

اطلاع‌رسانی:

نشریه فنی مرکز اسناد و مدارک علمی (مرکز اطلاعات و مدارک علمی ایران)

دوره چهارم شماره ۲۱

تجزیه و تحلیل مدارک و نظریه زبان‌شناسی^۱

نوشته: ژان کلود گاردن

ترجمه: آنوش هارطونیان

در این مقاله از یکی از جنبه‌های دانش اطلاع‌رسانی یعنی تجزیه و تحلیل مدارک و رابطه آن با نظریه زبان‌شناسی سخن می‌رود. منظور از "تجزیه و تحلیل مدارک"^۲ در این مورد بخصوص استخراج مفاهیم از مدارک نوشته شده است. البته مدارک نوشته شده را نباید منحصر به "مدارک علمی نمود، بلکه سایر مدارک مانند متون - مدارک تاریخی - اساطیر - فرهنگ عامه - نوشته‌های مذهبی و مانند اینها نیز در این رده مورد توجه هستند. منظور از ارائه این مقاله در درجه اول بحث درباره مدارک علمی است و در درجه دوم اشاره به گروه‌های مختلف متون تا آنجا که در حدود معین به روش تحلیل مدارک مربوط می‌شوند.

در مورد استخراج مفاهیم از مدارک، که قبلاً ذکر شد، هر کس مجاز است که برای چنین منظوری نظری را که خود با روش موجود مناسب میدانند بپذیرد، که خود سبب بی‌نظمی می‌گردد. این بی‌نظمی بعلت آن است که بیشتر روشهای مشروحه تحت عنوان وسیع آماده‌سازی مدارک یا ارائه مدارک، در دانش اطلاع‌رسانی، مانند زمینه‌های دیگر، بصورت یک وسیله جستجوی مفهوم در نظر گرفته می‌شود. بعنوان مثال این موضوع در روشهای آماری نیز صدق می‌کند، روشی که در آن

¹ Jean- Claude gardin. Document analysis and linguistic theory. Journal of documentation, volume 29, number 2, June 1973. p. 137- 163

² برای بررسی و تحقیق وسیع تری در زمینه "زبان‌شناسی و دانش اطلاع‌رسانی" می‌توان به منابع زیر رجوع کرد: مقاله ای تحت همین عنوان و بوسیله C. Montgomery (1972) - همچنین مقاله‌هایی که توسط M. Kay و K. Sparch jones (1971) نوشته شده است و نیز گزارشی که بوسیله همین نویسنده برای F.I.D چاپ شده است. مقاله دیگری نیز بزبان فرانسه و در همین زمینه توسط M.Coyaud (1972) نوشته شده است.

³ تجزیه و تحلیل مدارک در زبان فرانسه "Analyse documentaire" است که در اینجا سعی شده است در قالب همان معنی بکار برده شود. این نام از اسم گروهی که از پانزده سال پیش شروع بکار نموده اند یعنی گروه "entre d`analyse documentaire pour l`archeologie" اخذ شده است.

حالت های مختلف (مانند تعداد دفعات، فاصله آماری و غیره) بمنظور آشکار ساختن مفاهیم چند کلمه معنی دار یا چند جمله از یک مجموعه بکار می آید. این روشها قبل از اینکه در سالهای ۶۰ در دانش اطلاع رسانی متداول شود در روان شناسی، ادبیات، جامعه شناسی و غیره مورد استفاده قرار می گرفت. البته روشهای مذکور را اکنون می توان به اعتباری بعنوان قسمتی از وسایل تحلیلی زبان شناسی تلقی کرد، چنانکه امروز "زبان شناسی آماری" یکی از کاملترین رشته های علوم زبان شناسی بشمار می آید. در این جا از اشاره به "معنا شناختی آماری" بخاطر پرهیز از مشکلات و نقایصی که برای منظور ما پیش خواهند آورد چشم پوشی می شود.

باری موضوع مورد بحث نویسنده را می توان بنحو صریحتر چنین بیان نمود: بررسی همکاری دانش زبان شناسی با کیفیت روشهایی که در آماده سازی مدارک در دانش اطلاع رسانی از آنها استفاده می شود. برای سهولت کار می توان این روشها را به دو گروه وسیع قسمت کرد: (الف) روش های طبقه بندی، که در آن حاصل تحلیل مدارک، فهرستی است از واحدهای زبان شناختی که از متون زبان طبیعی استخراج و به روشهای مختلف مرتب شده اند (مانند نمایه ها و هماهنگ سازها). (ب) روشهای تبدیلی، که در آن متون زبان طبیعی (مدارک یا پرسشها) موضوع عملیات پیچیده تری هستند مانند نمایه سازی - تجزیه و تحلیل معنایی و ترکیبی - ارائه آنها در زبان اطلاع رسانی و غیره. مفهوم این دودستگی برای متخصصان دانش اطلاع رسانی آشنا است و احتیاج به تفسیر ندارد، همینطور احتیاجی نیست که تاکید شود که روشهای ترکیبی که متشکل از روشهای طبقه بندی و تبدیلی است، معمول ترین روشها هستند.

۱- نمایه ها و هماهنگ سازها

ابتدا می پردازیم به بیان ساده ترین روشهای نمایه سازی: عناصر (غالباً کلمات) در زبان طبیعی از مجموعه موجود انتخاب و بنحو مشخصی مرتب می گردند، بعد به واحدهای قرینه (از یک نوع یا انواع مختلف) تقسیم می شوند (مانند عبارات، جمله ها و غیره) و سرانجام بصورت طبقه بندی شده که غالباً در بردارنده یک دستگاه ارجاعی برای واحدهای گشوده تری در مجموعه است تظاهر میکند (مدارک، صفحات، پاراگرافها و غیره). روش و نتیجه کار هر دو مشخص هستند و بعنوان نمونه می توان نمایه های گردان^۱ را جزو این گروه ذکر کرد. در اینجا مساله قابل توجهی که مطرح میشود اینست که آیا زبان شناسی در توسعه قواعد اطلاع رسانی مثمرتر بوده است یا نه؟

۱-۱ با مطالعه مراحل ذکر شده در بالا، بشرح کوتاهی در زمینه مساله انتخاب می پردازیم. باید گفت راه حل هائی که در دانش اطلاع رسانی مورد استفاده قرار می گیرد بستگی خیلی کمی با زبان شناسی دارد. قبل از هر چیز باید واحدهای دارای معنی و واحدهای فاقد معنی را از هم تشخیص دهیم. بعد ببینیم در زبان طبیعی کدام بیان رساننده مقصود شناخته می شود یعنی کدام مدخلها در یک نمایه موجود ضرور و کدام غیر ضرور است. جواب این سؤال به میزان رسابودن نمایه، مقصود نمایه ساز و همچنین به خود مجموعه بستگی دارد. در این زمینه، در

^۱ KWIC

زبان‌شناسی، روش خاصی موجود نیست، مگر آنکه استفاده از فهرست کلمات غیرمجاز^۱ را بعنوان یک روش زبان‌شناسی تلقی کنیم. که در این صورت زبان‌شناسان هر زبان، فهرستی استاندارد شده از لغاتی که نباید بعنوان مدخل‌های نمایه بکار رود تهیه می نمایند. به چنین فهرستی نه تنها باید حروف ربط، قیود، حرف اضافه و نظایر آنرا اضافه کرد، بلکه تعداد زیادی از اصطلاحات که در عبارات علم معانی وجود دارد و برای استفاده کنندگان از فهرست در درجه اهمیت کمتری قرار دارند، بآنها اضافه شود. طبیعی است که فهرست کلمات غیر مجاز تهیه شده توسط دانشمندان دانش اطلاع رسانی یک فهرست تجربی و دور از نظریه های زبان شناسی است.

با توجه به مرحله دیگری از انتخاب، نظریه مشابهی می توان اظهار نمود: بعضی کلمات بهتر است به تنهایی بکار برده شوند، ولی در موارد دیگر لازم است که یک یا چند واژه بصورت ترکیبی بکار روند تا گویای مفهوم معینی باشند. طبیعی است که در حالت اخیر جابجائی واژه هائی که در ترکیب بکار رفته اند موجب تغییر مفهوم و پیدایش مفاهیم جدید خواهند شد. یک چنین زمینه ای اساس کار واژه سازی است. درباره لغت نامه - اصطلاحات - اختصارات و غیره مطالب زیادی نوشته شده است که در مسائل علمی وابسته به نمایه سازی در دانش اطلاع رسانی می تواند مفید باشد. ولی به هر حال نمی توان باین امر اکتفا کرد. از آنجا که نمایه های مربوط به مدارک منحصر^۲ نمایه هائی هستند سوگرفته و مشخص در یک زمینه خاص، تصمیم گیری درچنین مواردی باز بر اساس ملاحظیات تجربی و جدا از روشهای پیشرفته فرهنگ نویسی خواهد بود.

سومین مرحله انتخاب، گزینش است. در این مرحله انتخاب واژه ها بشکل طبیعی آن بعنوان مدخل‌های نمایه انجام می شود. نمونه هائی از آنرا می توان در دانش اطلاع رسانی یافت. در گام اول، روش تبدیل از اشکال به پایه ها بطور وضوح با تجزیه و تحلیل ریخت شناختی^۳، همانطور که در زبان شناسی نظری بکار برده شده است، مرتبط است. معهدا لازم به یادآوری است که با وجود اینکه متخصصان تجزیه و تحلیل ریخت شناسی در سده شانزدهم برای هر زبان علمی مانند زبانهای انگلیسی - روسی - آلمانی - فرانسه - و به مقاصد مختلف (مانند ترجمه ماشینی، لغت نویسی، تجزیه و تحلیل ترکیبی و غیره) قواعدی به دست داده اند، باز بنظر نمی رسد که این قواعد بتوانند بعنوان قواعد استاندارد شده جهت نمایه سازی در دکومانتاسیون در هر یک از زبانهای فوق بحساب آیند.

۲-۱ بعد از انتخاب مدخل نمایه ها، مساله طرز قرار گرفتن آنها بصورتی که کار تحقیق را آسانتر نماید مطرح می شود. آسانترین راه، روش الفبائی کردن آنهاست. این روش تا حدی در

¹ Stop-list (فهرستی است از کلماتی که بعلت فقدان ارزش اطلاعاتی در نمایه بکار نمی روند).

² Morphological

زبان شناسی یک روش ابتدائی است. بعضی نمایه ها از الگوی دیگری که بر پایه روابط معنایی استوار است متابعت می نمایند. رده بندی مدخلها از روی ترکیب معانی صورت میگیرد، مانند واژه نامه مفهومی^۱ که شناخته شده است، و یا واژه نامه های زبان طبیعی^۲.

پایه های نظری واژه نامه های زبان طبیعی هنوز کاملاً ساخته نشده اند و نمی توان ادعا کرد که ترتیبات معنایی اصطلاحات زبان طبیعی مستخرج از مدارک علمی از لحاظ زبان شناسی دارای ارزش خاصی است. البته روشهایی پیشنهاد شده است که هدف اساسی چنین گروه ها و طبقه بندیهایی را در بر می گیرد. این روشها از سوئی به نظرات و تحقیقاتی که زمانی بوسیله هریس^۳ متداول شده بود و تا اندازه ای خارج از بحث زبان شناسی است متکی است، و از سوئی دیگر به مقیاسهای آماری مانند آنچه در طبقه بندی خودکار مورد استفاده قرار میگیرد. البته طبقه بندی خودکار رابطه ای غیر زبان شناختی با سمانتیک دارد.

۳-۱ بعد از انتخاب و طبقه بندی مدخلهای نمایه، مساله رده بندی مفاهیم و طریقه ارائه آنها جهت استفاده کنندگان مطرح میشود. در این مرحله نمایه ساز از میان اصطلاحات بشماری که مربوط به متن می باشد نسبت به تقدم و تاخر آنها تصمیم میگیرد. در این صورت است که دو مساله پیش میآید: (۱) کدام اصطلاح با توجه به ضرورت دانش اطلاع رسانی بهترین است؟ (۲) و آیا از روشهای زبان شناسی می توان برای رده بندی مفاهیم زبان طبیعی از چنین اصطلاحاتی استفاده کرد؟ جواب سؤال (۱) از طرفی به ذوق و سلیقه و از طرفی دیگر به تشخیص عملی بودن آن بستگی دارد. در قسمت (۲) بویژه مواقعی که برای اصطلاحات اختیار شده محدودیت قائل باشند امکان انتخاب کم میشود و این بدین جهت است که گر چه زبان شناسان هنوز از مفاهیم: جمله، عبارت، بند و غیره استفاده می کنند معهدا در مورد وجود چنین اختلافات آشنائی کمتری دارند و بطور یقین این بیانی است از این حقیقت که: (۱) متخصصان دانش اطلاع رسانی اصطلاحات صریح را بیشتر ترجیح میدهند، نظیر عناوین مدارک و جملات چکیده های تلگرافی. (۲) نکته دیگر اینکه بعلت عدم روابط مشخص مجبورند بر اساس روابط تجربی تصمیم گیری نمایند.

۴-۱ در هیچ یک از سه بحث فوق بنظر نمی آید که زبان شناسی بتواند کمک زیادی به دکومانالیست ها بنماید. قواعد پیشرفته تجزیه و تحلیل متون در زبان شناسی بطور کلی بمنظور آمادگی قدرت درک یک شخص در تکلم و فهم زبان مشخص است. در صورتیکه هدف نمایه سازها در دانش اطلاع رسانی بیشتر سودمند نمودن نتیجه سخن رانی ها و نطق ها در جهان مباحث است تا کیفیات دیگر آنها. دستورالعملها و یا قواعدی که بر حسب روابط تجربی بوجود آمده اند نه تنها در

¹ Conceptual dictionaries

² رجوع شود به: (Roget for English, watburg and halling for French, etc.)

³ Harris

این مورد پذیرفته هستند بلکه بدلیل صرفه جوئی در وقت توصیه هم می شوند. در این مورد اصول نظری مبنی بر متابعت دستورالعملها از قواعد دستوری زبان طبیعی وجود ندارد.

۲- تجزیه و تحلیل محتوی

در موارد مختلف این بحث پیش آمده که دانش اطلاع رسانی جز آنکه فعالیت خود را محدود به نمایه سازی و هماهنگ کردن آنها نماید، به تنهایی نمی تواند در زبان طبیعی عمل نماید. یکی از فعالیت‌های اساسی دانش اطلاع رسانی ایجاد ارتباط بین مدرک و پرسش است. این امر مستلزم این است که هر دو قسمت (مدرک و پرسش) با در نظر داشتن مناسبات معنایی و صور مختلف آن در زبان طبیعی دوره نوشته شوند. براحتی میتوان نشان داد که جهت انجام فعالیت‌های مربوط به مدارک اطلاع رسانی در زبان طبیعی (مدارک) و همچنین جهت برآورده ساختن نیازهای اطلاعاتی زبان طبیعی (پرسشها)، داشتن دانشی مشابه آنچه در بالا ذکر شد لازم است. از جمله قواعد شناخته شده ای که دکومانتالیست ها مورد استفاده قرار میدهند، طبقه بندی - اصطلاح نامه ها - واژه نامه و غیره. سئوالی که مطرح می شود این است که آیا زبان شناسی نقش مهمی در توسعه این قواعد داشته است یا نه؟

۱-۲ قواعد معنایی

قبل از هر چیز به معنای وسیعی که عنوان بالا دارد اشاره می کنیم. در کنار "قواعد معنایی" لغاتی کم و بیش اختصاصی نظیر: هم معنایی، چند معنایی و غیره را می توان نام برد. این قواعد مدت زیادی متداول ترین قواعد طبقه بندیها به توسعه دانش، به گسترش شبکه ارتباطات جهانی و تا اندازه ای نیز به خود واژه ها (سرعنوان ها و غیره) بستگی دارد. با اختصاصی شدن نوع طبقه بندیها، اصطلاحاتی که برای در برگرفتن طبقه بندیها و فهرست کردن مورد استفاده قرار گرفت و مفهوم وسیع "زبان اطلاع رسانی" برای در برگرفتن طبقه بندیها و فهرست اصطلاحات و توصیف گرها بوجود آمد. بعدها مساله ارتباط زبان طبیعی با زبان اطلاع رسانی پیش آمد که واژه نامه^۱ جواب پیشنهادی این مساله بود که با ارتباط های ارجاعی مشخص می نمود کدام شاخه فرعی از لغات طبیعی باید در زبان اطلاع رسانی بکار برده شوند. رده ها و زیر رده ها در روشهای سیستماتیک و توصیف گرها در اصطلاح نامه ها بعنوان نشانه هائی هستند که مفاهیم زبان طبیعی را به اشکال مختلف نمایان می سازند.

روابط معنایی در دانش اطلاع رسانی ایجاد دو مساله می نماید: (۱) مسائلی که مربوط به انتخاب مفاهیم لغوی است و باید در فرهنگ معنایی گنجانده شود، (۲) طرز برقراری ارتباط بین مفاهیم ذکر شده در قسمت (۱). سئوالی که پیش می آید این است که آیا زبان شناسی نظری برای متخصصان دانش اطلاع رسانی که با دو مساله فوق مواجه هستند مفید بوده است؟

^۱ Thesaurus

مسائلی که در مورد نمایه سازی و هماهنگ نمودن آنها پیش می‌آید نظیر مسائل و مشکلات ذکر شده در بالا است. انتخاب مدخلهای فرهنگ لغات بیشتر بر اساس غیر استدلالی است تا از روی روابط قراردادی. در قسمت دوم این مساله روشهایی جهت استخراج روابط معنایی از اصطلاحات زبان طبیعی از طریق مطالعه وضع این اصطلاحات در یک مجموعه موجود می باشد. نقائصی که بر این روشها وارد میشود عبارتست از: بی ثباتی تمام روشهای توزیع (اعم از نظری و عملی) در زبان شناسی. دستور العملهایی که برای تالیف واژه نامه در دانش اطلاع رسانی تهیه شده چندان از قواعد نظری زبان شناسی پیروی نمی نمایند.

۲-۲ قواعد دستوری

متخصصان تجزیه و تحلیل روشهای گوناگونی جهت نشان دادن وجود یا طبیعت روابط نحوی که بین همه و یا پاره ای از اصطلاحات وجود دارد، بکار برده اند. با وجود اهمیتی که دانش اطلاع رسانی در این زمینه دارد، عده ای از نویسندگان در تمیز بین محتوی و شکل موفق نبوده اند. در اینجا لازم است که به دو نکته که هر کدام پرسشی را طرح می کنند، اشاره کنیم: (۱) با توسل به اصطلاحات و رعایت روابط ترکیبی بین آنها می توان درک مفاهیم اصلی یک مدرک را که گاه ممکن است نارسا باشد تکمیل نمود. (۲) توجه به قواعد مربوط به تجزیه و تحلیل مفاهیم از دیدگاه زبان شناسی به آماده سازی مدارک کمک نسبی می نماید.

هر دو فرضیه هنوز کاملاً برای کسانی که دست اندر کار تجزیه و تحلیل مدرک هستند روشن نشده است. ارتباط نامحدود کلمات کلید و توصیف گرها تنها راه پرداختن به تحلیل مدارک نیست. از سوی دیگر، همانطور که از این مقاله می توان پی برد، قواعد و دستورهای گمراه کننده ای در زبان طبیعی هست که خیلی بندرت بطور علمی در تجزیه و تحلیل مدارک از آنها استفاده میشود. این اظهار نظر یا ادعا بدون شک بتا مخالفت بسیاری از نویسندگان دستور زبان و قواعد مواجه خواهد شد. در زیر بطور اختصار به چند شاهد اساسی در این زمینه می پردازیم.

اگر چه روش های اطلاع رسانی به اطلاعات ترکیبی اهمیت میدهند لیکن قواعد زبان طبیعی، حتی قدیمی ترین آنها، نمی تواند چهار چوب مناسبی برای این منظور باشد. تنها دانستن این مطلب که بسیاری از روش های اطلاع رسانی جدید برای قواعد نظری اخیر مفید نمی باشند کافی نیست، بلکه دانستن این مطلب نیز مهم است که هیچ کدام از روش های قرار دادی تجزیه و تحلیل مدارک که در صد سال اخیر بکار رفته است، در استفاده از قواعد دستوری موجود در زبان طبیعی زمان خود جهت روابط ترکیبی مناسب نبوده اند.

این موضوع را صرفاً نمی توان یک پیشامد دانست و کوتاهی در تجزیه و تحلیل مدارک و استفاده اندک از آن را مورد انتقاد قرار داد. نمونه های متعدد و متناقض از ساختمان های ارتباطی وجود دارد که استفاده از آنها به همان اندازه که برای علمای دستور غیر منتظره است در کار تهیه متن نیز نامناسب می باشند.

در واقع قواعد دستوری تجزیه و تحلیل مدارک رابطه نزدیک تری با مفاهیم منطقی دارند تا با زبان شناسی محض معمول ترین این قواعد عبارت از نشانه هائی هستند که در تجزیه و تحلیل مدارک خصوصاً با بکار گرفتن اصطلاح نامه ها از آنها استفاده می شوند مانند نشانه هائی که برای تعیین نوع ارتباط چند توصیف گر و یا جنبه های مختلف آنهاست. این نشانه ها مسلماً دور از قواعد تجزیه و تحلیل نمی باشند ولی متمم صریحی نیز چه در گذشته و چه در زمان حال برای قواعد زبان شناسی بحساب نمی آیند. بعبارت دیگر هنوز قواعد دستوری در هیچ زبانی، فهرست کامل و قابل توجهی در زمینه های مختلف مثلاً: یک رابطه نسبی بین "مقدمه" X و "نتیجه" Y بوجود نیاورده است و نظریه پردازان سبک های نحوی و ترکیبی را که در ساختن قواعد زبان شناسی دارای اهمیت بسیار و در تجزیه و تحلیل مدارک دارای اهمیت کمتری است آشکار نمی سازند.

۳- روش و نحوه تجزیه و تحلیل مدارک

۱-۳ غرض از این مبحث توجه به سه جنبه است: (۱) تعریف تجزیه و تحلیل مدارک که بمعنی استخراج مفاهیم از متن است، مستلزم مرجعی است که از متن گرفته می شود و بوسیله نمادهائی^۱ که حتماً لازم نیست در متن وجود داشته باشد معین و معرفی میگردند یعنی مفاهیم و مناسبات مشترک آنها، بوسیله نمادهای غیر مشخصی (مانند توصیف گرها، نشانه های ارتباط و غیره) معین میگردند. باین نمادها که برای بیان محتوی بکار میروند در اصطلاح Metalanguage^۲ گفته می شود. گر چه اختیار نمودن واژه ای که خود دارای معنی مشخص و خاصی است خالی از اشکال نیست، لیکن بیان مناسب دیگری که معرفی کننده این دسته از نمادها باشد وجود ندارد که هم خارج از زبان مدرک مورد تجزیه و تحلیل باشد و هم رساننده مفهوم مدرک در جهات مختلف و به مقاصد متفاوت نیز باشد، بعنوان مثال: بازیابی اطلاعات - طبقه بندی و غیره در دانش اطلاع رسانی، مطالعات مربوط به فرهنگ عامه و تفسیرها).

[ابتدای صفحه](#)

^۱ Symbols

^۲ رجوع شود به: (Gardin, 1969)