# تهیه بانک پرحجم واژگان برای کاربرد در نظام های متن سازی و سازه یابی و بازیابی اطلاعات

**اطلاع رسانی شماره ۲ دوره ۱۲ زمستان ۱۳۷۵**

**نویسندگان: اس. پی. ان. کانلن ایونز؛ تی.اهلسود؛ آر. استراتز***

**ترجمه: سید مهدی سمائی**

**مرکز اطلاعات و مدارک علمی ایران**

**کلیدواژه ها:**

☐ **بازیابی اطلاعات** ☐ **بانک واژگان** ☐ **پردازش زبان**

☐ **سازه یابی** ☐ **متن سازی** ☐ **هوش مصنوعی**

**چکیده:**

در این مقاله ثابت می شود که می توان با ترکیب مطالب تعدادی منبع ماشین خوان (machin-readable) بانک واژگان ساخت. همچنین درباره اطلاعاتی بحث می شود که در واژه هست و در بازیابی اطلاعات[1] (information retrieval) و دیگر حوزه های پردازش زبان نظیر رابطه های بانک اطلاعات (database interfaces) و نظامهای خودکار برگه آرائی به کار می رود. ساختار بانک واژگان که در نظام مدیریت بانک اطلاعات رابطه ای اراکل (Oracle Relational Database) (Management System) طرح جدولهای موجود در بانک مذکور نیز توصیف خواهد شد. دستیابی به بانک علاوه بر فهرست الفبائی سنتی از راه ریشه به مشق و بر عکس و از راه رابطه واژگانی و معنائی کلمه ها نیز میسر است به طوری که بانک هم به صورت تزلر روس و هم به شکل فرهنگ لغت به کار می رود. بانک مورد نظر دارای فهرستی بی انتها (open-ended) و تعریف دار (Self-defining) است. ویژگی واژه های هر جدول در خود بانک نیز مشخص شده است. این بانک را می توان به راحتی و با یک رابط (SQL) که افزایش واژه های جدید به جدولها را آسان می کند گسترش داد.

**۱- مقدمه**

محققان با پر حجم شدن بانکهای مدارک و پیچیده شدن نظامهای بازیابی اطلاعات در جستجوی راههایی هستند که به استفاده کنندگان مبتدی در یافتن اطلاعات کمک کنند. تجربه ساختن سازمان تزاروس های حجیم برای کمک به نظامهای بازیابی اطلاعات روز به روز بیشتر می‌شود. محققانی نیز که درصدد ساختن نوعی نظامهای بازیابی اطلاعات هستند که قادر به سازه‌یابی (parsing) و فهم پرسش استفاده کنندگان و متون باشند. تلاش دیگر محققان در ساختن نظامهای اطلاع رسانی دفاتر است که به خوبی قادر به فهم نامه های اداری و بایگانی خودکار و حتی دادن جواب مناسب به آنها باشند. رابطه های بانک اطلاعات و نظامهای تجاری نیز هر روز پیچیده تر می شود. منظور از همه این اقدامها به وجود آوردن راههای مختلف پردازش زبان بشر (natural language processing) است.

* - Conlon, S. P. N., Evens, M., Ahlswede, T., Strutz, R. (1993). Developing a large lexical database for information retrieval, parsing, and text generation systems, Information processing and Management, vol 29, no 4, pp. 415-431.

نظامهای پردازش زبان که قادر به کار در عالم واقع هستند باید توانائی پرداختن به مسائل گوناگون را داشته باشند. این نظامها باید ترکیبهای بدساخت (ill-formed) و مبهم را تشخیص دهند و آنچه را که برای ارائه برون دادی پویا لازم است دارا باشند. برای حل این مسائل نیاز به اطلاعات صریح و مفصل و مبسوطی درباره کلمات و عبارات داریم. نظامهای موثر پردازش زبان در عالم واقع نیاز به بانکهای پرحجم و پیچیده واژگان دارند.

با اینکه بانک واژگان بخش اصلی نظامهای پردازش زبان است در واژگان نظامهای رایج پردازش زبان اطلاعات محدودی موجود است. بدیهی است که چنین واژگانی فقط نیاز بازیهای کامپیوتری یا برخی حوزه های خاص را برآورده می کند. نظامی که بتواند زبانهای متعدد را پردازش کند نیاز به اطلاعات گسترده درباره معنی کلمات و رابطه بین آنها و قواعد ترکیبشان دارد. چنین اطلاعاتی باید چنان واضح و منظم سامان داده شود که نظامهای پردازش زبان بتوانند به اطلاعات نحوی و معنایی کلمات پی ببرند و از آن برای سازه یابی و متن سازی (text generation) و بازیابی اطلاعات استفاده کنند.

از میان مشکلات گوناگونی که در بازیابی اطلاعات و ساختن و تحلیل نحوی متن هست ابهام مشکلات خاصی ایجاد می کند. بسیاری از مشکلات سازه یابی به این دلیل ایجاد می شود که تعداد زیادی از کلمات زبان انگلیسی چند معنی دارد. برنامه سازه یاب قادر به انتخاب معنی مناسب نیست مگر آنکه اطلاعات واضحی از معنی کلمات در پایگاه اطلاعات باشد. ابهام انتخاب واژه برای متن سازی را هم پیچیده می کند. حل این گونه مشکلات نیاز به داشتن اطلاعات مفصّل از معنی کلمات دارد. طبقه بندی کلمات و محدودیتهای باهم آیی (Selectional restrictions) یا رجحان (preferences) و روابط واژگانی و معنائی نیز جزئی از معنی کلمه است.

پیش بینی می شود که واژگان نیاز به اطلاعاتی بسیار پیچیده دارد تا نظامهای پردازش زبان بشر در عالم واقع را تقویت کند. این مقاله درباره ساختن بانک واژگان برای کاربرد در پردازش زبان است. بخش ۲ در خصوص هدف تحقیق و انواع اطلاعات واژگانی لازم برای کاربرد در شیوه‌های گوناگون پردازش زبان است. از آنجائیکه اطلاعات دادگان ما از منابع مختلف است در بخش ۳ درباره انواع اطلاعاتی که از این منابع فراهم شده بحث می کنیم و نیز روشهایی را مورد گفتگو قرار می‌دهیم که برای تبدیل انواع اطلاعات به شکل واژگان اختراع کرده ایم و می تواند مورد استفاده نظام پردازش زبان باشد. بخش ۴ اصول طراحی بانک واژگان شرح داه می شود. در بخش ۵ ساختمان بانک واژگان را توصیف می کنیم و چند نمونه از جدولهای داده ها را ارائه می دهیم و در بخش ۶ توضیحاتی درباره اطلاعات واژگانی و معنائی تهیه شده به دست می دهیم.

در بخش ۷ درباره قابلیتهای بانک بحث می شود و در بخش ۸ بانک واژگان از دیدگاه استفاده کنندگان تشریح می گردد. مقاله با ارائه خلاصه ای در بخش ۹ به پایان می رسد.

## ۲- اهداف تحقیق

ما در کار ساختن یک بانک پرحجم واژگان هستیم که دارای اطلاعاتی صریح راجع به نحو و معنی شناسی باشد. این بانک شامل اطلاعاتی درباره تقریباً ۵۰۰۰۰ مدخل است که هر مدخل خود ممکن است چند معنی داشته باشد. اطلاعات مذکور به اجزاء اصلی کلام تقسیم شده است به صورت جدول اسم و فعل و صفت و قید درآمده است.

این بانک برای استفاده در بازیابی اطلاعات و فهم زبان و متن سازی طراحی شده است. ما در راه رسیدن به هدفمان نیاز به گردآوری و ساماندهی اطلاعات حجیمی داریم که انسان در سیر یادگیری زبان از آن استفاده می کند. در اینجا چند مورد از کاربرد پردازش زبان و انواع اطلاعات واژگانی مورد نیاز آنها ذکر می شود.

**۲-۱- بانک واژگان برای استفاده در بازیابی اطلاعات**

کار اصلی نظام بازیابی اطلاعاتی کمک به استفاده کنندگان است تا موضوع دلخواه خود را در مجموعه های انبوه بیابند. مهمترین کمکی که بانکی پرحجم واژگان می تواند به نظامهای موجود بازیابی اطلاعات بکند در حوزه بهسازی پرس و جو (query enhancement) است.

شیوه کار به این ترتیب است که اصطلاحات فهرست شده مناسب به صورت تزاروسی که دارای اصطلاحات وابسته (related terms) است به پرسش استفاده کننده اضافه می شود. تحقیقات لسک (Lesk 1969) ثابت کرده که افزودن آماری اصطلاحات وابسته سودمند است. از روی تجربه فاکس (Fox 1980) و وانگ (Wang et al 1985) در مورد مجموعه های کوچک نیز ثابت شده که افزودن اصطلاحات وابسته بر حسب معنی سودمندتر است. از تحقیقات مقدماتی فاکس (1980) چنین بر می آید که افزودن اصطلاحات در محیط دو کنشی (interactive) سودمندتر از همه موردهای قبلی است. تزاروس در این حالت اصطلاحات را به استفاده کنند ارائه می کند و استفاده کننده می تواند یک اصطلاح به پرسش خود اضافه یا با فشار دکمه آن را رد کند. برای تهیه این نوع بهسازی پرس و جو برای مجموعه های پرحجم مدارک پرواژه باید اطلاعاتی از روابط بین واژه ها داشت که خود از طرفی نیاز به بانک واژگان پرحجم دارد. بانک واژگان در اصل برای شناسائی چنین اطلاعاتی تدارک دیده شده است. (Ahlswede & Evens, 1988)

هر نظام هوشمند بازیابی اطلاعات که قادر به سازه یابی و فهم پرسش استفاده کنندگان و متن ذخیره شده در نظام باشد باید علاوه بر داشتن اطلاعات درباره روابط بین واژه ها اطلاعات دقیقی نیز از نحو و معنی شناسی واژه ها داشته باشد (Smith, 1987; Waemer, 1987)

اگر کاربرد بازیابی اطلاعات را در مورد بانک واژگان قضائی در نظر بگیریم می بینیم که این نوع اطلاعات دقیق و صریح درباره طبقه وسیع کلمات بیش از بیش ضروری می آید (Maggs & sprowl, 1987)

در نظام بانک واژگان قضائی همه مدارکی که در مرافعات حقوقی و جنائی وجود دارد در یک بانک ذخیره شده که در این مورد بازیابی اطلاعات در کل متن لازم است. گاهی بخشهای معتبر قانون و مرافعات مورد نظر را از مراکز مدارک قانونی نظیر وست لاو (west law) و لکسیس (Lexis) می یابند و به بانک اضافه می کنند. در این گونه نظامها بایگانی خودکار مکاتبات مزایای فراوانی دارد اما برای رسیدن به این هدف نیاز به واژه های بسیاری داریم تا موضوع مکاتبه درک و مدرک در جای مناسب بایگانی شود.

**۲-۲- نیاز به واژگان برای نظامهای فهم زبان**

نظامهای فهم زبان ابتدا باید جمله را سازه یابی کنند تا بتوانند معنی جمله هائی را که استفاده کنندگان به آنها ارائه می دهند بفهمند. سازه یاب نیاز به واژگان دارد که دارای اطلاعاتی صریح درباره حداکثر ممکن کلمات باشد (perrault & Grosz 1986). کامینگ (Cumming) ثابت کرده است که چون استفاده کنندگان کلمات متعددی را برای صحبت درباره هر موضوعی به کار می برند بنابراین نظامهای فهم زبان

نیاز به واژگان گسترده دارند تا هم از نظر ساختمان واژه و هم از لحاظ ساختار نحوی قابل فهم باشد. همچنین این نظامها نیاز به اطلاعاتی درباره املاء و صرف و نحو معنی شناسی دارند. املاء شامل اطلاعاتی در خصوص حروف (graphemes) و خط تیره‌گذاری (Hyphemation) و تلفظ و شکلهای مختلف هر کلمه است. صرف شامل اطلاعاتی درباره صیغگان (paradigms) و روابط تکواژها است. نحو شامل اطلاعاتی راجع به انواع کلمه و ساختمان وابسته ها (argument structure) است. تعاریف و روابط معنائی نیز در معنی شناسی جای می‌گیرند. پنس و بینوت (Jensen & Binot 1987) برای ابهام زدائی عباراتی که دارای حرف اضافه اند, از تعاریف موجود در فرهنگ لغت کامپیوتر استفاده می کنند. محدودیت‌های باهم آیی (selectional restrictions) نیز مهم است. اینها اطلاعاتی است مربوط به اینکه افعال چه فاعل و چه مفعولی را می توانند با هم بیاورند. نظامهای فهم زبان برای انتخاب معنی مناسب کلمات, از محدودیتهای باهم آیی استفاده می کنند (Small et al., 1988).

نظامی که در صدد فهم متن است, باید اطلاعات انبوه تزاروس گونه نیز داشته باشد. نیاز به اطلاعات تزاروس گونه به این علت است که ارجاع جمله های متن به یکدیگر با گروه اسمی انجام می شود و نه مانند متون روائی (narrative) با ضمیر (Blach, 1985). یعنی اینکه بافتار (texture) متن با نوعی رابطه بین کلمات انجام می شود که انسجام واژگانی (Lexical cohesion) نام دارد. (Halliday & Hasan., 1976). موریس و هرست (Morris & Hirst., 1991) ثابت کرده اند که فهم روابط واژگانی به خصوص در فهم متون علمی اهمیت دارد.

### ۲-۳- نیاز به واژگان برای نظامهای متن سازی

به نظر کامینگ (Cumming., 1987) در واژگان نظامهای متن سازی حتی باید بیشتر از واژگانی که برای سازه یابی به کار می رود اطلاعات وجود داشته باشد. مثلاً سازه یابی که به ساختار نحوی گروه اسمی big red Wooden trucks می پردازد فقط احتیاج به تشخیص این نکته دارد که big و red و wooden هر سه صفت trucks است. در حالیکه جمله ساز (generator) ای که اشیاء را توصیف می کند باید بداند که big مربوط به اندازه و red مربوط به رنگ و wooden مربوط به جنس است در نتیجه این صفتها را با هم و با نظم خاصی بیاورد. یعنی اندازه را قبل از رنگ و رنگ را قبل از جنس بیاورد.

به گفته کامینگ نظامهای متن سازی به اطلاعات واژگانی بیشتری از نظامهای فهم متن‌سازی احتیاج دارند. نظامهای فهم متن کلاً به همه اطلاعاتی که برای نظامهای فهم زبان لازم است نیاز دارند. دستگاه جمله ساز باید قادر به انتخاب فعل و وابسته های مناسب آن باشد و بتواند اصطلاحات (idioms) و دیگر کلمه های همنشین (collocations) را به درستی انتخاب کند و قید را در جای مناسب بگذارد. (Jacobsan 1964). نظامهای متن سازی همچنین احتیاج به اطلاعاتی دارند تا بتوانند روابط بین کلمات را برای ساختن متن منسجم نشان دهند و واژه ها را به صورتی درست و پویا انتخاب کنند (Lee & Evens., 1991, 1992)

### ۲-۴- تحقیقات انجام شده در مورد واژگان

اولین پایگاه واژگانی _ که کاربرد اصلی آن هم در ترجمه ماشینی نبود _ فرهنگ ماشینی زبان ایتالیائی بود که کلتزولاری در دانشگاه پیزا آن را تهیه کرد (Calzolari., 1984; 1988). این فرهنگ با کمک فرهنگهای ماشین خوان موجود و صرف نیروی بسیار تدوین شده بود. استفاده از فرهنگهای ماشین

خوان در آمریکا را امسلربنا گذاشت (Amsler., 1981). او اسم و فعل را با استفاده از فرهنگ جیبی مری یا م وبستر (Merriam Webster pocket Dictionary) به صورت ماشینی طبقه بندی کرد. بر دو همکارانش در آی بی ام اطلاعات واژگانی بسیاری را جمع‌آوری کردند و ابزارهای پردازش آنها را فراهم آورند. میلر و همکارانش (Miller et al., 1991) نیز یک شبکه پرحجم از کلمات و روابط آنها به نام وردنت (Wordent) ساختند که براساس روانشناسی زبان و یک منبع واژه گرانبها بود.

عمده ترین تحقیقات درباره واژه ها در انگلستان را چند ناشر دوراندیش انجام دادند. انتشارات لانگمن (longman) اطلاعات نحوی را در فرهنگ انگلیسی معاصر لانگمن به مدخلها اضافه کرد و نسخه های ضبط شده آن را برای تحقیقات زبانشناسان در دسترس قرار داد (procter., 1978). انتشارات کالینز (Collins) با همکاری دانشگاه بیرمنگام یک مجموعه بیست میلیونی کلمه ای از متون انگلیسی تهیه و یک نظام پایگاه واژگانی برای واژه نگارانی تدونی کرد که در صدد مدخل سازی از این مجموعه بودند (Sinclair., 1987) فرهنگ کوبیلد کالینز (Collins COBUILD Dictionary) از همین منبع استخراج شد و خود اساس تدوین یک سلسله فرهنگ دوزبانه شد (Sinclair., 1987) نشر دانشگاهی آکسفورد نیز یک بانک پرحجم ساخت تا مقدمه ای باشد برای ماشینی کردن فرهنگ انگلیسی آکسفورد. بیشتر این کار در دانشگاه واترلو انجام شد (Stubbs & Tompa., 1984; Raymond & Tompa., 1987) مرکز تازه تاسیس فرهنگ انگلیس آکسفورد در دانشگاه واترلو برای شرکت در این کار یک رشته کنفرانس برگزار کرد و زبانشناسان و متخصصان کامپیوتر و ناشران را گردهم آورد تا درباره مشکلات اساسی این حوزه بحث کنند. دانشگاه نایجمگان (CELEX., 1986) در هلند نیز محل تهیه واژگان انگلیسی و آلمانی برای استفاده در تحقیقات فرهنگستانها و صنایع است.

## ۳- منابع اطلاعات واژگان سازی

همانگونه که توضیح داده پایگاههای واژگانی که به پردازش زبان کمک می کنند باید دارای اطلاعات زیادی باشند. برای تهیه پایگاههای واژگانی پیچیده باید انواع دانش را _ از جمله دانشی که حاصل کار زبانشناسان و روانشاسان و متخصصان کامپیوتر و فیلسوفان است _ در این پایگاه درج کرد. ساختار منابع واژگان با یکدیگر تفاوت دارد و باید با روشهایی متفاوت به آنها پرداخت. ساختار منابعی که برای ساختن واژگانی به آنها نیازمندیم در جدول شماره ۱ آمده است.

| نوع گویشور | گروه زبانشناسان | نام فرهنگ |
|---|---|---|
| گویشور بومی | ایندیانا<br>برندیس<br>یاکوبسون<br>دیگران | فرهنگ دانشجوی وبستر<br>فرهنگ معاصر زبان انگلیسی |

جدول شماره ۱- منابع اطلاعات واژگان سازی

### ۳-۱- فرهنگهای ماشین خوان

برخی زبانشاسان که در حوزه زبانشناسی با کامپیوتر تحقیق می کنند فرهنگهای ماشین‌خوان را منبع کار خود کرده اند. در این پژوهش اطلاعات دو فرهنگ ماشین خوان در واژگان ضبط شده است: یکی

فرهنگ دانشجوی وبستر (Webster's Seventh collegiate Dictionary) و دیگری فرهنگ انگلیسی کالینز (Collins Dictionary of the English Language)

این فرهنگها بسیار غنی و اطلاعات متنوعی دارند اما بیشتر برای استفاده انسان تهیه شده‌اند تا ماشین. فرهنگهای مذکور هم دارای اطلاعات تلویحی (implicit) هستند و هم اطلاعات صریح (explicit) در آنها وجود دارد. اطلاعات صریح شامل اجزاء کلام و تلفظ و ریشه شناسی و تکواژگونه و مترادفها و ارجاعات متقابل (cross references) است. نظامهای کامپیوتری اغلب قادرند مستقیماً از اطلاعات صریح استفاده کنند اما استفاده از اطلاعات تلویحی _ از جمله اطلاعات که به صورت تعریف است _ برایشان مشکل است. برای تبدیل دانش موجود در فرهنگهای ماشین‌خوان به صورتهایی که در بانک واژگان قابل استفاده باشد باید آنها را چنان تحلیل و صوری و بی‌ابهام کرد که برای استفاده در ماشین آماده شود. این کار با تحلیل نیمه خودکار انجام شده است. مثلاً تعریف واژه Chip با هم نویسه (homograph) شماره ۱ و معنی شماره ۱ در فرهنگ انگلیسی کالینز به صورت جدول شماره ۲ آمده است.

| |
|---|
| قطعات نازک سیب زمینی که در روغن فراوان سرخ شده باشد<br>Chip 1 1 |

جدول شماره ۲- تعاریف (اطلاعات تلویحی)

پس از تحلیل نیمه خودکار این تعریف می توان اطلاعاتی صریح درباره واژه Chip به دست داد. بنابراین ویژگیهای Chip را چنین می توان آورد:

الف. عینی

ب. شمردنی

ج. بیجان

د. باقاعده

ه. مفرد

و. اسم عام

ز. خنثی (جنس دستوری ندارد)

اطلاعات واژگانی Chip 1 1 در بانک واژگان به صورت جدول شماره ۳ در می آید:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| بیجان | عینی | خنثی | اسم عام | شمردنی | مفرد | باقاعده | Chip 11 |

جدول شماره ۳- اطلاعات واژگانی

می بینیم که فرهنگهای ماشین خوان اطلاعات صریح زیادی را به صورت تعریف فراهم می کنند و اگر قادر به استخراج این اطلاعات باشیم می توان از آنها در پردازش زبان استفاده کرد. جزئیات مربوط به شیوه استخراج نیمه خودکار اطلاعات صریح از تعاریف فرهنگهای ماشین خوان در مقاله زیر چاپ کانلن و ایونز (Colon & Evens., in press) آمده است.

### ۳-۲- دیگر منابع اطلاعات واژگان سازی

به گفته ملک کاولی (MC Cawley., 1986) در فرهنگها اطلاعات مربوط به چگونگی ترکیب کلمات و تشکیل جمله از قلم افتاده است. یکی از مثالهای او واژه "شوهر" است که باید یک گروه اسمی وابسته داشته باشد. مثلاً جمله ای نظیر "جان شوهر مری است" پذیرفتنی (acceptable) است ولی "جان شوهر است" پذیرفتنی نیست. ما در ساختن بانک واژگان اطلاعات را از منابع متعدد زبانشناسی جمع کرده ایم. این منابع شامل فهرست کلمات موجود در طرحها هوس هولدر (Householder) در دانشگاه ایندیانا (Alexandra & Kunz., 1964; Brdgeman et al., 1965; Houeholder et al., 1965) و فهرست افعال گریشماو و جکندوف (Grimshaw & Jacken off., 1985) در دانشگاه برندیس و فرهنگ جایگاه قیود یا کوبسون (۱۹۶۴) است. ساختار برخی از این منابع به اختصار در اینجا توصیف می شود:

#### ۳-۲-۱- فهرست کلمات ایندیانا

ایندیانا چهار فهرست اصلی کلمه دارد که مربوط به چهار بخش اصلی اجزاء کلام (parts of speech) یعنی اسم و فعل و صفت و قید است. هر فهرست چهارچوبی دارد و فهرست کلماتی که می توان آنها را در چهارچوب مذکور به کاربرد در فهرست اسم ایندیانا چهارچوبهائی هست که در آنها جمله های پیرو و فهرست اسمهائی که می توانند در آن چهارچوبها قرار گیرند آمده است. نمونه ای از فهرست اسم ایندیانا در جدول شماره ۴ آمده است.

| |
|---|
| (از چپ به راست) ساخت التزامی + حرف ربط + ساخت مجهول + گروه اسمی<br>مثال: The suggeston has been made that he win. |

جدول شماره ۴- نمونه فهرست اسم ایندیانا

ایندیانا ۷۲ فهرست فعل دارد که شامل ۱۷۱۸ فعل است. در هر فهرست متمم هائی که فعلها می‌گیرند طبقه بندی شده است. مثلاً فهرست شماره ۱ شامل فعلهائی است که منطبق با قاعده جدول شماره ۵ است.

| |
|---|
| جمله خبری + حرف ربط + گروه فعلی + گروه اسمی<br>مثال: John said that his father was coming |

جدول شماره ۵- نمونه فهرست فعل ایندیانا

هفت فهرست هم وجود دارد. صفتهای هر فهرست جمله پیرو خاص خود را می گیرد. فهرست شماره ۱ و ۲ و ۳ شامل واژه هائی است که می توان آنها را در معنی های مختلف با حرف اضافه For به کار برد. فهرست شماره ۴ و ۵ شامل کلمه هائی است که به ترتیب با حرف اضافه IN و On به کار می روند.

در فهرست شماره ۶ و ۷ صفتهائی قرار دارند که با حرف اضافه With می آیند. یک نمونه از فهرست شماره ۳ به این صورت است:

He is responsible for cleaning the house.

چون قواعد قیدها بسیار پیچیده است فهرستهای قید ایندیانا اطلاعات گوناگونی نظیر نمونه زیر دارد:

قید + فعل شخص زماندار (Finite Verb) + اسم فاعل:

چهارچوب D

مثال: He answered Sensibly

### ۳-۲-۲- فهرست فعل برندیس

فهرست فعل برندیس دارای اطلاعات مفصلی درباره زیر مقوله‌ها (Subcategorisation) است. ساختار متمم فعل admire نمونه ای است که در جدول شماره ۶ آمده است.

| ADMIRE VERB DO POSSING THAT DO-P4-ING-OC DO-P4-NP |
|---|
| DO = مفعول مستقیم |
| Possing = We admire his helping مالکیت در جمله ای نظیر ing |
| THAT = متمم زمانبر که با tath به فعل وصل شده است |
| P4 = for حرف اضافه |
| ING = اسم مصدر |
| OC = مفعول فرمانی (۲) |
| NP = نشانه آن دسته گروههای اسمی که در عبارتهای حرف اضافه دار به کار می روند |

جدول شماره ۶ – ساختار متمم فعل admire

دراین جدول نشان داده شده است که مفعول مستقیم فعل admire می تواند گروه اسمی معمولی یا گروه اسمی ing دار باشد. اگر متمم فعل مذکور جمله باشد با حرف ربط that یا for به آن متصل شود. اگر حرف ربط for باشد بند (clause) بعد از آن باید اسم مصدر یا گروه اسمی باشد.

### ۳-۲-۳- فرهنگ جایگاه قیود یاکوبسون

یاکوبسون یا بررسی امکانات همنشنی قید در متون اطلاعاتی را درباره جایگاه قید فراهم کرد. او در فرهنگ خود قیدها را طبقه بندی و جایگاه قید در متون و میزان قرار گرفتن آن را در جایگاه مشخص کرده جدول شماره ۷ اطلاعاتی به شیوه مذکور درباره قید little آمده است.

| LITTLE:قید مقدار ; |
|---|
| مثال : (۷=٪۶۴) جایگاه میانی: This area is little used by day |
| مثال: (4=%36)جایگاه میانی : It Varies little from Government to Government. |

جدول شماره ۷ نمونه ای از فرهنگ یاکوبسون در مورد جایگاه قیود

از این مدخل درمی یابیم که Little یک قید مقدار است و در ۶۴٪ جمله های مورد نظر در جایگاه میانی و در ۳۶٪ آنها در پایان و پس از فعل آمده است. اطلاعات بیشتر در خصوص ویژگیهای معنائی و نحوی قیدها را در مقاله ۱۹۷۸ یا کوبسون می توان یافت (Jacobson., 1978)

اطلاعات فهرست ایندیانا و برندیس و فرهنگ یاکوبسون به خصوص در متن سازی به کار می آیند.

### ۱- اصول طراحی بانک واژگان

قصد ما ساختن یک بانک واژگان دارای اطلاعات انبوه است که لازمه نظامهای بازیابی اطلاعات و دیگر نظامهای پردازش زبان مانند سازه یابها و نظامهای متن سازی و غیره است. در هر بانک کامل واژگان اطلاعات دهها و حتی صدها هزار مدخل باید وجود داشته باشد. بانک واژگان کارآمد باید چنان باشد که انتخاب و یافتن خودکار کلمات مناسب را برای نظامهای پردازش زبان آسان کند. برد (Byrd., 1989) و بوگورو و لوین (Boguraev & Levin., 1990) برخی از ویژگیهای بانک واژگان مناسب را چنین توصیف کرده اند:

**الف.** دستیابی به آن با روشهای گوناگون میسر باشد: نظامهای پردازش زبان بتوانند به صورتهای مختلف کلمات را بیابند یعنی بانک واژگان کارآمد هم باید ساختاری روشن و مشخص نظیر آنچه در بخش پیش توصیف شد داشته باشد و هم بتوان از راههای مختلف به اطلاعات واژه های موجود در آن دست یافت. برای یافتن کلمات در بانک واژگان باید راه‌حلهائی فراهم کرد که هم بر صورت و هم بر معنی استوار باشد. برنامه ها و کسانی که از بانک واژگان استفاده می کنند جستجو را با ریشه کلمات و صورتهای مرکب و برخی صورتهای وابسته به ریشه کلمات آغاز می کنند. اجزاء مختلف کلام و کلماتی که ویژگیهای مشترکی با آنها دارند به گونه ای در جدولهای جداگانه در حافظه کامپیوتر ذخیره شده است که می توان برنامه را دقیقاً روی گره کلمات مورد نیاز متمرکز کرد. بنابراین یک جدول اصلی هست که هر عنصری را می توان به صورت الفبائی در آن پیدا کرد. همچنین جدولهائی برای هر جزء کلام و یک جدول مخصوص برای فعلهائی که متمم جمله ای وجود دارد و الی آخر. وقتی ریشه را به بانک بدهیم بانک می تواند صورتهائی را که از روی ریشه ساخته شده است ارائه کند. بانک همچنین می تواند در صورت ارائه صورت مرکب ریشه آن را بیابد. بانک واژگان دارای جدولهای مربوط به رابطه واژه ها و وابستگی معنائی بین آنها است تا به این وسیله بتوان اصطلاحات وابسته به هر کلمه را یافت.

**ب.** مدخلها از روی معنی جدا شده باشد: معمولاً کلمات به اجزاء مختلف کلام تعلق دارند. کلمه همچنین ممکن است فقط به یک طبقه از اجزاء کلام تعلق داشته باشد ولی چند معنائی باشد. واژه نگاران معانی جداگانه کلمات را با شماره های مختلف مشخص می کنند. دراین بانک واژگان اطلاعات مربوط به معنی کلمات به صورتی جداگانه و واضح آمده است به گونه ای که نظامهای پردازش زبان بتوانند معنی صحیح را انتخاب و آن را به درستی به کار ببرند.

**ج.** برای گروهها (phrases) و اصطلاحات (idioms) مدخلهای جداگانه وجود داشته باشد: برخی مدخلها از چند کلمه ساخته شده است و معنی آنها معمولاً با مجموع معنی تک‌تک کلمات تشکیل دهنده آنها تفاوت دارد. اگر این مدخلها به شکل یک مجموعه کلی و واحد در بانک فهرست نشود سازه یابها و نظامهای متن سازی نمی توانند آنها را تمیزدهند.این دسته مدخلهای موجوددرواژگان را اهلسود (Ahlswede et al., 1988) و مارکوویچ (Markowitz et al., 1988) به تفصیل شرح داده اند.

**د.** قابلیت استنتاج داشته باشد: نظامهای پردازش زبان باید قادر به یافتن روابط معنائی بین مدخلها باشند. بانک واژگان مورد نظر ما هم وابستگی های معنائی را نشان می دهد (اینکه مثلاً جوجه یک نوع مرغ و به خصوص مرغ کم سن و سال است) و هم رابطه ای نحوی را مشخص می کند (نظیر تشخیص این نکته که فاعل فعل خواب دیدن باید انسان و جاندار باشد).

**ه.** قابلیت گسترش داشته باشد: برد (۱۹۸۹) و بوگورو و لوین (۱۹۹۰) ثابت کرده اند که یکی از مهمترین ویژگیهای بانک واژگان نامحدود بودن آنها است. واژگان را باید به صورتی طراحی کرد که در آینده بتوان تعداد واژه ها را به سادگی افزایش داد کیفیت اطلاعات آن را مطلوبتر کرد. گویشوران یا حتی نظامهای خودکار اکتساب واژه (automatic Lexical acquistion systems) باید بتوانند اطلاعات موجود در بانک را افزایش یا تغییر دهند. فرم های اس کیوال برای جمع آوری اطلاعات از گویشوران بومی زبان انگلیسی تهیه شده است تا افزودن مدخلهای جدید به بانک آسان شود (Evens et al., 1989).

## ۵- ساختار بانک واژگان

یک مشکل اصلی نظامهای پردازش زبان چند معنی بودن برخی کلمات است. بنابراین واژگان باید برای هر معنی کلمه اطلاعاتی داشته باشد. وانگهی سازمان بانک واژگان بر اجزاء کلام بنا شده است و با هر جزء کلام انواع گوناگون اطلاعات معنائی و نحوی همراه است. اطلاعات در نظام مدیریت رابطه ای بانک اطلاعات را گردآوری شده است. در جدولهائی که متعاقباً می آید سعی بر این است که نوع داده های ذخیره شده و چند مثال مختصر ارائه شود. اما باید به خاطر داشت که جدولهای واقعی اغلب صدها یا هزاران مدخل دارد. اغلب جدولها کلیدهای مرکبی دارند که از زنجیره کلمات و شماره هم نویسه (homograph) و شماره معنی در فرهنگ انگلیسی کالینز تشکیل شده است. شماره هم نویسه را با علامت (H#) و شماره معنی را با علامت (S#) نشان می‌دهند. جزئیات این کار در رساله دکتری پین نگرن (Pin-Ngern., 1990) وجود دارد.

### ۵-۱- جدول اصلی

جدول اصلی شامل فهرست الفبائی مدخلهاست. هم نویسه ها و شماره معنی در کلمه برای ابهام زدائی در مدخلهای جداگانه آمده است به کلمه هائی که شماره هم نویسی (homography) ندارند رمز (Code) صفر داده شده است. ساختار جدول اصلی در جدول شماره ۸ نشان داده شده است.

| کلمه | شماره هم نویسه H# | معنی شماره S# | مقوله | منبع اطلاعات |
|---|---|---|---|---|
| Away | 1 | 1 | قید | |
| Bake | 1 | 1 | فعل متعدی | |
| Chip | 1 | 1 | اسم | |
| mechanical | 1 | 1 | صفت | |

جدول شماره ۸- ساختار جدول اصلی

در بانک واژگان فقط مدخلهای یک کلمه ای نیست بلکه مدخلهای گروهی نیز در آن وجود دارد. یعنی مدخلهای چند کلمه ای مانند Wide-awake و pick up. در فرهنگ دانشجویی وبستر حدود یازده هزار مدخل اصلی گروهی هست که شانزده درصد کل مدخلهای این فرهنگ است. این ساختهای گروهی جزء به جزء بررسی شده است. (Ahlswede et al., 1988; Markowitz et al., 1988) مثلاً تعاریفی که از مدخلهای گروهی تخصصی در فرهنگ بوده بررسی و از این تعاریف تخصصی هر حوزه استخراج شده است. علاوه بر جدول اصلی برای هر جزء کلام نیز جدول جداگانه آمده است. هر جدول شامل انواع اطلاعات خاص درباره آن جزء کلام است. در اینجا ساختار جدولهای هر جزء کلام به طور خلاصه تشریح شود.

### ۵-۲- جدول اسم

تقسیم اسمها به طبقه های فرعی هم از لحاظ نحوی مهم است و هم از لحاظ معنائی. بنابراین در بانک واژگان چنین ویژگیهائی را باید در مدخل اسمها درج کرد:

**الف.** با قاعده و بی قاعده بودن: صورت جمع اسمهای با قاعده پیش بینی شدنی است. صورت جمع اسمهای بی قاعده را باید در واژگان درج کرد.

**ب.** مفرد و جمع بودن: نظام پردازش زبان با این ویژگی می تواند ضمیر متناسب با اسم را به کار ببرد.

**ج.** شمردنی یا اسم جنس بودن (mass noun) اسمهای عام: با این ویژگی می توان نوع حرف تعریف و گاهی نوع صفتی را که با اسم به کار می رود تعیین کرد.

**د.** ویژگی عام یا خاص یا اسم جمع بودن: با این اطلاعات می توان حرف تعریف و فعل مناسب با اسم را انتخاب کرد.

**ه .** جنس دستوری (gender): با جنس دستوری می توان ضمیری را که به اسم ارجاع داده شده معین کرد.

**و.** اسم معنی (abstract) یا اسم ذات (concrete) بودن: این دو گروه صفتهای متفاوت می گیرند. مثلاً صفت "سبز" فقط با اسم ذات می آید (پیراهن سبز).

**ز.** انسان و جاندار و بیجان بودن: با این سه مشخصه می توان نوع فعل و ضمیر و غیره را که با اسم به کار می روند مشخص کرد. مثلاً فعل "خوردن" فقط فاعل انسان و جاندار و مفعول خوردنی می گیرند.

اطلاعات بند "ه" و "و" و "ز" بسیار مهم است زیرا در نظامهای متن سازی نیاز به محدودیتهای باهم آیی یا رجحان را برآورده می کند. محدودیتهای با هم آیی در فهم متن نیز مهم است. زیرا به تعیین معنی متناسب با بافت کمک می کند. جدول شماره ۹ ساختار جدول اسم و چند مثال مربوط را نشان می دهد.

| کلمه | شمارهٔ هم نویسه | شمارهٔ معنی | مفرد جمع | باقاعده بیقاعده | شمردنی اسم جنس | اسم عام اسم جمع | جنس دستوری | اسم معنی ذات | جاندار بیجان |
|---|---|---|---|---|---|---|---|---|---|
| Hen | ۱ | a۱ | مفرد | باقاعده | شمردنی | اسم عام | مونث | ذات | جاندار |
| Pasta | ۱ | ۱ | مفرد | باقاعده | اسم جنس | اسم عام | بی جنس | ذات | بیجان |
| Person | ۱ | ۱ | مفرد | بی قاعده | شمردنی | اسم عام | بی جنس | ذات | انسان |
| Scissors | ۰ | ۱ | جمع | باقاعده | شمردنی | اسم عام | بی جنس | ذات | بیجان |

جدول شمارهٔ ۹- جدول اسم

برای اسمهای بی قاعده جدول جداگانه آمده است که صورتهای جمع را نشان می دهد (جدول شماره ۱۰) دلیل وجود جدول اسمهای بی قاعده در بانک واژگان این است که نظامهای پردازش زبان نمی توانند صورتهای جمع اسمهای بی قاعده را با قاعده کلی پیش بینی کنند.

| صورت جمع | شماره معنی #S | شماره هم‌نویسه #H | کلمه |
|---|---|---|---|
| Geese | 1 | 1 | Goose |
| Men | 1 | 1 | Man |
| Mice | 1 | 1 | Mouse |
| Oxen | 1 | 0 | Ox |
| Sheep | 1 | 1 | Sheep |
| People | 1 | 1 | person |

جدول شماره ۱۰- جدول اسمهای بی قاعده

**۵-۳- جدول فعل**

فعلها یکی از پیچیده ترین بخشهای اجزاء کلام است زیرا به بسیاری از ویژگیهای آنها باید توجه کرد. باید دانست که آیا فعل باقاعده است یا بی قاعده و پویا (dynamic) است یا ایستا (Stative) و لازم است یا متعدی و اینکه آیا فعل متمم جمله ای می گیرد یا خیر و آیا می توان آن را به صورت مجهول درآورد یا نه. اگر فعل متمم جمله ای می گیرد یا خیر و آیا می توان آن را به صورت مجهول درآورد یا نه. اگر فعل کنشی (performative) است از چه طبقه ای است (طبقه به زعم مک کاولی ۱۹۷۹).

متاسفانه با تحلیل تعاریفی که فرهنگهای ماشین خوان به دست داده اند نمی توان فعلها را به صورت خودکار طبقه بندی کرد. از این رو در حال حاضر این اطلاعات را با دست عمل می آورند. جدول شماره ۱۱ مربوط به ساختمان فعلها است.

| کنشی | مجهول | متمم جمله ای | متعدی لازم | پویا ایستا | باقاعده بی‌قاعده | شماره معنی S# | شماره هم‌نویسه H# | کلمه |
|---|---|---|---|---|---|---|---|---|
| امری دارد | دارد | دارد | متعدی | پویا | بیقاعده | ۱ | ۱ | Forbit |
| خیر | دارد | دارد | متعدی | ایستا | باقاعده | ۱ | ۰ | realize |

جدول شماره ۱۱- جدول فعلها

فعلهای بی قاعده نیز جدا از فعلهای باقاعده و در جدول شماره ۱۲ آمده است تا نظامهای پردازش زبان بتوانند صورتهای گوناگون فعلها را تشخیص دهند.

| اسم مفعول | گذشته | شماره معنی #S | شماره هم نویسه #H | کلمه |
|---|---|---|---|---|
| Swum | Swam | ۱ | ۰ | swin |
| Run | Ran | ۱ | ۱ | Run |
| Put | Put | ۱ | ۱ | Put |

جدول شماره ۱۲- جدول فعلهای بی قاعده

در جدول شماره ۱۳ نمونه فعلهایی آمده است که متمم جمله ای می گویند. نوع حرف ربط ( to, for, that و غیره) نوع متمم (مصدری و التزامی و غیره) و طبقات استلزامی (implication classes) نظیر فعلهای واقعی (fative) و استلزامی (implicative) و فقط اگر (only-if) و شرطی منفی و استلزامی منفی و ضدواقعی (counter-factive) نیز مشخص شده است (Joshi & Weisechedel., 1976). از روی جدول همچنین می توان دریافت که آیا فعلها در فهرست ایندیانا هست یا خیر.

| فهرست ایندیانا | طبقه استلزامی | نوع متمم | حرف اضافه | شماره معنی | شماره | کلمه |
|---|---|---|---|---|---|---|
| وجود دارد | استلزامی | مصدری | To | ۱ | ۰ | agree |

جدول شماره ۱۳- جدول متمم های جمله ای

در موارد وابسته ها (arguments) فعل باید نقش نحوی (فاعل, مفعول غیرمستقیم, مفعول مستقیم, مفعول حرف اضافه) و حالت (case) آن مشخص شود. حالت شامل کنشگر (agent) و کنش پذیز (patient) و بهره ور (beneficiary) و غیره است. همچنین نوع محدودیتهای باهم آیی مانند جانداری و بیجانی و انسان بودن که جای خالی وابسته ها را پر می کند باید مشخص شود و این نکته نیز تصریح شود که آیا وابسته ها اجباری است یا اختیاری و حذف شدنی. بسیاری از زبانشناسانی که با کامپیوتر تحقیق می کنند بر این عقیده اند که چهارچوب حالت (case frame) – که شامل اطلاعات مربوط به حالت است – یکی از مهمترین سازه های بانک واژگان برای پردازش زبان است (Fillmore, 1968; Hirst, 1986). در جدول شماره ۱۴ اینگونه اطلاعات درج شده است.

| محدودیتهای باهم آیی | اجباری و اختیاری بودن | حالت | نقش نحوی | شماره معنی S# | شماره هم‌نویسه H# | کلمه |
|---|---|---|---|---|---|---|
| انسان یا سازمان | اجباری | کنشگر | فاعل | 1 | 1 | Sell |
| شیء | اجباری | تم | مفعول مستقیم | 1 | 1 | Sell |
| انسان یا سازمان | اختیاری | کنش پذیر | To+گروه اسمی | 1 | 1 | Sell |
| پول | اختیاری | وسیله | For+ گروه اسمی | 1 | 1 | Sell |

جدول شماره ۱۴- جدول چهارچوب حالت

ردیف اول افقی این جدول نشان می دهد که فاعل فعل Sell یک کنشگر است و آمدن آن اجباری است و از نظر محدودیتهای باهم آیی باید انسان یا سازمان باشد. ردیف دوم افقی مشخص می کند که مفعول مستقیم فعل که در نقش تم (theme) قرار می گیرد هر چیزی می‌تواند باشد. از ردیف سوم افقی در می یابیم که قبل از خریدار معمولاً حرف اضافه to می آید و خریدار کنش پذیر یا بهره ور عمل است و از نظر محدودیتهای باهم آیی انسان یا سازمان است و می توان آن را حذف کرده ردیف چهارم افقی معین می کند که وسیله اختیاری معمولاً از نوع پول است و با حرف اضافه for می آید. باید توجه کرد که

از روی ردیف آخر عمودی می توان دریافت که اطلاعاتی صریح و روشن درباره طبقه بندی انواع محدودیتهای باهم آیی موجود است (ر. ک بخش ۶).

کلاً می توان گفت جانشینهای نقش حالت باید به گونه ای طبقه بندی شود که از نظر مفهوم با آنچه در ردیف محدودیتهای باهم آیی می آید مربوط باشد.

برای اطلاعات دستوری بیشتر از داده ها و فهرست واژه هائی که تعدادی از زبانشناسان جمع آوری کرده بودند استفاده شده است. آنچه متعاقباً می آید جدولهائی است که برای سازمان اطلاعات دستوری از آنها استفاده شده است. این اطلاعات دستوری را گروه زبانشناسان دانشگاه ایندیانا و دانشگاه برندیس جمع آوری کرده بودند.

فهرست فعلهای ایندیانا در دو جدول ذخیره شده است. جدول شماره ۱۵ تعریفهائی است که برای اطلاعات جدول شماره ۱۶ داده شده است.

| فهرست | الگوی کاربرد فعل | مثال |
|---|---|---|
| ۳ | that indicative +passive+It Sentence | … It was said that |

جدول شماره ۱۵- فهرست ایندیانا

| شمارهٔ فهرست | شماره معنی #S | شمارهٔ همنویسه #H | کلمه |
|---|---|---|---|
| ۳۵ | ۱ | ۰ | Sesm |
| ۴/۵ | ۱ | ۱ | Request |

جدول شماره ۱۶- فعلهای ایندیانا

جدول فهرست ایندیانا (جدول شماره ۱۵) شامل شماره فهرستها و الگوی کاربرد فعلها و یک مثال برای هر یک از ۷۲ فهرست است. جدول شماره ۱۶ مشخص می کند که فعل متعلق به کدام فهرست است. این جدول شامل تمام فعلهائی است که در فهرست فعلهای ایدیانا وجود دارد و شماره فهرست هر فعل در آن ذکر شده است. استفاده کننده می تواند با اطلاعات جدول شماره ۱۵ و ۱۶ جمله های درست بسازد.

اطلاعات مربوط به فعل که از فهرست فعلهای برندیس گرفته شده درجدول شماره ۱۷ و ۱۸ آمده است. یکی از جدولها برای فعلها و اطلاعات جزء به جزء زیرمقوله ای شان و دیگری برای تعبیر رمزها (codes) است (ر. ک. جدول شمار ۶ برای توضیح رمزها).

| فعل | رمز وابسته ها |
|---|---|
| ADMIRE | NP-۴P –OC DO –ING – ۴P –DOPOSSING THAT DO |

جدول شماره ۱۷- فهرست برندیس

| معنی رمز | رمز |
|---|---|
| مفعول مستقیم | DO |
| حرف اضافهٔ for | P4 |

جدول شماره ۱۸- معنی رمزها

برای متن سازی اطلاعات اضافی زیادی درباره فعل لازم است. اتکینز (Atkins et al., 1988) و لوین (۱۹۸۷) بسیاری از این دست اطلاعات را ذکر کرده اند. مثلاً برخی ویژگیهای مهم فعلها از این قرار است که آیا فعل سببی (causative) است یا آغازی (inchoative) و آیا مربوطه به خوردن است (خوردن, نوشیدن) یا تغییر موقعیت را نشان می دهد (حرکت کردن, غلتیدن, چرخیدن).

**۵-۴- جدول صفت**

چون صفت اسم را توصیف می کند باید صفتها را به گونه ای طبقه بندی کرد که نظامهای پردازش زبان صفت مناسب هر اسم را تشخیص دهند. جدول شماره ۱۹ جدول اصلی صفتها است.

| کلمه | شماره-همنویسه#H | شماره معنی #S | پویا-ایستا | مدرج –غیرمدرج | ذاتی-غیر ذاتی | مجموعه معنایی | جایگاه وقوع |
|---|---|---|---|---|---|---|---|
| Brave | ۱ | ۱ | پویا | مدرج | ذاتی | ذهنی | معمولی |
| `mechanical | 1 | 1a | ایستا | غیرمدرج | ذاتی | کارکردی | پیش توصیفگر |
| asleep | 1 | 1 | ایستا | مدرج | غیرذاتی | مشتق از اسم | غیراسنادی |

جدول شماره ۱۹- طبقه صفتها

در این جدول پویا و ایستا بودن و مدرج و غیرمدرج بودن ذاتی اسم بودن (inherent) و غیرذاتی بودن (non-inherent) صفتها مشخص شده است. همچنین مشخص شده است که صفتها متعلق به چه مجموعه معنائی هستند. به این معنی که آیا صفتها تاکیدی است یا میان گروهی (post-determiner) آیا به صورت ذهنی و عینی سنجش پذیر است و آیا مربوط به سن و رنگ است و یا مشتق از اسم (denominal) است و جنس و شباهت را تعیین می کند. مربوط به منشاء یا سبک و روش بودن نیز از دیگر ویژگیهائی معنائی صفتها است.

برخی از اطلاعات مربوط به مجموعه معنائی صفت در تعیین جایگاه صفت و اسمی که صفت در جمله به آن ملحق می شود تاثیر اساسی دارد. گرچه بیشتر صفتها هم در جایگاه اسنادی (attributive) و هم در جایگاه گزاره (predicative) قرار می گیرد برخی صفتها وجود دارد که غیراسنادی است. مثلاً می‌توان گرفت This class is asleep اما نمی توان ترکیب the asleep class را ساخت. برخی صفتها نیز غیرگزاره ای است. یعنی می توان گفت electrical engineer ولی نمی توان گفت the engineer is electrical جدول شماره ۲۰ شامل صفتهایی است که تصریف (inflection) آنها پیش بینی شدنی نیست.

| کلمه | شماره همنویسه#H | شماره معنی#S | صفت تفضیلی | صفت برتر |
|---|---|---|---|---|
| Good | 1 | 1 | Better | Dest |
| far | 1 | 1 | Farther | Farthest |

جدول شماره ۲۰- جدول تعریف صفتها

### ۵-۵- جدول قید

قید پیچیده ترین بخش اجزاء کلام است. قید می تواند اسم و فعل و صفت و قید و کل جمله را توصیف کند. طبقه بندی قیدها مهم است زیرا قیدهای هر طبقه کلمه های خاصی را توصیف می کنند و در جایگاه های متفاوتی در جمله می نشینند. کونلن (conlon et al., 1990) شرح مفصلی از قیدهای موجود در واژگان داده است. در این بخش فقط جدول طبقه بندی قیدها (جدول شماره ۲۱) و جدولی که شامل اطلاعات مربوط به جایگاه قیدها است (جدول شماره ۲۲) ارائه می شود.

جدول شماره ۲۱ قید زمان و روش و جهت و سنجش و ربط و علت و شمار و مکان و امتداد را نشان می دهد.

| طبقه<br>طبقه | شماره<br>هم معنی #S | شماره<br>هم نویسه #H | کلمه<br>کلمه |
|---|---|---|---|
| زمان | ۰ | ۰ | Often |
| روش | ۱ | ۱ | Fast |
| درجه | ۱ | ۱ | far |

جدول شماره ۲۱- جدول طبقه قیدها

جدول شماره ۲۲ چهار ردیف عمودی دارد سه ردیف اول مربوط به کلمه و شماره هم نویسه و شماره معنی و چهارمی فهرست جایگاهها است.

| جایگاه | شماره<br>معنی #S | شماره<br>هم نویسه #H | کلمه |
|---|---|---|---|
| جایگاه آغازی۱، آغازی ۲، میانی۱، میانی۳، میانی۶، پایانی ۳، پایانی ۵ | ۱ | ۰ | Often |
| جایگاه میانی۳، پایانی۱، پایانی۴، پایانی۵ | ۱ | ۱ | Fast |
| جایگاه پایانی ۱، پایانی۵ | ۱ | ۱ | far |

جدول شماره ۲۲- جدول جایگاه قیدها

### ۶- روابط واژگانی معنائی

چنانکه ذکر شد نظامهای پردازش زبان و هوش مصنوعی (artificial intelligence) به همان اطلاعات پیچیده ای نیاز دارند که انسان درباره روابط بین کلمه ها می داند و با آن می تواند زبان را به شیوه موثری پردازش کند. نظامهای مطلوب پردازش زبان باید اطلاعات را به شکل نمادهای معنائی مانند منطق گزاره ها (predicate calaulus) و چهارچوب (frames) و شبکه معنائی (semantic networkd) در بانک دانش (knowledge base) خود ذخیره کنند برای انجام این کار باید دانش واژگانی و چگونگی ارتباط آنها با یکدیگر را به صورتی صریح و واضح ارائه کرد. مثلاً باید در بانک واژگان این نکته درج شود که "پیشی" گربه و گربه پستاندار و پستاندار حیوان است. نظامهای پردازش زبان از روی این مثال باید بفهمند که "پیشی" حیوان است. مثلاً فرهنگ دانشجوی وبستر یکی از معنی های canary را چنین برآورده است:

"پرنده ای کوچک و کاملاً زرد" با خلاصه کردن تعریف مذکور به صورت خودکار این رابطه به دست می آید: "قناری پرنده کوچک" و "قناری زرد رنگ". اهلسود معنایی حدود ۲۰۰ واژه را با روش نیمه خودکار از فرهنگ دانشجوی وبستر و فرهنگ انگلیسی کالینز به دست آورده است (Ahlswede, 1985; 1988; Ahlswede. et al., 1986; Markowitz et al., 1986). او همچنین توانسته با استفاده از فرهنگهای مذکور رابطه های بیش از صد هزار جفت واژه را استخراج کند. این رابطه ها ظاهراً یک گونه مهم از روابط پایه ای معنی در زبان است (R. N. Smith, 1985). اطلاعات جامعتر درباره رابطه ای معنائی بین واژه ها در کتاب ایونز (Evens, 1988) وجود دارد.

یک شبکه روابط واژگانی و معنائی مدخلها را در بانک واژگان به یکدیگر مربوط می کند. رایجترین روابط معنایی در بانک مورد نظر ما آنهائی است که در فرهنگهای موجود در بازار معمول است. نظیر هم معنائی (synonymy) و تضاد (antonymy) و طبقه بندی (taxonomy). اما از روابط معنائی دیگر نیز در این بانک استفاده شده است. مانند عضو بودن (Member) که رابطه گوسفند به گله از این نوع است و رابطه جزء به کل (part-whole) مثل رابطه چرخ با اتوبوس و رابط شیء به جنس (made-of) که در رابطه با چاقو با فولاد وجود دارد. روابط واژگانی زیادی نیز هست. نظیر رابطه صفت با اسم که صفت و اسم با فعل که اسم حالت یا عمل را به فعل مربوط به آن متصل می کند (مانند رابطه "مرگ" با "مردن" و "پذیرش" با "پذیرفتن"). بانک واژگان مورد نظر ما دارای جدول روابط و واژه ها و عبارتهائی است که با این روابط به هم وصل شده است. روابط واژگانی و معنائی در جدول شماره ۲۳ آمده است.

| کلمه | قرینه | انعکاسی | متعدی | یک به یک |
|---|---|---|---|---|
| Taxonomy | خیر | بله | بله | خیر |
| Synonymy | بله | بله | بله | خیر |
| child | خیر | خیر | خیر | خیر |

جدول شماره ۲۳- جدول روابط بازرگانی و واژگانی معنائی

در این جدول نام بیش از ۲۰۰ رابطه اسم با اسم اشاره شده است. همچنین مشخص شده که آیا هر رابطه متقارن (symmetric) است یا انعکاسی (reflexive) و متعدی (transitive) و یک به یک (one-to-one). مثلاً هم معنائی یک رابطه انعکاسی است و طبقه بندی یک رابطه متعدی است که متقارن نیست.

محققان هوش مصنوعی روابط طبقه بندی را معمولاً به صورت "یک ... است" نشان می‌دهند "مثلاً "قرقاول مرغ است" و "مرغ پرنده است"). هم معنائی رابطه مهم دیگری است که نظامهای پردازش زبان می توانند با آن معانی واژه ها را به هم وصل کنند. فرهنگ دانشجوی وبستر و فرهنگ انگلیسی کالینز اطلاعات مربوط به هم معنائی را به صورت صریح و تلویحی فهرست کرده اند. در فرهنگ وبستر حدود ۴۶۰۰۰ جفت واژه هم معناست. از آنجائیکه طبقه بندی و هم معنائی بسیار رایج است و اغلب به شکلهای خاصی به کار می رود جدولهای جداگانه‌ای برای آنها ساخته شده است. جدول شماره ۲۴ جدول رابطه هم معنائی است.

| شماره معنی S#۲ | شماره هم‌نویسه H# ۲ | کلمه ۲ | مقوله | شماره معنی S#۱ | شماره هم‌نویسه H# | کلمه ۱ |
|---|---|---|---|---|---|---|
| ط ۱ | ۱ | Sudden | صفت | 2a | ۰ | Abrupt |
| ۱ | ۱ | Graph | اسم | 2b | 1 | chart |

جدول شماره ۲۴- جدول هم معنائی

برای هر کدام از اجزاء اصلی کلام یک جدول روابط وجود دارد که اولین کلمه در فهرست سه گانی (triple) متعلق به آن جزء اصلی کلام است.

(معنی ۲, هم نویسه ۲, کلمه ۲, رابطه, معنی۱, هم نویسه ۱, کلمه ۱) فهرست سه گانی اسمها.

(معنی ۲, هم نویسه ۲, کلمه ۲, معنی ۱, هم نویسه ۱, کلمه ۱) فهرست سه گانی فعلها

(معنی ۲, هم نویسه ۲, کلمه ۲, رابطه, معنی ۱, هم نویسه ۱, کمله ۱) فهرست سه گانی صفتها

(معنی ۲, هم نویسه ۲,کلمه ۲, رابطه, معنی ۱, هم نویسه ۱,کلمه ۱) فهرست سه گانی قیدها

مثالی از جدول سه گانی اسمها:

| | | | | | | |
|---|---|---|---|---|---|---|
| ۱ | ۱ | مردن | عمل | ۱ | ۰ | مرگ |
| ۱ | ۰ | گوسفند | بچه | ۱ | ۱ | بره |

اطلاعات مربوط به روابط واژگانی و معنائی به کار بازیابی اطلاعات و دیگر حوزه های پردازش زبان می آید.

## ۷- قابلیتهای بانک واژگان

از مدخلهای اصلی فرهنگ انگلیسی کالینز حدود ۶۱۰۰۰ واژه و عبارت جمع آوری شده است. برنامه هایی برای شناسائی واژه های مشتق و عبارتهای درونه گیری شده (embeded) در مدخل کلمات دیگر نوشته شده است. در آغاز سعی شده که برای اینگونه واژه ها و عبارتها مدخل ساخته شود. از معنی بیشتر اینها تعریفی داده نشده است. احساس ویراستاران فرهنگ انگلیسی کالینز احتمالاً این بوده که خواننده خود می تواند به معنای مورد نظر برسد. برنامه های در دست تهیه است که با آنها اطلاعاتی صریح – که اغلب در مورد رابطه واژه ها و عبارتهای مذکور به هسته آنها است _ از معنی این گونه عناصر در بانک درج شود.

بسیار از قیدهای مختوم به LY که در مدخل صفتها و قیدها و اسمها وجود دارد و خود از همین مقوله ها مشتق شده است از جمله عناصر مذکور است و مورد توجه خاصی قرار گرفته است (conlon et al., 1990). صورتهای گذشته ساده و اسم مفعول فعلهای بی قاعده نیز در فرهنگ کالینز شناسائی و مدخلهای مختصری برای آنها ساخته شده است. فهرست جمع های بی قاعده هم با هسته مربوط به آنها در بانک آمده است. افزودن واژه های مشتق و عبارتهای درونه گیری شده و فعلها و ترکیبات دارای ادات و صورتهای گذشته ساده و اسم مفعول بی قاعده که همگی از فرهنگ کالینز گردآوری شده است باعث شده

که شماره واژه ها و عبارتهای مستقل به ۱۵۹۱۹۳ برسد. تعداد معانی مستقل کمتر و تنها ۱۲۱۶۸۴ است زیرا اگر چه مدخلهای اصلی دارای یک یا چند معنی و همراه با تعاریف جداگانه است برای صورتهای بی قاعده و کلمات مشتق و عبارتهای درونه‌گیری شده معنی مستقل داده نشده است.

افزودن عناصر از منابعی نظیر فهرست ایندیانا و فهرست فعلهای برتدیس و فرهنگ قیدهای یاکوبسون گرچه مدخلهای موجود را غنی تر کرده تعداد واژه ها و عبارتها را بیشتر نکرده است. بزرگترین مشکل این است که قسمت انطباق واژه های دیگر فهرستها با معانی متناسب در فرهنگ انگلیسی کالینز است. سعی بر این بوده که این کار به صورت خودکار انجام شود ولی اغلب دخالت انسان در آن ضروری است. برای هر منبعی باید سیاست جداگانه ای انتخاب کرد و همین کار را پیچیده می کند.

برای بهبود قابلیتهای بانک به جای استفاده از تعداد اندکی جدول بزرگ از تعداد زیادی جدول تقریباً کوچک استفاده شده است. به منظور سریع تر کردن بازیابی اطلاعات و پیوسته کردن عملیات یک تشخیص دهنده (identifoer) واحد برای هر زنجیره سه گانی هم نویسه و معنی و شمار نصب شده است. استفاده کننده از وجود این تشخیص دهنده بی اطلاع است اما تشخیص‌دهنده به طور قابل توجهی زمان بازیابی اطلاعات را کم می کند.

## ۸- روش استفاده از بانک واژگان

استفاده کنندگان برای بازیابی اطلاعات و ساختن تزاروس و سازه یابی و متن سازی به اطلاعاتی درباره روابط واژگانی و معنائی نیاز دارند. تجارب ساختن تزاروس در مقاله هائی دیگر ذکر خواهد شد. امید می رود که از بانک واژگان در سازه یابی پرسشهای استفاده کنندگان نیز استفاده شود اما این کار هنوز در مراحل ابتدائی است.

مجموعه ای فرعی (subset) از اصطلاحات پزشکی این بانک را در موسسه تکنولوژی دانشگاه ایلی نویز برای بخش فهم متن یک نظام آموزشی هوشمند استفاده می کنند (Sev et al., 1991). مجموعه فرعی تقریباً مشابهی را هم برای متن سازی به کار می برند (Lee & Evens., 1992) .

اغلب متقاضیان تاکنون خواهان مجموعه های فرعی این بانک بوده اند و گمان می رود که چنین سیری _ احتمالاً به جز مواردی که بازیابی کلی اطلاعات در کار باشد _ ادامه یابد. به همین دلیل فکر تهیه ابزارهائی برای ساختن یک واژگان کارآمد برای زبان فرعی مطرح شده است. یکی از این ابزارها برنامه صرف (morphplogy) مربوط به جمع های با قاعده و صیغه های فعل است که به مدخل مجموعه های فرعی اضافه می شود.

در مدخلهای فرهنگ انگلیسی کالینز سه نوع اطلاعات مربوط به زبان فرعی هست و امید می رود که در نهایت بتوان از هر سه نوع استفاده کرد. حوزه تخصصی برخی مدخلها صریحاً ذکر شده است. مثلاً جلو مدخل hematocele لفظ اختیاری pathol و جلو مدخل hematic حرف اختصاری Med آمده است. برای تهیه مجموعه فرعی کامل پزشکی باید مشخص شود که آسی‌شناسی (pathology) جزئی از پزشکی (Medicine) است. یعنی باید الفاظ اختصاری نشان‌دهنده حوزه های تخصصی را جمع کرد و آنها را به صورت سلسله مراتب در آورد. مدخلهای دیگری وجود دارد که از ترکیب اطلاعات زبان فرعی با اطلاعات مربوط هم معنائی حاصل شده است. بنابراین مدخل اسید استیل سالی سیلیک به صورت "نام شیمیایی آسپیرین" آمده است. در این موارد باید دو نوع اطلاعات مذکور را تمیز داد و هر دو را به شکلی صریح و روشن در حافظه یادداشت کرد. در دیگر موارد اطلاعات زبان فرعی به صورت تعریف می آید که

برای شناسائی این تعریفها باید آنها را به دقت طبقه بندی کرد. این عمل بیشتر در خصوص واژه های عمومی و معمول صادق است. شاید به این دلیل که فرض کرده اند اغلب استفاده کنندگان فرهنگها – هر چند ممکن است تلفظ و معنی دقیق آنها را ندانند – با واژه های مذکور آشنا هستند. بنابراین گرچه hematocele و hematin حوزه تخصصی مشخص دارد واژه hemorhange ندارد و تعریف آن را باید به صورت "خونریزی فراوان از زخم یا جراحت" آورد تا از واژه هم آوایش که یک اصطلاح پزشکی است متمایز شود.

موریس و هرست (Morris & Hirst., 1991) مثالهایی برای اثبات اهمیت روابط معنائی در فهم متن داده اند. یکی از مثالها چنین است:

Mary ate a peach, She liked fruit.

در این مثال سازه باب بادی رابطه بین peach (هلو) و fruit (میوه) را پیدا کند تا بتواند نقش جمله دوم را بفهمد. فهم چنین رابطه ای مثلاً برای پاسخ به سئوال "چرا مری هلو خورد" لازم است. در مقاله نویسندگان مذکور تحلیل مفصلی از زنجیره واژه های موجود در یک نوشته جی تیتل (Jay Teitel) وجود دارد. نویسندگان توضیح داده اند که ساختار واژه ها چگونه ساختار منظوری (intentional structure) متن را آشکار می کند. برای اثبات این نکته پاراگرافی از یکی از آثار تیتل را ذکر و سپس آنرا تحلیل می کنیم:

In front of me lay a virgin crescent cut of pine brush. A dozen houses were going up, in various stages of contruction, Surrounded by hummocks of dry earth and stands of precariously tall trees nude halfwaye up their trunks.

کشف رابطه های واژگانی بین pine و brush و trees و trunks برای سازه یابی فهم پاراگراف ضروری است. سازه یاب به بانک واژگان مراجعه و درخواست یافتن رابطه tree با trunks را می کند. بانک واژگانی به بررسی جدول اصلی و سپس فهرست سه گانی اسمهای می پردازد (ر. ک. بخش ۶) و به رابطه جزء به کل tree با trunk اشاره می کند.

اهمیت رابطه های واژگانی در متن سازی حتی بیشتر از اهمیت آن در فهم متن است. فرض کنیم که برنامه پزشکی (lee, W., & Evens, M., 1992) بخواهد توضیحی در مورد خونریزی مغزی بدهد. این برنامه می تواند با چنین توضیحی آغاز کند: "کومالین و کی سی آی برای خانم تامپسون تجویز شده است." آنچه برنامه در مرحله بعد به آن نیاز دارد مقدار دارو است پس باید یک عبارت برای کومادین بسازد. برنامه بانک به جستجوی می پردازد تا اصطلاحاتی را که هم طبقه کومادین پیدا کند. کومادین در جدول طبقه بندی شده اسامی پیدا می شود (ضدانعقاد دارو) روی صفحه می آید. تولیدکننده متن در صورتی که داروی دیگری به تازگی برای این بیماری ذکر نشده باشد یکی از اصطلاحات فوق الذکر را انتخاب می کند.

**۹- نتیجه:**

این مقاله درباره ساختن یک بانک پرحجم واژگان است که به روش نیمه خودکار عمل می‌کند. در این مقاله درباره منابع اطلاعاتی که از آنها برای ساختن بانک استفاده شده و برخی ویژگیهای مهم و ساختار کلی بانکی بحث شده است. همچنین طرح دقیق و مدخلها و نمونه هر یک در این نوشته گنجانده شده است.

هدف از ایجاد بانک واژگان ساختن یک ترازوس برای نظامهای بازیابی اطلاعات بود. سپس برای کمک به متن سازی و سازه یابی و تحلیل نامه های تجاری اطلاعات واژگانی بیشتری به واژگان اضافه شد.

به یقین با پیچیده تر شدن نظامهای پردازش زبان به اطلاعاتی بیش از این نیاز خواهد بود. در آینده انواع دیگر اطلاعات نیز اضافه خواهد شد تا بانک واژگان مذکور جامعتر شود.

**توضیحات:**

۱- منظور از لفظ "زبان" در این مقاله زبان بشر است-م

۲- این مساله در یکی از نظریه های زبانشناسی به نام فرمانش و مرجع گزینی (government & Binding) و به صورت نظریه ای فرعی به نام کنترل (control) مطرح است. به این معنی که گاهی فاعل پنهانی (non-overt) را با یک مقوله تهی به نام PRO نشان می‌دهد. مثلاً در جمله "John promised Mary to go" مقوله تهی را John باید پر کند و چون John در اینجا فاعل جمله اصلی نیز هست بنابراین در این جمله کنترل فاعلی (Subject control) وجود دارد. در جمله "John Persuaded Mary to go" بر عکس کنترل مفعولی وجود دارد چون Mary کننده کار در جمله پیرو است ضمن اینکه مفعول جمله پایه نیز هست-م

**فهرست منابع**


Ahlswede, T. E. (1985). A tool kit for lexicon building. *proceedings of the 23 rd annual meetiny of the association for computational linyuistics*, Chicago, 268-276.

Ahlswede, T.E.(1988) .*Syntactic and semantic analysis definitions in a machine readable dictionary*. Unpublished doctoral dissertation. Department of Computer Science, Illionois lnstitute of Technoiogy, Chicago, IL.

Ahlswed, T.E., & Evens, M.(1988) Generating a relational lexicon from a machine-readable dictionary. *International journal of Lexicography*, 1(3).214-237.

Ahlswed, T,E., Evens, M., Markowitz, J., & Rossi, K. (1986).Buiding a lexical database by parsing Websters Seventh collegiate Dictionary. In G. Johannesen (Ed), Advances in lexicology*: Proceedings of the University of the University of Waterloo Centre for the New Oxford English Dictionary* (pp. 65-78). Waterloo Centre for the New OED.

Ahlswed, T. E. Anderson, J., Evens, M., Li, S. M., Neises, J., Pin-Ngern, S.,&., Markowitz, J. (1988). Automatic Construcrtion of a phrasal thesaurus for an information retrieval system. *Proceedings of RIAO88 (Recherche d'lnformation Assistee par Ordinateur), PP.597-608.*

Alexander, D.,& kunz, W.J.(1964). Some Classes of Verbs in English, Bloomington, In: Linguistics Research Project, indiana University.

Amsler, R.A.(1981). A taxonomy for English nouns and Verbs. *Proceedings of the 19th Annual Meeting Of the ACL, PP.133-138.*

Amsler, R.A. (1984) Machine- readable dictionaries. In M.E. Williams (Ed*), Annual review of information science and technology (PP.61-209) Silver Spring, MD: ASIS.*

Atkins, B.S., Kegl, J.,& Levin, B.(1988). Anatomy of the verb entry: From linguistic theory to lexicographic Practice. *International Journal of Lexicography, Vol.l.*

Austin, J. (1962) .*How to do things with words*. London: Oxford University Press.

Black, J. (1985). An exposition on understanding expository text. In B. Britton & J.Black (Eds*), Understanding expository text (PP.249-267).* Hillsdale, NJ: Erlbaum.

Boguraev, B.,& Levin, B.(1990). Models for lexical Knowledge bases. *Proceedings of the Sixth Annual Conference of the University of waterloo Centre for the New Oxford English Dictionary*: *Electronic Text Research (PP.65-78).* University of Waterloo, Canada.

Birdgeman, L., Dillinger, D., Higgens, C., Seaman, P.D., & Shank, F. (1965). *More Classes of vebs in English.* Blooming'ton, IN: lndiana University Linguistics Club.

Byrd, R.J. (1989). Discovering relationships among word senses. *Proceedings of the Fifth Annual Conference of the University of Waterloo Centre for the New Oxford English Dictionary (PP.67-79).*

Byrd, R.J., Calzolari, N., Chodorow, M., Klavans, Neff, M., & Risk, O. (1987). Tools and methods for computational lexicology. *Computational Linguistics, 13 (3/4), 219-240.*

Calzolari, N. (1984). Mechine-readable dictionaries, lexical databases and the lexical system. *Proceedings of the International Conference on Computational liguistics(P.460).*

Calazolari, N. (1988). the dictionary and the thesaurus can be combined. In M. Evens (Ed), Relational models of the lexicon (pp.75-96). combrige: cambridge University press.

Carver, D.J. (1974). *Collins English dictionary. London*: Collins Publishers.

CELEX (1986). *CELEX Newsletter*. University of Nijmegen. Available from: Centre for Lexical Information, Wundtlaan I, 6525 XD Nijmegen, the Netherlands.

Conlon, S. Pin-Ngern, M.,& Ahlswede, T. (1990). Generating a lexical database for adverbs. *Proceedings of the Sixth Annual Conference of the University of Waterloo Centre for the New Oxford English Dictionary:* Electronic Text Research (pp.95-109). Waterloo: Centre for the New OED, Canada.

Conlon, S. Pin-Ngern, & Evens, M. (in press). A lexical database for nouns to support parsing, text Generation, and information retrieval. *Research in humanities computing.* Oxford: Oxford University Press.

Cumming, S (1987). The lexicon in text generation. *Proceedings of the Lexicon Workshop of the 1987 Linguistics Summer lnstitute (PP.151-181).* Stanford, CA.

Evens, M. (Ed) (1988). Relational models of the lexicon. Cambridge: cambridge University press. Evens, M., pin-Ngern, s., Ahlswede, T., li, s. M., & Markowits, j.1989)6. Acquring information from informants for a lexical database. Proceedings first international Lexical Acquisilion Workshop.

Fillmore, C.J. (1968) The case for case. In E. Bach & R. Harms (Eds.) *Universals in linguistic theory(pp.0-88).* New York: Holt, Rinehart, & Winston.

Fox, E.A. (1980) Lexical relations: enhancing effectiveness of information retrieval systems. *ACM SIGIR Forum, 15 (3), 5-36.*

Fox, E. A., Nutter, J. T., Ahlswede, T. E., Evens, M., & Markowits, J. (1988). Building a large thesaurus for information retrieval. *Proceeding 2nd ACL Applied (pp. 101-108).*

Gove, P. B. (Ed.) (1983). *Webster's seventh new collegiate dictionary*. Sprinfield, MA: G. & C. Merriam Company.

Grimshaw, J., & Jackendoff, R. (1985). *Report to the NSF on grant IST-81-20403.* Waltham, MA: Department of Linguistics, Brandeis University.

Halliday, M.A.K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Hirst, G. (1986). Why dictionaries should list case structure. *Proceeding of the Conference on Advances in lexicology (pp. 147-163).* University of Waterloo, November.

Householder, F., Wolck, W., Matthews, P.H., Tone, J., & Wilson, J. (1965). *Preliminary classification of adverbs in English*. Bloomington, IN: Indiana Linguistics Club.

Jacobs, P. (1988). Phred: A generator for natural language interfaces. In D. McDonald & L. Bolc (Eds.), Natural language generation systems (pp. 312-354). Berlin: Springer-Verlag.

Jacobson, S. (1964). *Adverbial positions in English*. Disseration, Uppsala, Stockholm: AB Studentbok.

Jacobson,S.(1978). *On the use, and syntax of English preverbal adverbs.* Stockholm: Almqvist & Wilksell international.

Jensen, K., & Binot, J.L. (1987). Disambiguating Prepositional phrase attachments by using online dictionary definitions. *Computational Linguistics, 13 (3/4),251-260.*

Joshi, A.K., & Weischedel, R.M. (1976). Some frills for modal tic-tac-toe: Semantics of predicate complement constructions. *IEEE Transactions on Computers, C-25 (4), 374-389.*

Lee, W., & Evens, M. (1991). Generating coherent text using lexical semantic relations. *Proceeding of the Midwest Artificial Inteligence and Cognitive Science Society (pp. 41-45).*

Lee W., & Evens, M. (1992). Generating cohesive text using lexical functions. *Proceedings of the International Workshop on Meaning Text Theory*. Darmstadt july,1992. Berlin: Springer Verlag.

Lesk, M.E. (1969). Word-word associtions in document retrieval systems. *American Documentation, 20 (1), 27-38.*

Levin, B. (1987). Approaches to lexical semantic representation. *Proceedings of the Lexicon Workshop of thy 1987 Linguistics Summer Institute (PP.39-87).*

Maggs , P., & Sprowl, J. (1987). *Computer applications in the law*. St. Paul, MN: West Publishing Company.

Markowits, J ., Ahlswede, T., & Evens, M. (1986). Semantically Significant patterns in dictionary definitions. *Proceedings of the 24th Annual Meeting of the Associations for Computational Linguistics (PP. 112-119).*

Markowitz, J., Pin-Ngern, S., Evens, M., Anderson, J., & Li, S.M. (1988). Generating lexical database entries for phrases. In D. L. Berg (Ed), *Information in Text: Proceedings of the Univesity of waterloo Centre for thy New Oxford English Dictionary (PP . 115 – 127).*

Mc Cawley, J. (1979). Adverbs Vowels and Other Objeds of Wonder. Chicago: the University of chicago press .

Mc Cawley , J. (1986). What linguists might contribute to dictioary making if they could get their act together . In V. Raskin & P. Bjarkman (Eds), *The real – World linguist (pp. 3 .10)*. Norwood, NJ: Ablex.

McDonald, D. (1991).On the place of Words in the generatios process. In C. paris ., W. Swartout, & W. Mann (Eds), *Natural language generation in artifical intelligence and Computational linguistics (PP.227- 247*). Dordrecht: Kluwer.

Miller, G., Beckwith, R ., Fellbaum, C., Gross, D. & Miller, K. (1991). Word net. *International journal of Lexicography, 4(1), 1 – 75.*

Morris, J., & Hirst, G. (1991). Lexical cohesion, the thesaurus, and the structure of text. *Computational Linguistics, 17(2), 21 – 48.*

Myers, G. (1991). Lexical cohesion and specialized Knowledge in scientific and popular scieence texts. *Discourse Processes , 14(1), 1 – 26.*

Nuuter, J. T., Fox, E.A., & Evens, M. (1990). Building a lexicon from machine – readable dictionaries for improved information retrieval . *Literary and linguistic Computing , 5 (2), 129-138.*

Perrault , R ., & Grosz, (1986). Natural language interfacas. *Annual review of computer science.* Palo Alto,CA: Annual Reviews, Inc. (pp. 47-82).

Pin-Ngern, S. (1990). *A lexical database for English to support information ratrieval, parsing , and taxt generation* .Unpublished ph. D. dissertation , Computer Science Department , Illinois lnstitute of Technology .

Procter , P., & (Ed) . (1978). Longman dictionary of contemporary English . London : longman Group.

Raymond ,D ., & Tompa, F. (1987). Typertext and the New Oxford English Dictionary . *Communications of the Associatin for Ccomputing Machinery, 31 (7), 971-879.*

Seu, J., Evens , M., Michael , J., & Rovich , A. (1991). Understanding ill-formed inpud to an intelligent tutoring system in an LFG framework. *Proceedings of the Third Midwest Artificial Intellgence and cognitive Science Society Conference (pp. 36-40).*

Sinclair , J.M. (Ed) (1987a). The collins COBUILD English Language dictionary. London : William Collins & sons , Ltd.

Sinclair, J.M.(Ed). (1987b*). Looking up :An account of the COBUILD project in lexical computing.* London :William Collins & Sons, Ltd.

Smadja, F. (1991a). From N-grams to collоctions: An evaluation of Xtract. *Proceedings of the 29th Annual Meeting of the ACL (pp . 279-284.*

Smadja, F. (1991b). Macrocoding the lexicon with co-occurrence Knowledge. In U. Zernik (Ed). *Lexical acquisition: Exploiting on-line resources to build a lexicon (pp. 165-190).* Hillsdale, NJ: Lawrence Erlbaum Associates.

Small , S., Cotrell, G., & Tanenhaus , M. (1988*). Lexical ambiguity resolution* . Los Altos, CA: Morgan Kaufmann.

Smith , L. (1987). Artificial intelligence and information retrieval. In M. Williams (Ed), *Annual review of information science and technology (pp. 41 – 77).* Amsterdam: Elsevier.

Smith, R. N. (1985). Conceptual primitives in the English lexicon. *Papers in Linguistics, 18(1), 99- 137.*

Stubbs, J., & Tompa , F. (1984). Waterloo and the New Oxford English Dictionary. *20th Annual Coference on Editorial problems (pp. 19 – 44).* Toronto New York , NY: AMS Press

Wang , Y. C., Vandendorpe, J ., & Evens , M. (1985). Relational thesauri in information retrieval. *JASIS, 36(1), 15 – 27.*

Warnar, A. (1987). Natural language processing . In M. Williams (Ed) , *Annual review of information science and technology (pp . 79 – 108 ).* Amsterdam :Elsevier.