دارند، کاملاً هم در این زمینه مختار نیستند. استفاده واژگان ممکن است در میان نویسندگان تنوع زیادی داشته باشد، اما در مجموع باید به نحو معقولی با آنچه خوانندگان می‌فهمند و انتظار دارند سازگار باشد. تحلیل محتوا از رویکردی خاص، برخوردار است که ناشی از نحوه تلقی از موضوع تحلیل می‌باشد. زنجیره‌های واژگانی به عنوان عضوی از خانواده عبارات قالبی گونه‌های کاربردی هستند که در متون دانشگاهی عملکردهای متفاوت گفتمانی ایفا می‌کنند. از آنجایی که این دسته‌ها اختصاص به گونه‌ای کاربردی دارند می‌توانند نقش مهمی در کارایی گفتمان متون دانشگاهی داشته باشند. در تحلیل محتوا واحد تحلیل می‌تواند کلمه، عبارت، مضمون، جمله‌بندی و مطالب مقاله باشد. سنجش‌های آماری متن عموماً متن را بر اساس سبک، توصیف می‌کنند و برای مقایسه یک متن با متن دیگر، آثار یک نویسنده، با نویسندگان دیگر و با همان نویسنده، مقایسه متن‌های نوشته

شده به یک زبان طبیعی با متون زبان دیگر و بررسی متون یک حوزه تخصصی به کار می‌روند (۱).

از اواخر ۱۹۷۰ میلادی، توجه بیشتری به عبارات قالبی در تولید و پردازش اطلاعات شده است (۲-۵) عبارات قالبی که معمولاً به عنوان ترکیبی از چند لغت تعریف می‌شوند به صورت کلی در قسمت واژگان مغز ذخیره شده و دوباره هنگام سخن گفتن بازیافت می‌شوند. گفته می‌شود این عبارات قالبی، امکان بیان روانتر را هم برای گوینده و هم برای مخاطب فراهم می‌آورد و در نتیجه باعث ایجاد گفتار روان می‌شود (۵۶). همچنین کاربرد مناسب عبارات قالبی در کسب توانش زبانی و در تولید زبان همانند سخنوران بومی بسیار مهم بوده است (۷). به علاوه واژه‌های «جمله‌بندی» و «زبان/عبارات قالبی» (۸، ۵) لغات جامعی هستند که معمولاً برای ارجاع به انواع مختلف ترکیب‌های چند لغتی به کار می‌روند. این لغات عبارتست از: خوشه یا دسته (۸، ۹)، ترکیب واژگان (۱۰، ۱۱) و زنجیره‌های واژگانی (۱۲، ۱۳). در طول ۲۵ سال گذشته، جمله بندی یکی از زمینه‌هایی بوده که در زبان‌شناسی کاربردی رشد سریعی داشته و این نشان می‌دهد که توالی لغات در کاربرد طبیعی زبان متداول است. از دیگر سو این مطلب در پردازش و ذخیره اطلاعات علمی در نظام‌های بازایی نیز مورد توجه قرار نگرفته است (۱۳، ۱۴).

عبارات قالبی تاریخچه‌ای طولانی در زبان‌شناسی کاربردی دارند و به زمان Jespersen و Firth بر می‌گردد. در آن زمان این افراد اصطلاح هم‌آیندی واژگان (Collocation) را عمومی کردند. البته این اعتقاد وجود دارد که دانش با همایی واژگان یا تمایل واژه برای همایی با یک یا چند واژه، ارتباط را آسان‌تر کرده و به نویسندگان کمک می‌کند متن‌هایی بنویسند که فصیح و منسجم هستند (۱۳).

بیشتر زبان‌شناسان علاقه‌مند بوده‌اند تا روی ترکیب‌های چند لغتی کار کنند. آنها این ترکیب را ترکیب لغات متناوب می‌نامند. مطالعات انجام شده در این زمینه در مورد رواج و حضور غالب عبارات قالبی توضیح داده و به دو ویژگی از این عبارات اشاره می‌کند: روانشناسی زبانی و تعامل اجتماعی.

مطالعات به بررسی سیاق نگارش متون آکادمیک پرداخته است (۱۵، ۱۶).

زبان‌شناسی پیکره‌ای (Corpus linguistics) اصطلاح «زنجیره واژگانی» را روشن‌تر توضیح می‌دهد. این واژه برای اولین بار در فرهنگ Longman با عنوان گرامر انگلیسی گفتاری و نوشتاری دیده شد (۱۷). با وجود این، مفهوم زنجیره واژگانی به دوره Salem بر می‌گردد (۱۸). او پژوهشی مبنی بر تجزیه و تحلیل اسناد حکومتی دولت فرانسه انجام داد. شخص دیگری که در این زمینه کار کرده Altenberg به عنوان اولین محقق است که ترکیب واژگان را در انگلیسی بررسی کرده است.

زنجیره‌های واژگانی را به عنوان اصطلاحات متناوبی تعریف می‌کنند که در کاربرد طبیعی زبان و بدون توجه به وضعیت ساختاری یا واژگانی یا سبکی شان با یکدیگر می‌آیند. زنجیره‌های واژگانی به عنوان بخشی از خانواده بزرگتر که عبارات قالبی هستند در نظر گرفته می‌شوند و گفته می‌شود که به عنوان واحدی کلی در مغز ذخیره شده و پردازش می‌شوند. در واقع زنجیره‌های واژگانی، ترکیبی از واژگان هستند که عملکرد خاص داشته و با نگرش ناشی از مجموعه یا پیکره زبانی بازیافت می‌شوند. Cortes (۲۰۰۲، ۲۰۰۴ میلادی) زنجیره‌های واژگانی در مقالات پژوهشی را دسته‌های تعیین شده می‌نامد (۱۳). Biber و همکارانش ترکیب‌های لغاتی که متناوباً با یکدیگر می‌آیند را زنجیره واژگانی می‌نامند (۱۷)، در حالی که Scott آنها را «خوشه» نامید (۱۹). در انجام یک سری از مطالعات در زمینه بررسی زنجیره‌های واژگانی و مقایسه آنها در بین گونه‌های کاربردی Biber و همکارانش یافتند که ساختارهای گرامری این زنجیره‌های واژگانی ویژگی خاص گونه‌های کاربردی است (۱۷).

برای مثال، Cortes در سال ۲۰۰۴ میلادی مطالعه‌ای انجام داد که در آن او زنجیره‌های واژگانی موجود در مقالات چاپ شده و نوشته‌های دانشجویان در حوزه تاریخ و زیست‌شناسی را بررسی کرد. نتایج مطالعه نشان داد که نویسندگان مقالات

عملکردهای این عبارت قالبی به ویژه در هنگام انجام کارهای دشوار و هم‌زمان زبانی و انجام آنها در مدت زمان کم یک سری مزایا به هنگام پردازش داده‌ها به همراه دارند. یکی از دلایل آن این است که توالی یا دسته‌ای از لغات می‌توانند به عنوان یک واحد در مغز ذخیره و پردازش شوند و این در صورتی است که سرعت پردازش این واحدها خیلی بیشتر از پردازش لغاتی است که طبق قوانین زبانی کنار یکدیگر قرار گرفته‌اند (۵).

Wray در سال ۲۰۰۲ میلادی مدلی منسجم برای دو عملکرد عبارات قالبی که در بالا به آن اشاره شد ارائه کردند و برای هر کدام زیر مجموعه‌هایی نیز در نظر گرفتند. برای عملکرد روان‌شناسی زبانی این عبارات سه زیر مجموعه در نظر گرفتند که عبارتند از: ۱- میان برهای پردازشی ۲- کاربردپذیری‌های زمان ۳- استفاده درست و اداره اطلاعات (۵). در پژوهشی پیرامون عبارات قالبی به منظور فهم بهتر اینکه آن سخنوران بومی و غیر بومی عبارات قالبی را سریعتر از عبارات غیر قالبی پردازش می‌کنند یا نه، عبارات قالبی خیلی سریعتر و راحت‌تر از عبارات غیر قالبی پردازش می‌شوند. به علاوه آنها یافتند سخنوران غیر بومی توانستند از مزایای پردازش این نوع عبارات بهره‌مند شوند. آنها همچنین بیان داشتند که اگرچه سرعت خواندن عبارات قالبی برای سخنوران غیر بومی در مقایسه با سخنوران بومی کمتر بود این عبارات حتی با این سرعت پائین پردازش هم برای غیر بومیان مفید بودند. ویژگی دوم مربوط به عملکرد تعامل اجتماعی عبارات قالبی است.

Wray در مدل خود سه زیر مجموعه برای این عملکرد نیز در نظر گرفتند که عبارتند از ۱- استفاده درست و اداره دیگران است که مورد رضایت‌مندی به لحاظ فیزیکی، احساسی و نیازهای شناختی است. ۲- مطالبه هویت جداگانه یا گروهی است که مربوط به هویت شخصی سخنور است. ۳- داشتن هدف به منظور کنترل قدرت درک و دریافت شنونده است (۵). تحقیق در مورد بررسی عبارات قالبی در متون دانشگاهی هدف بسیاری از مطالعات بوده است. تعدادی از این

نوشتار تاثیر می‌گذارد. همچنین، این لغات ترکیبی باعث افزایش توانایی کاربردی دانشجویان می‌شود. پیامد آموزشی جلب توجه این است که دانشجویان به طور نظام‌مند تشویق می‌شوند تا به این زنجیره‌ها در متون معتبر در زبان دوم توجه کنند. Lewis توجهی به راهکارها برای حفظ و ذخیره این زنجیره‌ها برای کاربردهای بعدی نمی‌کند بلکه بیشتر به ایجاد آگاهی از این زنجیره‌های واژگانی توجه دارد و معتقد است که این یادگیری از طریق تقلید این واژگان در درون یا بیرون از کلاس افزایش می‌یابد. پس، دانشجویان انگلیسی برای اهداف پزشکی (English for Medical Purposes (EMP)) از فراهم آوردن لیستی از این زنجیره‌های واژگانی بهره‌مند خواهند شد چرا که کسب توانایی بالا یا دانش زیاد در زمینه زنجیره‌های واژگانی به دانشجویان کمک می‌کند تا در گفتار و نوشتار زبان دوم بسیار ماهر و توانمند باشند. از طرفی معرفی این ترکیب‌های واژگانی به نظام‌های پردازش و بازیابی اطلاعات و شناسایی و توجه به آنها در هنگام ذخیره سازی اطلاعات در نظام بازیابی باعث بهبود نتایج در بازیابی از پایگاه‌های اطلاعاتی تخصصی خواهد شد. این پژوهش با هدف معرفی رایج‌ترین زنجیره‌های واژگانی در مقالات پژوهشی حوزه پزشکی، تعیین ساختار زنجیره‌های واژگانی به کار رفته در مقالات، و نقشی که زنجیره‌های واژگانی در مقالات پژوهشی پزشکی ایفا می‌کنند، انجام شده است.

روش بررسی

این مطالعه از نوع تحلیل محتوا بر اساس ساختار دستوری است، به منظور جمع‌آوری مقالات پژوهشی مورد نیاز برای ایجاد پیکره زبانی، پایگاه اطلاعاتی Science Direct استفاده شد. این پایگاه اطلاعاتی یکی از معتبرترین پایگاه‌های اطلاعاتی تمام متن در حوزه‌های علمی است و ۷۹۰ مقالات پژوهشی پزشکی به کار رفته در مطالعه حاضر از این پایگاه بازیابی شدند. در شاخه پزشکی و دندانپزشکی این پایگاه، ۳۳ زمینه پزشکی وجود دارند. به دنبال روشی که Wang و همکارانش (۲۰۰۸ میلادی) در مطالعه خود به

پژوهشی از زنجیره‌های واژگانی بیشتری در نوشته‌های خود استفاده می‌کردند و این زنجیره‌ها در نوشته‌های دانشجویان بسیار کمتر دیده می‌شد. بعضی دیگر از مطالعات بر بررسی تفاوت و شباهت زنجیره‌های واژگانی در بین انواع نوشته‌های یک رشته پرداخته‌اند (۱۷، ۲۲-۲۰).

جلالی به مطالعه زنجیره‌های واژگانی در انواع نوشته‌های دانشگاهی مثل مقالات پژوهشی، پایان‌نامه‌های کارشناسی ارشد و رساله‌های دکتری در رشته زبان‌شناسی کاربردی پرداخت (۲۳). در حالی که مطالعات دیگر در این زمینه به بررسی کاربرد زنجیره‌های واژگانی در رشته‌های مختلف پرداختند (۹، ۱۲، ۱۳). Hyland در سال ۲۰۰۸ میلادی زنجیره‌های واژگانی در انواع نوشته‌ها مثل مقالات پژوهشی، پایان‌نامه‌های کارشناسی ارشد و رساله‌های دکتری در چهار رشته مهندسی برق، مطالعات بازرگانی، زبان‌شناسی کاربردی و زیست‌شناسی بررسی کرد (۹). مطالعات دیگر نیز روی ساختار زنجیره‌های واژگانی در قسمت‌های مختلف متن متمرکز شده و عملکرد و نقش این زنجیره‌ها را در بخش‌های مختلف بررسی کردند (۱۷، ۲۰-۲۲). مطالعه تحقیقات انجام شده بر روی مقالات پژوهشی پزشکی نشان داد که بعضی از تحقیقات به بررسی ترکیبات دو لغتی و فهرست لغات آکادمیک متون پزشکی (۱۷) پرداخته‌اند اما مطالعه‌ای به بررسی و تحلیل محتوا بر اساس زنجیره‌های واژگانی در متن مقالات پژوهشی پزشکی نپرداخته است.

در نگارش مقالات علمی و پژوهشی از زنجیره‌های واژگانی متفاوتی استفاده می‌شود. بنابراین، دانشجویان رشته‌های مختلف باید از این زنجیره‌های واژگانی آگاه باشند تا از این طریق توانایی آنها در فهم مطالب و تولیدات علمی آنها افزایش یابد. این دانش و توانایی از طریق دانستن فهرستی از این زنجیره‌های واژگانی یا از طریق خواندن متونی که شامل این زنجیره‌ها می‌باشند به دست می‌آید.

دانشجویان مرتباً با این زنجیره‌های واژگانی در متون علمی برخورد می‌کنند و عدم توانایی آنها در فهم این ترکیبات بر موفقیت دانشجویان و پژوهشگران در تولید و فهم گفتار و

کاربرد، در مطالعه حاضر مقالات چاپ شده از سال ۲۰۰۹-۲۰۱۱ میلادی تمام ۳۳ زمینه پزشکی برای ایجاد پیکره زبانی به کار گرفته شد. در هر سال، دو شماره از هر جلد مجله به صورت تصادفی انتخاب شد. به علاوه، در این مجلات، مقالاتی انتخاب شده که دارای ساختار ۴ قسمتی مقدمه، روش، نتیجه و بحث (IMRAD, Introduction, Methods, Research [and] Discussion) بودند و آنهایی که این ساختار را نداشتند از جامعه حذف شدند. برای این اساس، ۲۱ مقاله از هر زمینه پزشکی انتخاب شد (۴). هر مقاله به صورت متوسط شامل ۳۰۰۰ لغت بود. در نهایت، ۷۹۰ مقاله، پیکره زبانی به اندازه ۲۴۲۰۹۱۴ لغت تشکیل داد. بسامد تکرار ۲۰ بار در هر یک میلیون لغت و تکرار این بسامد حداقل در ۵ متن مختلف به عنوان معیار تحلیل محتوا در نظر گرفته شدند. در بیشتر مطالعات لغات ۴ تایی در نظر گرفته می‌شوند چرا که زنجیره‌های ۴ لغتی، زنجیره‌های ۳ لغتی را نیز در خود دارند. به علاوه، زنجیره‌های ۴ لغتی، در داده یک میلیون لغتی بیش از ۱۰ بار اتفاق می‌افتند و ساختارها و نقش‌های بیشتری نیز ایفا می‌کنند. بنابراین، در مطالعه حاضر زنجیره لغات ۴ تایی در نظر گرفته شدند. نرم‌افزارهای کامپیوتری که در این مطالعه به کار گرفته شدند عبارتند از: Word smith tools 5, Antconc3.23, ABBYY Fine Reader to professional edition و ABBYY Fine Reader سیستم نویسه‌خوان نوری

قادر است فایل‌های PDF و دیگر فایل‌ها را به شکلی قابل ویرایش تغییر دهد. این نرم افزار فرمت PDF را به خدمت txt تبدیل می‌کند تا متون قابل خواندن توسط نرم افزار Antconc باشند. فایل‌های txt به دست آمده به نرم افزار Antconc داده شدند و طول زنجیره‌های واژگانی ۴ در نظر گرفته شد. همچنین، واژه‌های متفاوتی چون حروف تعریف، افعال to be، حروف اضافه، صفات اشاره به این نرم‌افزار داده شد. به علاوه، بسامد ۲۰ در یک میلیون لغت نیز در نظر گرفته شد. بنابراین، بسامد ۴۸ برای پیکره زبانی ۲۴۲۰۹۱۴ لغتی به نرم افزار داده شد. با دادن این مشخصات به نرم افزار Antconc3.23، زنجیره‌های واژگانی ۴ لغتی از مقالات پزشکی به دست آمد که نرم افزار به طور خودکار آنها را بر اساس بسامد مرتب می‌کرد. سپس نرم افزار Word Smith استفاده شد تا تعداد متن‌های با زنجیره‌های دارای بسامد ۲۰ را نشان دهد. مرحله بعد دسته‌بندی زنجیره‌های واژگانی به لحاظ ساختاری و عملکرد بود. برای دسته‌بندی ساختاری زنجیره‌های واژگانی از دسته‌بندی Biber و همکارانش (۱۷) و برای طبقه‌بندی عملکردی زنجیره‌ها از دسته بندی Hyland (۲۰۰۸ میلادی) استفاده شد (۱۶). از آنجا که مطالعه حاضر بر گونه خاصی از متون علمی متمرکز می‌شود از دسته‌بندی و رده‌بندی Hyland استفاده شد (جدول ۱).

جدول ۱: تاکسونومی Hyland

زیر مجموعه‌ها	نقش‌های اصلی (Major functions)
مکان: برای نشان دادن زمان و مکان فرآیند: برای نشان دادن روش و هدف تحقیق اندازه‌گیری: برای توصیف میزان و اعداد توصیف: برای شرح دادن کیفیت و خصوصیات مواد موضوع: مرتبط با حوزه پژوهش	پژوهش محور (Research-oriented): این زنجیره‌ها به نویسنده کمک می‌کنند تا به فعالیت‌ها و تجربیاتشان از جهان واقعی نظم دهند
علائم انتقال: ارتباط جمعی یا تضادی بین عناصر متن را نشان می‌دهد علائم نتیجه: ارتباط استنتاجی و سببی بین عناصر متن را نشان می‌دهد	متن محور (Text-oriented): این زنجیره‌ها در مورد ساخت متن و معنای عناصر آن هستند

ادامه جدول ۱: تاکسونومی Hyland

زیر مجموعه‌ها	نقش‌های اصلی (Major functions)
<p>علائم ساختاری: نشانه‌های انعکاسی که متن را سازماندهی کرده و خواننده را در طول متن راهنمایی می‌کند</p> <p>علائم طرح‌ریزی (چینش): بیان مطالب از طریق استفاده عباراتی که شرایط خاص و محدود را نشان می‌دهند.</p> <p>ویژگی‌های نقطه نظر: این عبارات نگرش و ارزیابی نویسنده را نشان می‌دهد.</p> <p>ویژگی‌های درگیر کردن خواننده در متن: خواننده را مستقیماً در متن خطاب قرار می‌دهد.</p>	<p>شرکت‌کننده محور (Participant-oriented): این زنجیره‌های واژگانی نشان دهنده دیدگاه نویسنده یا خواننده متن است.</p>

جدول ۳ ارائه شده‌اند. این زنجیره‌ها مبنی بر اینکه با اسم شروع شدند یا حروف اضافه یا فعل، دسته‌بندی شدند. همانطور که قبلاً گفته شد، برای دسته‌بندی ساختاری زنجیره‌های واژگانی از طبقه بندی شده توسط Biber و همکارانش (۱۹۹۹ میلادی) استفاده شد (۱۷). دسته‌بندی زنجیره‌ها نشان داد که بزرگترین دسته ساختاری مربوط به عبارات حرف اضافه‌ای است که حدود ۴۴/۵ درصد کل زنجیره‌ها را تشکیل می‌دهد. عبارات اسمی با تناوب ۱۸۴۲ (حدود ۲۰/۴۲ درصد) و عبارات فعلی با تناوب ۱۶۷۰ (حدود ۱۸/۵ درصد) زنجیره‌های دیگر این پیکره زبانی هستند. کمترین تناوب مربوط به دسته *verbphrase that claus* هستند که حدود ۱/۷ درصد زنجیره‌های واژگانی را تشکیل می‌دهند. دسته‌بندی این زنجیره‌ها بر اساس ساختارشان در جدول ۳ ارائه شده است.

همانگونه که پیشتر گفته شد، برای تجزیه و تحلیل زنجیره‌های واژگانی از طبقه بندی Hyland (۲۰۰۸ میلادی) استفاده شد. این طبقه بندی شامل سه دسته اصلی است: زنجیره‌های پژوهش محور، متن محور و شرکت‌کننده محور که هر کدام از این‌ها شامل زیر مجموعه‌هایی هستند.

جدول ۴ نتایج طبقه‌بندی عملکردی و زنجیره‌های واژگانی را ارائه نموده و تعداد زنجیره‌های واژگانی، تناوب آنها و تعداد متن‌هایی که در آنها تکرار شدند را ارائه می‌دهد.

یافته‌ها

به طور کلی ۱۰۲ زنجیره ۴ واژه‌ای در پیکره زبانی ۲۰۱۴، ۴۲۰ لغتی مقالات پژوهشی پزشکی به دست آمد. این زنجیره‌های واژگانی به همراه بسامد و تعداد متونی که در آنها به کار رفته‌اند را ارائه می‌دهد (جدول ۲).

آنچه که از جدول می‌توان فهمید این است که ۰/۳ درصد از پیکره زبانی مقالات پزشکی را زنجیره واژگانی *in the present study* است که ۴۵۳ می‌باشد و این زنجیره واژگانی در ۲۲۰ متن مختلف تکرار شده است. این بسامد بالا به این معنا است که در هر یک میلیون لغت این زنجیره واژگانی ۲۲۶ بار تکرار شده که تقریباً ۱۱ برابر بسامد استاندارد ۲۰ در هر یک میلیون لغت می‌باشد. همچنین زنجیره‌هایی مثل *plays an important role, should be noted that, the presence of a , was used for the* کمترین بسامد را دارا هستند.

به علاوه، طبق جدول ۲، تعداد متون با زنجیره‌های کم بسامد نیز بسیار کم می‌باشد. و این زنجیره‌ها با بسامد ۴۸ در ۳۹ متن مختلف تکرار شده‌اند که ۹ بار کمتر از تعداد متن‌هایی است که بیشترین بسامد در آنها تکرار شده است.

بعد از اینکه زنجیره‌های واژگانی به دست آمدند، به لحاظ ساختاری دسته‌بندی شدند. در این مطالعه تنها ۹ دسته‌بندی اصلی برای ساختار زنجیره‌های واژگانی استفاده شد که در

جدول ۲: زنجیره‌های واژگانی به همراه بسامد و فراوانی متون در جامعه آماری این پژوهش

ردیف	زنجیره های واژگانی	فراوانی	تعداد متن	بسامد	زنجیره های واژگانی	فراوانی	تعداد متن
۱	in the present study	۴۵۳	۲۲۰	۵۲	may be due to	۷۲	۶۴
۲	on the other hand	۲۵۸	۱۸۲	۵۳	with the exception of	۷۲	۵۷
۳	in the presence of	۲۲۴	۱۲۰	۵۴	are more likely to	۷۲	۴۹
۴	At the end of	۲۰۱	۱۱۹	۵۵	for each of the	۷۰	۵۷
۵	at the time of	۱۸۵	۱۲۳	۵۶	as a function of	۶۹	۳۱
۶	were more likely to	۱۷۷	۷۷	۵۷	an increase in the	۶۸	۵۱
۷	on the basis of	۱۶۶	۱۰۰	۵۸	results of this study	۶۷	۵۵
۸	the end of the	۱۶۵	۱۱۲	۵۹	for the treatment of	۶۷	۵۴
۹	It is possible that	۱۶۳	۱۱۳	۶۰	in the treatment of	۶۶	۴۹
۱۰	as well as the	۱۵۹	۱۲۵	۶۱	presence or absence of	۶۵	۴۹
۱۱	The results of the	۱۴۸	۱۱۷	۶۲	are presented in Table	۶۴	۵۳
۱۲	of the present study	۱۴۸	۱۱۲	۶۳	was used to determine	۶۴	۵۷
۱۳	as shown in Fig	۱۴۸	۸۰	۶۴	aim of this study	۶۳	۶۲
۱۴	in the control group	۱۳۹	۶۵	۶۵	are summarized in Table	۶۳	۵۲
۱۵	In the current study	۱۲۷	۶۷	۶۶	were obtained from the	۶۳	۵۷
۱۶	are shown in Table	۱۲۶	۱۰۱	۶۷	Are summarized in table	۶۳	۵۲
۱۷	this study was to	۱۲۵	۱۱۶	۶۸	and the number of	۶۱	۴۸
۱۸	It is important to	۱۲۳	۱۰۰	۶۹	in terms of the	۶۱	۴۱
۱۹	studies have shown that	۱۰۴	۸۹	۷۰	In this study were	۶۱	۵۷
۲۰	in the absence of	۱۲۱	۸۴	۷۱	as well as in	۶۰	۵۶
۲۱	In the case of	۱۲۱	۸۲	۷۲	has been reported to	۶۰	۴۹
۲۲	more likely to be	۱۱۹	۶۱	۷۳	In this study, the	۵۹	۵۲
۲۳	been shown to be	۱۱۸	۹۶	۷۴	the fact that the	۵۹	۵۱
۲۴	was found to be	۱۱۷	۸۸	۷۵	with the use of	۵۹	۴۶

ادامه جدول ۲: زنجیره‌های واژگانی به همراه بسامد و فراوانی متون در جامعه آماری این پژوهش

ردیف	زنجیره های واژگانی	فراوانی	تعداد متن	ردیف	زنجیره های واژگانی	فراوانی	تعداد متن
۲۵	in the United States	۱۰۹	۷۷	۷۶	the extent to which	۵۸	۳۵
۲۶	is one of the	۱۰۸	۹۱	۷۷	as compared to the	۵۸	۲۵
۲۷	In this study, we	۱۰۷	۹۳	۷۸	purpose of this study	۵۶	۵۳
۲۸	an important role in	۱۰۶	۸۶	۷۹	In our study, the	۵۶	۵۲
۲۹	as a result of	۱۰۳	۸۴	۸۰	present study was to	۵۵	۵۴
۳۰	have been shown to	۹۸	۷۵	۸۱	important role in the	۵۵	۵۰
۳۱	one of the most	۹۴	۸۶	۸۲	may be related to	۵۵	۴۵
۳۲	As shown in Table	۹۳	۶۶	۸۳	It should be noted	۵۵	۴۳
۳۳	The results of this	۹۲	۷۵	۸۴	the basis of the	۵۵	۴۲
۳۴	in accordance with the	۸۵	۷۵	۸۵	the presence or absence	۵۴	۴۰
۳۵	The purpose of this	۸۲	۷۴	۸۶	a large number of	۵۳	۴۶
۳۶	at the same time	۸۲	۷۰	۸۷	by the fact that	۵۲	۴۷
۳۷	the total number of	۸۲	۶۲	۸۸	is consistent with the	۵۲	۴۵
۳۸	It has been shown	۸۱	۶۷	۸۹	were less likely to	۵۲	۳۱
۳۹	in the context of	۷۹	۵۵	۹۰	the presence of the	۵۱	۳۸
۴۰	was used as a	۷۷	۶۶	۹۱	a role in the	۵۱	۴۵
۴۱	were found to be	۷۷	۵۷	۹۲	at the beginning of	۵۰	۴۱
۴۲	was defined as the	۷۷	۶۳	۹۳	higher than that of	۵۰	۳۱
۴۳	a wide range of	۷۷	۶۶	۹۴	could be due to	۵۰	۲۷
۴۴	with respect to the	۷۶	۶۲	۹۵	play an important role	۴۹	۴۵
۴۵	during the study period	۷۶	۴۳	۹۶	was obtained from the	۴۸	۴۳
۴۶	The number of patients	۷۵	۴۳	۹۷	the size of the	۴۹	۴۰
۴۷	in the number of	۷۵	۵۷	۹۸	The time of the	۴۸	۴۲
۴۸	In addition to the	۷۴	۶۹	۹۹	plays an important role	۴۸	۴۵
۴۹	Be due to the	۷۴	۶۴	۱۰۰	should be noted that	۴۸	۴۰
۵۰	test was used to	۷۳	۶۶	۱۰۱	the presence of a	۴۸	۳۹
۵۱	The aim of this	۷۱	۷۱	۱۰۲	was used for the	۴۸	۴۷

جدول ۳: دسته‌بندی ساختاری زنجیره‌ها در جامعه آماری این پژوهش

درصد	فراوانی کل	تعداد زنجیره	نمونه‌ها	ساختار
۱۵/۴۲	۱۳۹۱	۱۹	The end of the, the results of the, the results of this	Noun phrase+ of
۵/۰۰	۴۵۱	۷	A role in the, the extent to which, the fact that the	Other noun phrase
۲۰/۵۰	۱۸۵۰	۱۸	In the presence of, in the development of, at the time of	Prepositional phrase+ of
۲۴/۰۳	۲۱۶۸	۱۹	At the same time, between the two groups, in the current study	Other prepositional phrases
۱۱/۲۴	۱۰۱۴	۱۳	Are shown in table, was used for the, are summarized in table	Passive+ prepositional phrase fragment
۴/۷	۴۲۲	۴	It is possible that, it is important to, it should be noted	Anticipatory it+ verb/adjective
۲/۵۹	۲۳۴	۳	Is one of the, is consistent with the, is possible that the	Be+ noun/adjectival phrase
۱/۷	۱۵۲	۲	Should be noted that, studies have shown that	Verb phrase+ that clause fragment
۳/۳۳	۳۰۱	۳	Are more likely to, can be used to, used to determine the	Verb/adjective+ to-clause fragment
۳/۳۱	۲۹۹	۳	As shown in figure, as compared to the	Adverbial clause fragment
۸/۱۸	۷۳۸	۱۱	This study was to, test was used to, as well as in	Others
۱۰۰	۹۰۲۰	۱۰۲		Total

جدول ۴: طبقه‌بندی عملکردی زنجیره‌های جامعه آماری این پژوهش

درصد	فراوانی	تعداد زنجیره	نوع زنجیره
۳۶/۵۳	۳۳۴۷	۳۷	Research-oriented bundles
۶/۱۸	۱۶۵۵	۱۲	مکان
۱۵/۱۸	۷۴۷	۱۲	فرآیند
۷/۰۷	۶۴۸	۱۰	اندازه‌گیری
۰۵/۲	۱۸۸	۲	توصیف
۱/۱۸	۱۰۹	۱	موضوع
۴۲/۴۷	۳۸۹۲	۴۰	Text-oriented bundles
۶/۰۱	۵۵۱	۴	علایم انتقال
۲۴/۱۲	۱۱۲۲	۱۲	علایم نتیجه
۷/۱۱	۶۵۲	۷	علایم ساختاری
۱۷/۱۰	۱۵۶۷	۱۷	علایم طرح‌ریزی
۲۱	۱۹۲۳	۲۵	Participant-oriented bundles
۱۹/۸۶	۱۸۲۰	۲۳	ویژگی‌های نقطه نظر
۴/۱۵	۳۸۱	۵	Attitude markers
۱/۷۷	۱۶۳	۳	Epistemic-certain
۹/۸۸	۹۰۶	۱۰	Epistemic- uncertain
۴/۰۳	۳۷۰	۵	Intention
۱/۱۲	۱۰۳	۲	ویژگی‌های درگیر کردن خواننده
۱۰۰	۹۱۶۲	۱۰۲	Total

علاوه بر عبارات گروهی و بندی، گروه دیگری از زنجیره‌ها هستند که Biber و همکارانش (۱۹۹۹ میلادی) آن را ترکیباتی می‌دانند که در هیچ کدام از ساختارها نمی‌گنجد (۱۷). این نوع از زنجیره‌های واژگانی ۸/۱۸ درصد کل زنجیره‌های پیکره زبانی را تشکیل می‌دهند و به عنوان مثال می‌توان به موارد زیر اشاره نمود.

As well as the, consent way obtained from, study was approved by

علاوه بر دسته‌بندی ساختاری و گرامری زنجیره‌های واژگانی، عملکرد، نقش و معنای آنها نیز باید بررسی شود چرا که نقشی که آنها در هر متن ایفا می‌کنند نشان دهنده گونه کاربردی است که در آن یافت می‌شوند (۹، ۲۴).

همانگونه که در جدول ۳ نشان داده شده است حدود ۳۶/۵ درصد زنجیره‌های واژگانی متعلق به گروه زنجیره‌های پژوهش محور هستند و این زنجیره‌ها در متون علمی به کار می‌روند تا به زمان، مکان، اندازه و خود مطالعه و فرایندهای مربوط به آن اشاره کنند. همانطور که در قبل اشاره شد، متناوب‌ترین زنجیره واژگانی در پیکره زبانی مقالات پزشکی in the present study است که حدود ۰/۱ درصد کل زنجیره‌ها را تشکیل می‌دهد. اگرچه متناوب‌ترین زنجیره واژگانی در پیکره زبانی مطالعه حاضر مربوط به دسته پژوهش محور است، می‌توان گفت مقالات پژوهشی پزشکی بیشتر بر مبنای زنجیره‌های متن_محور به ویژه «علایم طرح ریزی» (Framing Signal) هستند و کمتر از زنجیره‌های شرکت‌کننده محور استفاده می‌کنند. چرا که زنجیره‌های متن_محور حدود ۴۲/۵ درصد کل زنجیره‌های واژگانی در پیکره زبانی مقالات پژوهشی پزشکی تشکیل می‌دهند که در آن «علایم طرح ریزی» با تناوب ۱۵۶۷ دارای بالاترین تناوب هستند. زنجیره‌های شرکت‌کننده_محور کمترین تناوب را نشان می‌دهند. این زنجیره‌های واژگانی بیشتر روی خواننده و نویسنده متن متمرکز می‌شوند. در این گروه، زنجیره واژگانی مثل *to were more likely* متناوب‌ترین زنجیره است که نویسندگان در متون پزشکی به کار برده‌اند. این دسته از زنجیره‌های واژگانی حدود ۲۱ درصد زنجیره‌های

بحث

زنجیره‌های دارای بسامد زیاد، زنجیره‌های رایج یا بدون معنی خاص هستند، با این حال از نظر کاربرد در متون در یک متن مفصل مهم و تعیین کننده تلقی می‌شوند. بسامد نسبی زنجیره‌ها در یک زمینه موضوعی از یک نویسنده به نویسنده دیگر تفاوت چندانی ندارد که البته در همه انواع نوشته‌ها صدق نمی‌کند (۱).

همانگونه که در جدول ۳ نشان داده شده، در پیکره زبانی مقالات پزشکی حدود ۶۴/۹۵ درصد زنجیره‌های گروهی بوده که در میان آنها عبارات حرف اضافه‌ای پر بسامدترین بندها هستند. علاوه بر عبارات حرف اضافه‌ای، عبارات اسمی گروه دیگری از زنجیره‌های گروهی را تشکیل می‌دهند. حدود ۲۶/۸۷ درصد را زنجیره‌های بندی تشکیل می‌دهند و در میان آنها آن دسته از زنجیره‌هایی که با فعل مجهول به همراه عبارت حروف اضافه‌ای شروع می‌شدند پربسامدترین گروه هستند. همانگونه که پیش از این اشاره شد، متناوب‌ترین گروه از زنجیره‌های واژگانی عبارات حروف اضافه‌ای هستند که برای اشاره به زمان یا مکان به کار می‌روند که از آن جمله می‌توان به این موارد اشاره نمود:

As in the present study, on the other hand, at the end of, in the presence of, at the time of
زنجیره‌هایی که زیر مجموعه دسته عبارات اسمی+ حرف اضافه of قرار می‌گرفتند، دومین دسته‌ای بودند که بسامد بالایی را نشان دادند. آنها حدود ۲۰/۴۲ درصد کل زنجیره‌های موجود در پیکره زبانی را تشکیل می‌دادند.

در دسته بندی عبارات فعلی که شامل زیر مجموعه‌هایی مثل *anticipatory it +verb/adjective, passive+ prepositional phrase, verb phrase +that-clause fragment, copula be+ noun/adjectival verb/adjective+ to clause fragment, phrase* می‌شود دسته‌ای که با افعال مجهول شروع می‌شدند بیشترین بسامد که حدود ۱۴/۰۱ بود را به خود اختصاص دادند. این نتیجه نیز تأییدی است بر این حقیقت که افعال مجهول بیشترین ساختار استفاده شده در متون دانشگاهی هستند،

طرف دیگر در نظام‌های بازیابی اطلاعات سابقه‌ای مبنی بر کاربرد زنجیره‌های واژگانی به منظور بهبود نظام بازیابی وجود ندارد. البته تنها یک سابقه در جهت مخالف (آنهم البته در جهت بهبود کاوش) می‌توان یافت و آن وجود فهرست واژه-های عمومی در پایگاه داده نظام است که نظام این واژه‌ها را جستجو نمی‌کند. با عنایت به اینکه زنجیره‌ها از متون پژوهشی یک حوزه علمی استخراج شده‌اند و جامعه نمونه نیز جامعه بسیار بزرگی بوده است. به کارگیری این زنجیره‌ها و زنجیره‌های واژگانی مشابه حاصل پژوهش‌های دیگر، با ساختارهای نحوی و معنایی متنوع، در نظام‌های بازیابی اطلاعات تخصصی پیشنهاد می‌شود.

همان‌طور که می‌دانیم متن دارای قواعدی است که با استفاده از آن، کلمات در عبارتها و جملات معنی‌دار ترکیب می‌شوند. قواعد نحوی به روشن کردن معنی و اهمیت این کلمات کمک می‌کنند و عکس این نیز صادق است. ارتباط نحوی و معنای کلمات پیچیدگی تحلیل را حتی بیشتر می‌کند و این برای نظام‌های پردازش رایانه‌ای کار آسانی نیست، خلا پژوهشی در این نوع تحلیل‌ها نیاز به پژوهش‌های کاربردی دارد که البته ترکیب کلمات غالباً با توجه به زمینه یا حوزه‌ای که در آن به کار می‌روند معنی می‌شوند. الگوهای حاصل همچنین می‌توانند کلمات یا عباراتی را شناسایی کنند که در صورت جداسازی از متن به عنوان نمایه کاربرد داشته باشند.

تشکر و قدردانی

نویسندگان بر خود لازم می‌دانند از راهنمایی‌های ارزنده دکتر محمدرئوف معینی در بخش زبان‌شناسی پژوهش تشکر و قدردانی نمایند. همچنین از اهتمام و حسن نظر سردبیر، داوران و کادر علمی نشریه مدیریت اطلاعات سلامت کمال تشکر را دارند.

موجود در پیکره زبانی مقالات پژوهشی پزشکی را تشکیل می‌دهند و در میان آنها زنجیره‌های دارای ابهام شناختی (framing signal) بیشترین تناوب را دارا هستند. بنابراین بیشتر نویسندگان متون پزشکی از این دسته از زنجیره‌های واژگانی استفاده کرده‌اند تا خود را از متن دور نگه دارند.

زنجیره‌های «ویژگی‌های درگیرکننده» (engagement) در متون علمی استفاده می‌شوند تا خواننده را در متن درگیر کنند. برای مثال، در گفتمان گفتاری زنجیره‌های واژگانی مثل and you have to برای درگیر یا جذب فرد مخاطب استفاده می‌شود (۲۵). با وجود این در گفتمان نوشتاری علمی، شکل‌های دیگر ترجیح داده می‌شوند همانگونه که در مطالعه حاضر به دست آمده مثل noted that should be. همانگونه که در جدول ۳ نشان داده شده، این زنجیره‌ها حدود ۱ درصد کل زنجیره‌ها در پیکره زبانی مقالات پزشکی تشکیل می‌دهند. هدف این زنجیره‌ها هدایت خوانندگان برای فهم بهتر و برداشتی خاص از متن است.

نتیجه‌گیری

بنابر یافته‌های این پژوهش می‌توان گفت نویسندگان مقالات پژوهشی پزشکی بیشتر از زنجیره‌های متن محور استفاده می‌کنند تا خوانندگان خود را در طول متن هدایت کرده و به فهم بهتر مطالب کمک کنند. کاربرد زنجیره‌های متن محور آگاهی نویسنده از قواعد بلاغی گفتمان و انتظارات خوانندگان متون علمی را نشان می‌دهد.

پیشنهادها

همانطور که پژوهش حاضر نشان داد، یکی از نتایج فهرستی از زنجیره‌های واژگانی پرکاربرد در علوم پزشکی بودند. مبحث تحلیل محتوا را می‌توان در همین جا تمام شده فرض کرد. از

References

1. Meadow CT, Boyce BR, Kraft DH, Barry C. Text information retrieval systems. Hariri N, Trans. Tehran: chapar; 2011. [In Persian]
2. Hakuta K. Prefabricated patterns and the emergence of structure in second language acquisition. Language Learning 1974; 24(2): 287-97.
3. Nattinger J, DeCarrico J. Lexical phrases and language teaching. Oxford: Oxford University Press; 1992.

4. Wang J, Liang Sh, Ge G. Establishment of a medical academic word list. *English for Specific Purposes* 2008; 27(4): 442–58.
5. Wray A. *Formulaic language and the lexicon*. Cambridge: Cambridge University Press; 2002.
6. Erman B. Cognitive processes as evidence of the idiom principle. *International Journal of Corpus Linguistics* 2007; 12(1): 25–53.
7. House J. Developing pragmatic fluency in English as a foreign language. *Studies in Second Language Acquisition* 1996; 18(2): 225–52.
8. Schmitt N, Grandage S, Adolphs S. Are corpus-derived recurrent clusters psycholinguistically valid? In Schmitt N, Eds, *Formulaic Sequences: Acquisition, Processing, and Use*. Amsterdam: John Benjamins Publishing; 2004.
9. Hyland K. Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics* 2008; 18 (1): 1-9.
10. Altenberg B. On the phraseology of spoken English: the evidence of recurrent word-combinations. In Cowie AP, Eds, *Phraseology: theory, analysis and applications*. Oxford: Oxford University Press; 1998.
11. De Cock S, Granger S, Leech G, & McEney T. An automated approach to the phrasicon of EFL learners. In Granger SEds, *Learner English on computer*. London: Longman; 1998.
12. Biber D, Barbieri F. Lexical bundles in university spoken and written registers. *English for Specific Purposes* 2007; 26(3):263–86.
13. Cortes V. *Lexical bundles in academic writing in history and biology*. [Thesis]. United States; Northern Arizona University; 2002.
14. Firth JR. *Modes of meaning*. Oxford: Oxford University Press; 1951.
15. Cortes V. Lexical bundles in published and student disciplinary writing: examples from history and biology. *English for Specific Purposes* 2004; 23(4): 397–23.
16. Hyland K. As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes* 2008; 27(1):4–21.
17. Biber D, Johansson S, Leech G, Conrad S, Finegan E. *The Longman grammar of spoken and written English*. London: Longman; 1999.
18. Salem A. *Pratique des segments re'pe'te's*. Paris: Institut National de la Langue Franc,aise, 1987; In Tremblay A, Derwing B, Libben G, Eds, *Are lexical bundles stored and processed as single units? Proceeding of the 23rd Northwest Linguistics Conference; 2007, Victoria BC CDA*.
19. Scott M. *WordSmith Tools version 5*. Liverpool: Lexical Analysis Software; 2008.
20. Martinez I. Aspects of theme in the method and discussion sections of biology journal articles in English. *Journal of English for Academic Purposes* 2003; 2(2): 103–23.
21. Parvizi N. *Identification of discipline-specific lexical bundles in education*. [Thesis]. Kashan: Iran, University of Kashan; 2011.
22. Valipoor L. *A corpus-based study of words and bundles in chemistry research articles*. [Thesis]. Kashan: Iran, University of Kashan; 2010.
23. Jalali H. *Lexical bundles in applied linguistics: variations within a single discipline*. [Thesis]. Isfahan: Iran, University of Isfahan; 2009.
24. Marco MJ. Collocational frameworks in medical research papers: a genre-based study. *English for Specific Purposes* 2000; 19(1): 63-86.
25. Biber D, Conrad S, Cortes V. If you look at lexical bundles in university teaching and textbooks. *Applied Linguistics* 2004; 25 (3): 371–405.

Content Analysis of Medical Research Articles: A Corpus-based Study

Maryam Salami¹, Zahra Sadat Jalali², Maryam Pakdaman Naeini³,
Mohammad Alaei Arani⁴

Original Article

Abstract

Introduction: Content analysis applied for comparing texts, a writer's works with others or the same author, the texts written in a natural language with texts in other languages literature and review of specialized field. The present study aims to analyze the articles to extract the sentences structure in order to support medical researchers.

Methods: This study investigates 4-word lexical bundles in medical research articles. To this end, a corpus of 2,400,000 words was established based on 33 subject areas in medicine and frequency of 20 and distribution numbers of at least 5 were considered as criterion. Word smith tools 5, Antconc3.23, ABBYY Fine Reader to professional edition were used for the identification of lexical bundles. Then, identified lexical bundles were classified according to their functions and structures in the medical research articles.

Results: Totally, 102 lexical bundles with different structures and functions were identified in medical articles. The findings of the current study showed that most of the medical writers use text-oriented bundles in their articles.

Conclusion: This study presents new insights in better understanding of medical research articles related to specific area and further researches in written academic discourse and English for Academic Purposes. It also broadens new horizon on areas of text analysis and text statistical assessment.

Keywords: Analysis; Computerized linguistics; Journal Article; Corpus .

Received: 29 Sep, 2014

Accepted: 10 Jun, 2015

Citation: Salami M, Jalali ZS, Pakdaman Naeini M, Alaei Arani M. **Content Analysis of Medical Research Articles: A Corpus-based Study**. Health Inf Manage 2015; 12(5):607.

*- This article was resulted from an Independent Research.

1- Assistant Professor, Information science, Payame Noor University, Mashad, Iran

2- MA, English education, University of Kashan, Kashan, Iran

3- Ph.D, Candidate of Information Science, Payame Noor University, International Institute of Earthquake Engineering and Seismology (IIIES), Tehran, Iran

4- Ph.D, Candidate of Information Science, Payame Noor University, University of Kashan, Kashan, Iran (Corresponding Author)
 Email:alae62@gmail.com