



جمهوری اسلامی ایران
Islamic Republic of Iran

سازمان ملی استاندارد ایران

Iranian National Standardization Organization



استاندارد ملی ایران

۱۴۸۴۹

چاپ اول

بهمن ۱۳۹۱

INSO

14849

1st. Edition

Jan.2013

مدیریت منابع زبانی – چارچوب نشانه‌گذاری
واژگانی (LMF)

**Language resource management – Lexical
markup framework (LMF)**

ICS:01.020

به نام خدا

آشنایی با سازمان ملی استاندارد ایران

مؤسسه استاندارد و تحقیقات صنعتی ایران به موجب بند یک ماده ۳ قانون اصلاح قوانین و مقررات مؤسسه استاندارد و تحقیقات صنعتی ایران، مصوب بهمن ماه ۱۳۷۱ تنها مرجع رسمی کشور است که وظیفه تعیین، تدوین و نشر استانداردهای ملی (رسمی) ایران را به عهده دارد.

نام موسسه استاندارد و تحقیقات صنعتی ایران به موجب یکصد و پنجاه و دومین جلسه شورای عالی اداری مورخ ۹۰/۶/۲۹ به سازمان ملی استاندارد ایران تغییر و طی نامه شماره ۲۰۶/۳۵۸۳۸ جهت اجرا ابلاغ شده است.

تدوین استاندارد در حوزه های مختلف در کمیسیون های فنی مرکب از کارشناسان مؤسسه* صاحب نظران مراکز و مؤسسات علمی، پژوهشی، تولیدی و اقتصادی آگاه و مرتبط انجام می شود و کوششی همگام با مصالح ملی و با توجه به شرایط تولیدی، فناوری و تجاری است که از مشارکت آگاهانه و منصفانه صاحبان حق و نفع، شامل تولیدکنندگان، مصرف کنندگان، صادرکنندگان و وارد کنندگان، مراکز علمی و تخصصی، نهادها، سازمان های دولتی و غیر دولتی حاصل می شود. پیش نویس استانداردهای ملی ایران برای نظرخواهی به مراجع ذی نفع و اعضای کمیسیون های فنی مربوط ارسال می شود و پس از دریافت نظرها و پیشنهادات در کمیته ملی مرتبط با آن رشته طرح و در صورت تصویب به عنوان استاندارد ملی (رسمی) ایران چاپ و منتشر می شود.

پیش نویس استانداردهایی که مؤسسات و سازمان های علاقه مند و ذیصلاح نیز با رعایت ضوابط تعیین شده تهیه می کنند در کمیته ملی طرح و بررسی و در صورت تصویب، به عنوان استاندارد ملی ایران چاپ و منتشر می شود. بدین ترتیب، استانداردهایی ملی تلقی می شود که بر اساس مفاد نوشته شده در استاندارد ملی ایران شماره ۵ تدوین و در کمیته ملی استاندارد مربوط که سازمان استاندارد تشکیل می دهد به تصویب رسیده باشد.

سازمان ملی استاندارد ایران از اعضای اصلی سازمان بین المللی استاندارد^۱ (ISO) کمیسیون بین المللی الکتروتکنیک^۲ (IEC) و سازمان بین المللی اندازه شناسی قانونی^۳ (OIML) است و به عنوان تنها رابط^۴ کمیسیون کدکس غذایی^۵ (CAC) در کشور فعالیت می کند. در تدوین استانداردهای ملی ایران ضمن توجه به شرایط کلی و نیازمندی های خاص کشور، از آخرین پیشرفت های علمی، فنی و صنعتی جهان و استانداردهای بین المللی بهره گیری می شود.

سازمان ملی استاندارد ایران می تواند با رعایت موازین پیش بینی شده در قانون، برای حمایت از مصرف کنندگان، حفظ سلامت و ایمنی فردی و عمومی، حصول اطمینان از کیفیت محصولات و ملاحظات زیست محیطی و اقتصادی، اجرای بعضی از استانداردهای ملی ایران را برای محصولات تولیدی داخل کشور و/ یا اقلام وارداتی، با تصویب شورای عالی استاندارد، اجباری نماید. سازمان می تواند به منظور حفظ بازارهای بین المللی برای محصولات کشور، اجرای استاندارد کالاهای صادراتی و درجه بندی آن را اجباری نماید. همچنین برای اطمینان بخشیدن به استفاده کنندگان از خدمات سازمان ها و مؤسسات فعال در زمینه مشاوره، آموزش، بازرسی، ممیزی و صدور گواهی سیستم های مدیریت کیفیت و مدیریت زیست محیطی، آزمایشگاه ها و مراکز کالیبراسیون (واسنجی) وسایل سنجش، مؤسسه استاندارد این گونه سازمان ها و مؤسسات را بر اساس ضوابط نظام تأیید صلاحیت ایران ارزیابی می کند و در صورت احراز شرایط لازم، گواهی نامه تأیید صلاحیت به آن ها اعطا و بر عملکرد آن ها نظارت می کند. ترویج دستگاه بین المللی یکاها، کالیبراسیون (واسنجی) وسایل سنجش، تعیین عیار فلزات گرانبها و انجام تحقیقات کاربردی برای ارتقای سطح استانداردهای ملی ایران از دیگر وظایف این سازمان است.

1-International organization for Standardization

2-International Electro technical Commission

3-International Organization for Legal Metrology (Organization International de Metrologie Legal)

4-Contact point

5-Codex Alimentarius Commission

کمیسیون فنی تدوین استاندارد
« مدیریت منابع زبانی – چارچوب نشانه‌گذاری واژگانی (LMF) »

رئیس:

فیضی درخشی، محمدرضا
(دکترای کامپیوتر)

سمت و / یا نمایندگی

دانشگاه تبریز

دبیر:

سالک‌زمانی، لیلا
(فوق لیسانس زبان و ادبیات فرانسه)

کارشناس

اعضاء: (اسامی به ترتیب حروف الفباء)

بدرزاده، فریبا
(دانشجوی کامپیوتر)

اداره کل استاندارد و تحقیقات صنعتی استان
آذربایجان شرقی

بشیری، عباس
(فوق لیسانس علوم اقتصاد)

معاونت برنامه‌ریزی استانداری آذربایجان
شرقی

جاودانی، بهاره
(فوق لیسانس الکترونیک)

شرکت معیارآزمای ارس

سالک‌زمانی، مریم
(فوق لیسانس علوم تغذیه)

اداره کل استاندارد و تحقیقات صنعتی استان
آذربایجان شرقی

محمدی، آیدا
(لیسانس مترجمی زبان و ادبیات انگلیسی)

شرکت آتابای پلاست

معین، فروزان
(فوق لیسانس روابط عمومی و امور بین‌الملل)

اداره استاندارد و تحقیقات صنعتی مراغه

نصیری زنوز، وحید
(لیسانس کامپیوتر)

شرکت داده تامین خاورمیانه

فهرست مندرجات

صفحه

عنوان

ب	آشنایی با سازمان ملی استاندارد ایران
ج	کمیسیون فنی تدوین استاندارد
ه	پیش‌گفتار
و	مقدمه
۱	۱ هدف و دامنه کاربرد
۱	۲ مراجع الزامی
۱	۳ اصطلاحات و تعاریف
۸	۴ استانداردهای کلیدی مورد استفاده در LMF
۹	۵ مدل LMF
۱۶	پیوست الف (الزامی) بسط ساختواژه
۱۹	پیوست ب (اطلاعاتی) مثال‌های ساختواژه
۲۶	پیوست پ (الزامی) بسط فرهنگ لغت قابل خواندن توسط ماشین
۲۸	پیوست ت (اطلاعاتی) مثال‌هایی برای فرهنگ لغت قابل خواندن توسط ماشین
۳۰	پیوست ث (الزامی) بسط نحوی NLP
۳۳	پیوست ج (اطلاعاتی) مثال‌های نحو NLP
۳۷	پیوست چ (الزامی) بسط معنایی NLP
۴۰	پیوست ح (اطلاعاتی) مثال‌های معنایی NLP
۴۸	پیوست خ (الزامی) بسط نشانه‌گذاری‌های چندزبانه NLP
۵۲	پیوست د (اطلاعاتی) مثال‌های نشانه‌گذاری‌های چندزبانه NLP
۵۶	پیوست ذ (الزامی) بسط الگوهای صرفی NLP
۶۱	پیوست ر (اطلاعاتی) مثال‌های الگوهای صرفی NLP
۷۸	پیوست ز (الزامی) بسط الگوهای اصطلاحات چندکلمه‌ای NLP
۸۰	پیوست ژ (اطلاعاتی) مثال‌های الگوهای اصطلاح چند کلمه‌ای NLP
۸۲	پیوست س (الزامی) بسط بیان محدودیت
۸۴	پیوست ش (اطلاعاتی) مثال بیان محدودیت
۸۶	پیوست ص (اطلاعاتی) ارتباط با چارچوب نشانه‌گذاری واژه‌شناسی (TMF) و سایر سیستم‌های بازنمایی مبتنی بر مفهوم
۸۷	پیوست ض (اطلاعاتی) تعریف نوع سند (DTD) برای LMF
۹۱	پیوست ط (اطلاعاتی) کتابنامه

پیش‌گفتار

استاندارد " مدیریت منابع زبانی - چارچوب نشانه‌گذاری واژگانی (LMF)" که پیش‌نویس آن در کمیسیون‌های فنی مربوط تهیه و تدوین شده و در نودمین اجلاس کمیته ملی استاندارد اسناد، تجهیزات اداری و آموزشی مورخ ۹۰/۱۱/۲۳ مورد تصویب قرار گرفته است، اینک به استناد بند یک ماده ۳ قانون اصلاح قوانین و مقررات سازمان ملی استاندارد ایران، مصوب بهمن ماه ۱۳۷۱، به عنوان استاندارد ملی ایران منتشر می‌شود.

برای حفظ همگامی و هماهنگی با تحولات و پیشرفت‌های ملی و جهانی در زمینه صنایع، علوم و خدمات، استانداردهای ملی ایران در مواقع لزوم تجدید نظر خواهد شد و هر پیشنهادی که برای اصلاح و تکمیل این استانداردها ارائه شود، هنگام تجدید نظر در کمیسیون فنی مربوط مورد توجه قرار خواهد گرفت. بنابراین، باید همواره از آخرین تجدیدنظر استانداردهای ملی استفاده کرد.

منبع و مأخذی که برای تهیه این استاندارد مورد استفاده قرار گرفته به شرح زیر است:

ISO 24613: 2008, Language resource management – Lexical markup framework(LMF)

بهینه‌سازی تولید، حفاظت و ترویج منابع الکترونیکی واژگانی^۱، از جنبه‌های بسیار مهمی است که معمولاً بر فنآوری‌های زبان انسانی (HLT)^۲، خصوصاً پردازش زبان طبیعی^۳ (NLP) و فنآوری‌های ترجمه انسانی تأثیر می‌گذارد. یکی دیگر از جنبه‌های پراهمیت تأثیرگذار با بهینه‌سازی فرایند حاکم بر تلفیق آن‌ها در کاربردها مرتبط است. چارچوب نشانه‌گذاری واژگانی^۴ (LMF)، یک فرامدل انتزاعی است که چارچوب استاندارد شده و عمومی‌ای را برای ساخت واژگان کامپیوتری ایجاد می‌کند. LMF، اطلاعات زبان‌شناسی را به روشی کدگذاری می‌کند که بتوانند در کاربردها و منظوره‌های متفاوتی مورد استفاده مجدد قرار بگیرند. LMF، در واقع بازنمایی^۵ مشترک و متداول جنبه‌های واژگانی از جمله جنبه‌های ساختارزی^۶، نحوی^۷ و معنایی^۸ را ممکن می‌سازد.

LMF، به دنبال ایجاد مدلی برای ساخت و استفاده از منابع الکترونیکی واژگانی در ابعاد کوچک تا بزرگ می‌باشد تا به منظور کنترل تبادل داده‌ها، بین و در میان این منابع و همچنین تسهیل در ادغام تعداد زیادی از منابع مختلف الکترونیکی مجزا در همدیگر و تشکیل منابع الکترونیکی گسترده جامع مورد استفاده قرار گیرد. هدف اصلی LMF، ایجاد یک ساختار مدولار است که هم‌کنش‌پذیری^۹ محتوایی واقعی را در میان تمامی جنبه‌های منابع الکترونیکی واژگانی تسهیل بخشد.

هسته مرکزی LMF، سلسله مراتب اساسی اطلاعات یک مدخل^{۱۰} واژگانی از جمله اطلاعات مربوط به یک حالت را شرح می‌دهد. منابع مختلفی که جزئی از مفهوم LMF می‌باشند، به هسته مرکزی ضمیمه شده است. این منابع عبارتند از:

الف - رده‌های^{۱۱} داده‌ای خاصی که توسط انواع مختلفی از منابع مرتبط با LMF مورد استفاده قرار می‌گیرند، از جمله رده‌های داده‌هایی که به خود فرامدل مربوط هستند و هم چنین رده‌هایی که مربوط به پیوسته‌های هسته مرکزی می‌باشند.

ب - محدودیت‌های حاکم بر روابط میان رده‌های داده‌ای مذکور با فرامدل و پیوسته‌های آن.

پ - روش‌های استاندارد جهت بیان این رده‌ها و در نتیجه به منظور قراردادن آن‌ها در چارچوب ساختاری LMF و مرتبط ساختن آن‌ها با مدل‌های بسط^{۱۲} مربوط.

1-Lexical

2-Human language technologies

3-Natural language processing

4-Lexical Markup Framework

به منظور جلوگیری از اطاله کلام از سرواژه "LMF" به جای "چارچوب نشانه‌گذاری واژگانی" در سراسر متن استاندارد استفاده شده است.

5-Representation

6-Morphological

7-Syntactic

8-Semantic

9-Interoperability

10-Entry

11-Categories

12-Extention

ت - واژگانی که توسط LMF برای بیان اهداف اطلاعاتی مرتبط جهت شرح چگونگی بسط و گسترش روش LMF از طریق اتصال به منابع مختلف (پیوست‌ها) و روش‌های خاص برای تحلیل و طراحی چنین سامانه‌های به هم متصلی مورد استفاده قرار می‌گیرند.

بسط‌های هسته مرکزی که در پیوست‌های این استاندارد مدون شده‌اند، عبارتند از: فرهنگ‌های لغات قابل خواندن توسط ماشین و منابع واژگانی پردازش زبان طبیعی.

بسط‌های LMF در چارچوبی بیان شده‌اند که استفاده مجدد مولفه‌های^۱ مرکزی آن (همانند ساختارها، رده‌های داده‌ها و واژگان) را در تلفیق با مولفه‌های اضافی که برای یک منبع خاص الزامی می‌باشند، شرح می‌دهند.

انواع نمونه‌های مجزای معرفی شده از LMF می‌توانند شامل منابع الکترونیکی واژگانی به عنوان پایگاه داده‌های واژگانی نسبتاً ساده، فرهنگ لغات پردازش زبان طبیعی و ترجمه ماشینی و هم چنین پایگاه داده‌های الکترونیکی واژگانی چند زبانه، دو زبانه و تک‌زبانه باشند. LMF، سیستم و ساختار جامعی را برای تحلیل و طراحی منابع الکترونیکی واژگانی جدیدی فراهم می‌سازد، اما ساختارها و محدودیت‌های داده‌ها و واژگان مورد استفاده در طراحی منابع الکترونیکی واژگانی خاص را مشخص نمی‌کند. LMF هم چنین سازوکارهایی را برای تجزیه و تحلیل و نیز توصیف منابع موجود با بهره‌گیری از چارچوب توصیفی مشترک فراهم می‌آورد. به منظور دستیابی به هر دو هدف طراحی منابع واژگانی جدید و توصیف منابع واژگانی موجود، LMF به تبیین شرایطی می‌پردازد که به داده‌های موجود در یک منبع واژگانی امکان انطباق با ساختار آن را فراهم کرده، در نتیجه یک قالب میانی^۲ برای تبادل داده‌های واژگانی ایجاد می‌کند.

1-Components
2-Intermediate format

مدیریت منابع زبانی - چارچوب نشانه‌گذاری واژگانی (LMF)

۱ هدف و دامنه کاربرد

هدف از تدوین این استاندارد، توصیف چارچوب نشانه‌گذاری واژگانی، به عنوان فرامدلی برای بازنمایی داده‌های موجود در پایگاه داده‌های واژگانی مورد استفاده در کاربردهای کامپیوتری چند زبانه و تک زبانه می‌باشد.

LMF سازوکارهایی را در راستای امکان گسترش و تلفیق انواع مختلف منابع الکترونیکی واژگانی فراهم می‌سازد. این سازوکارها تا حد امکان به نمایش فرهنگ‌های لغت موجود خواهند پرداخت و در صورت عدم امکان، اطلاعات مشکل‌ساز شناسایی و تفکیک خواهند شد.

۲ مراجع الزامی

مدارک الزامی زیر حاوی مقرراتی است که در متن این استاندارد ملی ایران به آن‌ها ارجاع داده شده است. بدین ترتیب آن مقررات جزئی از این استاندارد ملی ایران محسوب می‌شود. در صورتی که به مدرکی با ذکر تاریخ انتشار ارجاع داده شده باشد، اصلاحیه‌ها و تجدید نظرهای بعدی آن مورد نظراین استاندارد ملی ایران نیست. در مورد مدارکی که بدون ذکر تاریخ انتشار به آن‌ها ارجاع داده شده است، همواره آخرین تجدید نظر و اصلاحیه‌های بعدی آن‌ها مورد نظر است. استفاده از مراجع زیر برای کاربرد استاندارد الزامی است:

- 2-1 ISO 639 (all parts), Codes for the representation of names of languages
- 2-2 ISO 1087-1, Terminology work — Vocabulary — Part 1: Theory and application
- 2-3 ISO 1087-2, Terminology work — Vocabulary — Part 2: Computer applications
- 2-4 ISO 12620, Terminology and other content and language resources — Data categories — Specification of data categories and management of a Data Category Registry for language resources 3)
- 2-5 ISO 15924, Information and documentation — Code for the representation of names of scripts

۳ اصطلاحات و تعاریف

در این استاندارد، علاوه بر اصطلاحات و تعاریف تعیین شده در استانداردهای ISO 1087-1 و ISO 1087-2، اصطلاحات و تعاریف زیر نیز به کار می‌رود:

۱-۳

حالت اختصاری^۱

حالتی (بند ۳-۱۴) که از حذف هر یک از قسمت‌های حالت کامل (بند ۳-۱۶) همان واژه قاموسی (بند ۳-۲۵) ایجاد می‌شود.

1-Abbreviated form

۲-۳

افزوده^۱

کلمه‌ای که همراه فعل می‌آید و برخلاف آرگمان نحوی^۲ (بند ۳-۴۳) وجود آن در جمله غیرضروری می‌باشد. یادآوری - قیده‌ها، افزوده‌هایی هستند که ممکن است به یک جمله اضافه شوند.

۳-۳

وند^۳

وند عبارت است از واژ وابسته^۴ (بند ۳-۸) که می‌تواند بخشی از یک حالت (بند ۳-۱۴) بوده و در فرآیند تصریفی^۵ (بند ۳-۲۰)، پیوندی^۶ (بند ۳-۵)، اشتقاقی^۷ (بند ۳-۱۲) و یا ترکیب^۸ (بند ۳-۹) نقش داشته باشد. یادآوری - وندها به شکل‌های پیش‌وندها (در ابتدای کلمه)، پس‌وندها (در انتهای کلمه)، میان‌وندها (در میان کلمه) و وندهای ترکیبی^۹ (ترکیبی از پیش‌وند و پس‌وند) کاربرد دارند.

۴-۳

وندافزایی^{۱۰}

فرآیندی که در آن یک وند (بند ۳-۳) به یک بن‌واژه^{۱۱} (بند ۳-۲۴) و یا یک ستاک^{۱۲} (بند ۳-۴۰) اضافه می‌شود.

۵-۳

پیوند

فرآیندی که در آن یک حالت پیوندی (بند ۳-۶) ساخته می‌شود.

۶-۳

حالت پیوندی

حالت کلمه (بند ۳-۴۷) که یک واژه قاموسی (بند ۳-۲۵) هنگام کاربرد در یک جمله و یا یک عبارت در یک زبان پیوندی (بند ۳-۷)، به خود می‌گیرد.

۷-۳

زبان پیوندی

زبانی که در آن یک واژه قاموسی (بند ۳-۲۵) در اشکال مختلف و گوناگون یک حالت کلمه (بند ۳-۴۷) نمایش داده شود که ممکن است شامل بیش از یک واژ (بند ۳-۳۱) باشد، در حالی که مرز بین واژها کاملاً معین است.

-
- 1-Adjunct
 - 2-Syntactic arguments
 - 3-Affix
 - 4-Bound morph
 - 5-Inflection
 - 6-Agglutination
 - 7-Derivation
 - 8-Composition
 - 9-Circumfixes
 - 10 -Affixation
 - 11-Lemma
 - 12-Stem

مثال - زبان‌های کره‌ای، ژاپنی، مجارستانی و ترکی زبان‌های پیوندی می‌باشند.

۸-۳

واژ وابسته

واژی که تنها با یک یا چندین واژ (بند ۳-۳۱) دیگر ظاهر می‌شود.

۹-۳

ترکیب کردن^۱

ترکیب کردن عبارت است از شکل‌گیری از یک واژه قاموسی (بند ۳-۲۵) که طی آن واژه قاموسی جدید [به همراه اطلاعات اجزای کلام^۲ (بند ۳-۳۷)] با اتصال دست کم دو واژه قاموسی در حالات (بند ۳-۱۴) اصلی آن‌ها یا با ایجاد تغییر شکل جزئی، ایجاد می‌شود.

یادآوری - ترکیب نبایستی با پیوند و اشتقاق که در آن واژه‌های وابسته به واژه‌های مستقل^۳ اضافه می‌شود، اشتباه گرفته شود.

۱۰-۳

ترکیب

واژه قاموسی (بند ۳-۲۵) همراه با اطلاعات اجزای کلام (بند ۳-۳۷) که از دو یا چند واژه قاموسی ساخته می‌شود.

۱۱-۳

حالت ترکیبی

حالتی (بند ۳-۱۴) که ماحصل فرآیند ترکیب کردن (بند ۳-۹) می‌باشد.

۱۲-۳

اشتقاق

به تغییری که در حالات (بند ۳-۱۴) یک واژه قاموسی (بند ۳-۲۵) برای ایجاد واژه قاموسی جدید ایجاد می‌شود، اطلاق می‌گردد که معمولاً با تغییر در ستاک کلمه (بند ۳-۴۰) یا اعمال فرآیند وندافزایی صورت می‌پذیرد.

۱۳-۳

حالت اشتقاقی

حالتی (بند ۳-۱۴) که ماحصل فرآیند اشتقاق است.

۱۴-۳

حالت

توالی^۴ واژها، حالت نامیده می‌شود.

۱۵-۳

واژ مستقل

1-Composition or Compounding

2-Part of speech

3-Free

4-Sequence

واژی که به تنهایی دارای معنی و مفهوم است.

۱۶-۳

حالت کامل^۱

بازنمایی^۲ کامل یک واژه قاموسی (بند ۳-۲۵) که حالت اختصاری (بند ۳-۶) برای آن وجود داشته باشد.

۱۷-۳

مشخصه دستوری^۳

ویژگی مربوط به حالت تصریفی (بند ۳-۱۹)، پیوندی (بند ۳-۶)، ترکیبی (بند ۳-۱۱) یا حالت اشتقاقی (بند ۳-۱۳) که خصیصه‌های دستوری یک حالت را توصیف می‌کند.

۱۸-۳

نگاره^۴

کوچکترین واحد نوشتاری که شامل حرف‌ها، تصویرنگاشت‌ها^۵، اندیشه‌نگاشت‌ها^۶، اعداد و نشانه‌های نقطه‌گذاری می‌باشد.

۱۹-۳

حالت تصریفی

حالت کلمه (بند ۳-۴۷) که یک واژه قاموسی (بند ۳-۲۵) هنگام استفاده در یک جمله و یا یک عبارت در یک زبان تصریفی (بند ۳-۲۱)، به خود می‌گیرد.

۲۰-۳

تصریف

فرآیندی که در آن یک حالت تصریفی (بند ۳-۱۹) ساخته می‌شود.

۲۱-۳

زبان تصریفی

زبان صرفی

زبانی که در آن یک واژه قاموسی (بند ۳-۲۵) در اشکال مختلف و گوناگون یک کلمه (بند ۳-۴۷) نمایش داده می‌شود، در حالی که هیچ مرز مشخصی بین واژه‌ها (بند ۳-۳۱) وجود ندارد، و واژه‌ها معمولاً به نحوی با هم آمیخته شده‌اند که حالت (بند ۳-۱۴) غیر قابل تفکیکی ایجاد شده است. مثال - زبان‌های اسپانیایی، ایتالیایی، فرانسوی و انگلیسی جزو زبان‌های تصریفی می‌باشند.

۲۲-۳

زبان میانجی^۷

-
- 1-Full form
 - 2-Representation
 - 3 -Grammatical feature
 - 4 -Graph
 - 5-Pictograms
 - 6-Ideograms
 - 7-Interlingua

یک زبان واسط انتزاعی است که در ترجمه ماشینی زبان‌های انسانی مورد استفاده قرار می‌گیرد.

۲۳-۳

زبان گسسته^۱

زبانی که در آن اکثریت بالایی از واژه‌ها (بند ۳-۳۱) از نوع واژه‌های مستقل (بند ۳-۱۵) هستند.

مثال - زبان چینی از نوع زبان گسسته می‌باشد.

۲۴-۳

بن‌واژه

حالت مرسوم^۲

حالت متعارفی (بند ۳-۱۴) است که برای اشاره به یک واژه قاموسی (بند ۳-۲۵) انتخاب می‌شود.

۲۵-۳

واژه قاموسی

یک واحد انتزاعی که عموماً با دسته‌ای از حالات (بند ۳-۱۴) که دارای معانی مشترکی هستند، همراه است.

۲۶-۳

مدخل واژگانی

ظرفی برای مدیریت یک یا چند حالت (بند ۳-۱۴) مختلف و احتمالاً یک یا چند معنای مختلف که برای

تعریف یک واژه قاموسی (بند ۳-۲۵)، مورد استفاده قرار می‌گیرد.

۲۷-۳

منبع واژگانی

پایگاه داده‌های واژگانی

پایگاه داده‌ای که شامل یک یا چند فرهنگ لغت (بند ۳-۲۸) می‌باشد.

۲۸-۳

فرهنگ لغت

منبعی که متشکل از چندین مدخل واژگانی (بند ۳-۲۶) برای زبان مفروض، می‌باشد.

یادآوری - فرهنگ لغت زبان خاص و/یا فرهنگ لغت آماده شده برای کاربردهای پردازش زبان طبیعی خاصی، می‌تواند شامل

زیردسته^۳ مخصوص یک زبان باشد.

۲۹-۳

فرهنگ لغت قابل خواندن توسط ماشین

MRD^۴

منبع واژگانی الکترونیکی (بند ۳-۲۷) به منظور انتقال دانش از انسان به ماشین طراحی شده است.

۳۰-۳

1 - Isolating language

2 - Canonical

3 - Subset

4 - Machine readable dictionary

واژه نامه ترجمه ماشینی

منبع واژگان الکترونیکی (بند ۳-۲۷) که در آن مدخل‌های واژگانی (بند ۳-۲۶) مجزا شامل مترادف‌هایی در دو یا چند زبان همراه با اطلاعات ساختواژی، نحوی و/یا معنایی جهت تسهیل پردازش خودکار و یا نیمه خودکار واژه‌های قاموسی (بند ۳-۲۵) درحین ترجمه ماشینی، هستند.

۳۱-۳

واژ

توالی نگاره‌ها (بند ۳-۱۸) و یا آواها (بند ۳-۳۸) را واژ می‌نامند.

۳۲-۳

الگوی ساختواژی

مجموعه‌ای از پیوندها و/یا عملیاتی^۱ که حالات مختلف یک واژه قاموسی (بند ۳-۲۵) را با توجه به زبان مربوط با یکی از فرآیندهای تصریفی (بند ۳-۲۰)، پیوندی (بند ۳-۵)، ترکیب (بند ۳-۹) و یا اشتقاقی (بند ۳-۱۲) می‌سازند.

۳۳-۳

ساختواژه

توصیفی از ساختار و تشکیل حالات (بند ۳-۱۴) می‌باشد.

۳۴-۳

اصطلاح چندکلمه‌ای

^۲MWE

واژه قاموسی‌ای (بند ۳-۲۵) که از توالی دو یا چند واژه قاموسی دیگر ساخته می‌شود، به نحوی که از ویژگی‌هایی برخوردار است که امکان پیش‌بینی آن‌ها بر اساس ویژگی‌های تک تک واژگان قاموسی (بند ۳-۲۵) یا حالت طبیعی ترکیب آن‌ها وجود ندارد.

یادآوری - اصطلاح چند کلمه‌ای می‌تواند یک ترکیب و یا بخشی از یک جمله و یا فقط یک جمله باشد. گروهی از واژگان قاموسی‌ای که یک اصطلاح چند کلمه‌ای را می‌سازند، می‌توانند پیوسته و یا ناپیوسته باشند. نایستی همیشه اصطلاح چند کلمه‌ای به عنوان اجزای کلام در نظر گرفته شود.

۳۵-۳

پردازش زبان طبیعی

NLP

رشته‌ای که فن و دانش پردازش داده‌های زبان‌شناسی را به وسیله یک کامپیوتر در بر می‌گیرد.

۳۶-۳

املاء (نوشتار)^۳

طریقه هجی کردن یا نوشتن واژگان قاموسی (بند ۳-۲۵) به شیوه متعارف را املا و یا رسم‌الخط می‌گویند.

1-Operations
2- Multiword expression
3-Orthography

۳۷-۳

اجزای کلام

رده واژگانی

کلاس کلمه

رده‌بندی اختصاص یافته به واژه قاموسی (بند ۳-۲۵) که بر اساس مشخصه‌های دستوری آن انجام می‌شوند. یادآوری - اجزای کلام معمول در زبان‌های اروپایی شامل: اسم، فعل، صفت، قید، حرف‌افزافه و غیره است.

۳۸-۳

آوا

به کوچک‌ترین واحد سیستم صوتی یک زبان، آوا اطلاق می‌گردد.

۳۹-۳

خط^۱

منظور از خط، دسته‌ای از حروف نگاره‌ای مورد استفاده برای حالت (بند ۳-۱۴) نوشتاری یک یا چند زبان مختلف می‌باشد.

۴۰-۳

ستاک

یک توالی از واژها (بند ۳-۳۱) که کوچک‌تر و یا برابر با حالت (بند ۳-۱۴) یک واژه قاموسی (بند ۳-۲۵) مستقل است و ممکن است از فرآیندهای تصریفی (بند ۳-۲۰)، پیوندی (بند ۳-۵)، ترکیب (بند ۳-۹) و یا اشتقاقی (بند ۳-۱۲) تاثیر بپذیرند.

۴۱-۳

چارچوب رده‌بندی فرعی^۲

ظرفیت^۳

دسته‌ای از محدودیت‌ها روی یک واژه قاموسی (بند ۳-۲۵) شامل ویژگی‌های آرگمان نحوی (بند ۳-۴۳) که می‌توانند یا باید با این واژه قاموسی مفروض رخ دهند.

۴۲-۳

فعل کمکی

فعلی که معنای عام بخصوصی به بستر^۴ می‌بخشد و فعلی که برای ایجاد یک واژه قاموسی (بند ۳-۲۵) با یک اسم ترکیب می‌شود.

۴۳-۳

آرگمان نحوی

یکی از عناصر عملکردی و ضروری یک عبارت که اجزای دیگر عبارت را توسط یک فعل شناسایی می‌کند.

1 -Script

2-Subcategorization frame

3-Valence (valency)

4-Context

۴۴-۳

آوانویسی^۱

حالتی (بند ۳-۱۴) ناشی از یک روش منسجم نگارش صداهای گفتاری است که شامل امکان تبدیل صداهای گفتاری توصیف شده در یک سیستم نوشتاری، به نمایش هم ارز آنهاست که توسط یک سیستم نوشتاری دیگر توصیف می‌شود.

۴۵-۳

نویسه‌گردانی^۲

حالتی (بند ۳-۱۴) ناشی از تبدیل یک سیستم نوشتاری به سیستم نوشتاری دیگر که معمولاً از طریق انطباق یک به یک میان حروف انجام می‌شود.

۴۶-۳

گونه

به یکی از حالات (بند ۳-۱۴) جایگزین یک واژه قاموسی (بند ۳-۲۵) اطلاق می‌گردد.

۴۷-۳

حالت کلمه

حالتی که یک واژه قاموسی (بند ۳-۲۵) هنگام استفاده در یک جمله یا یک عبارت، به خود می‌گیرد.

۴ استانداردهای کلیدی مورد استفاده در LMF

۱-۴ یونیکد^۳

LMF سازگار با یونیکد بوده و فرض می‌کند که همه داده‌ها با استفاده از نشانه‌گذاری نویسه‌ای یونیکد نمایش داده می‌شوند.

۲-۴ کدگذاری زبان

شناسه‌های^۴ زبانی مورد استفاده در منابعی که از LMF پیروی می‌کنند، باید با معیارهای مشخص شده در مجموعه استاندارد ISO 639 مطابقت داشته باشند. برخی موضوعات مربوط به ترکیب یک زبان و کدهای کشور مربوط، هم چنین هماهنگی در بخش‌های مختلف ISO 639، در استانداردهای خارجی^۵ حمایت شده با فنآوری‌های مربوط، مورد بررسی قرار گرفته‌اند. بهتر است کاربران به منظور حل مسائل مربوط به انتخاب و تطابق شناسه‌ها برای استفاده در محیط‌های کامپیوتری به نسخه اخیر "واحد وظایف مهندسی اینترنت، بهترین شیوه‌های متداول ۴۷، برچسب‌هایی برای شناسایی زبان‌ها"^۶ رجوع کنند.

1-Transcription

2-Transliteration

3-Unicode

4-Identifiers

5-External

6 -IETF Best Common Practices (BCP) 47, Tags for the Identification of Languages

۳-۴ کدگذاری خط

زمانی که کد خط بخشی از شناسه زبانی نباشد، شناسه‌های خط باید با معیارهای مندرج در استاندارد ISO 15924 مطابقت داشته باشد.

۴-۴ ثبت رده‌های داده‌ای (DCR)^۱

طراحان فرهنگ واژگان همسان با LMF باید از رده‌های داده‌های استاندارد ISO 12620 که در سایت اینترنتی www.isocat.org قرار گرفته است، استفاده کنند.

۴-۵ زبان مدل‌سازی یکپارچه (UML)^۲

LMF سازگار با ویژگی‌ها و قواعد مدل‌سازی UML است که توسط گروه مدیریت شیء (OMG)^۳ ارائه شده است. LMF از یک زیر مجموعه از UML که مربوط به توصیفات زبانی است، استفاده می‌کند.

۵ مدل LMF

۱-۵ مقدمه

مدل‌های LMF توسط کلاس‌های UML ارائه می‌شوند و هماهنگی‌ها میان کلاس‌ها و مجموعه‌ای از رده‌های داده‌ای استاندارد ISO 12620 تحت عنوان جفت‌های صفت - مقدار^۴ UML عمل می‌کنند. رده‌های داده‌ای برای ترسیم و آرایش نمودارهای UML که یک دید وسیع از مدل را نمایان می‌سازند، مورد استفاده قرار می‌گیرند. ویژگی‌های LMF در حالت شرح‌های متنی که معانی عناصر مدل‌سازی را شرح می‌دهند، می‌توانند بهتر از نمودارهای UML، اطلاعات کامل‌تری در باره کلاس‌ها، روابط و بسط‌ها ارائه کنند. در این فرآیند، توسعه‌دهندگان فرهنگ لغات باید از کلاس‌هایی استفاده کنند که در هسته مرکزی LMF (بند ۵-۲) مشخص شده‌اند. علاوه بر این، توسعه‌دهندگان به دلخواه می‌توانند از کلاس‌هایی که در بسط‌های LMF شرح داده شده‌اند، استفاده کنند (به پیوست‌های مربوط مراجعه شود). این توسعه‌دهندگان باید اقدام به تعریف انتخاب رده داده‌ای DCS همان گونه که برای روش‌های اجرایی انتخاب رده‌های داده‌ای LMF مشخص شده است، بنمایند.

۲-۵ بسط مرکزی LMF

هسته مرکزی LMF فرامدلی است که زیربنای انعطاف‌پذیری جهت ساخت مدل‌ها و بسط‌های LMF و پیوست‌ها فراهم می‌سازد (به شکل شماره یک مراجعه کنید).

۱-۲-۵ کلاس منبع واژگانی

منبع واژگانی کلاسی است که کل منبع را ارائه می‌کند. منبع واژگانی فقط و فقط یک بار ایجاد می‌شود. نمونه منبع واژگانی ظرفی است برای یک یا چند فرهنگ لغت.

1-Data Category Registry
2-Unified Modeling Language
3-Object management group
4-Attribute – value pairs

۵-۲-۲ کلاس اطلاعات کلی

اطلاعات کلی، کلاسی است که به ارائه اطلاعات اجرایی و صفات کلی دیگری می‌پردازد. یک رابطه مجتمع بین کلاس منبع واژگانی و کلاس اطلاعات کلی وجود دارد که اطلاعات اجرایی و صفات کلی منبع کامل را ترسیم می‌کند. در کلاس اطلاعات، زیر کلاس‌ها قابل تعریف نیستند.

یک نمونه از اطلاعات کلی باید دست کم صفات زیر را داشته باشد:

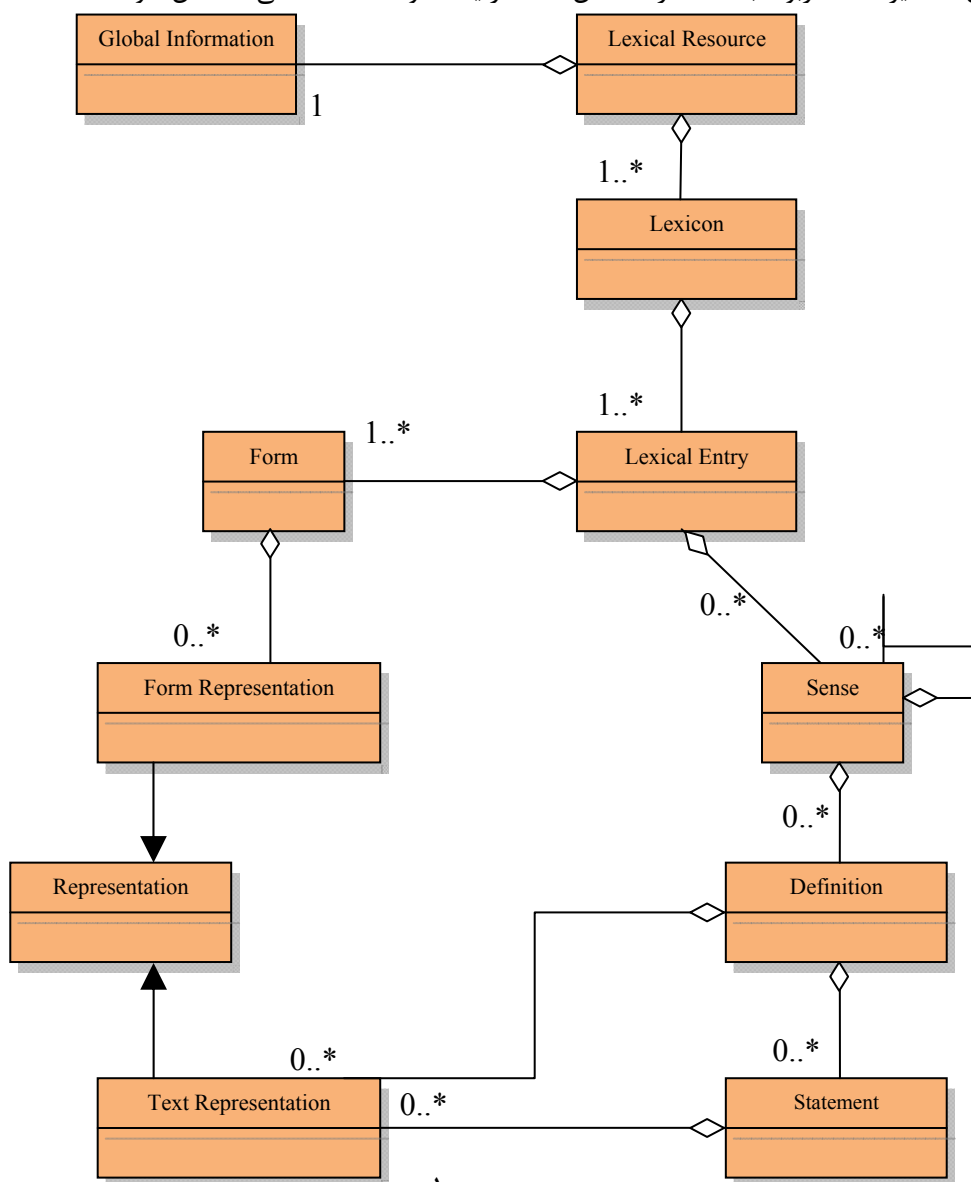
الف- /کدگذاری زبان/ این صفت مشخص می‌کند که چه استانداردی به منظور کدگذاری نام‌های زبانی داخل نمونه کلی منبع واژگانی مورد استفاده قرار گیرد.

نمونه‌ای از اطلاعات کلی می‌تواند دارای ویژگی‌های زیر باشد:

ب- /کدگذاری خط/ این ویژگی مشخص می‌کند که از کدام استاندارد به منظور کدگذاری نام‌های دست‌خط داخل نمونه کلی منبع واژگانی استفاده شده است.

پ- /کدگذاری حروف/ این ویژگی مشخص می‌کند که از چه نسخه یونیکد در داخل نمونه کلی منبع واژگانی استفاده شده است.

یادآوری - سایر نکات مربوط به استاندارد، ممکن است در یک نمونه اطلاعات کلی مشخص شود.



شکل ۱- هسته مرکزی LMF

۵-۲-۳ کلاس فرهنگ لغت

کلاس فرهنگ لغت، کلاسی است که دارای تمامی مداخل واژگانی یک زبان مفروض در منبع کلی می‌باشد. یک نمونه فرهنگ لغت باید دارای دست کم یک مدخل واژگانی باشد. در کلاس فرهنگ لغت امکان تعریف زیر کلاس وجود ندارد.

۵-۲-۴ کلاس مدخل واژگانی

کلاس مدخل واژگانی، کلاسی است که یک واژه قاموسی را در یک زبان مفروض بازنمایی می‌کند. مدخل واژگانی بخشی است که کلاس‌های حالت و مفهوم را مدیریت می‌کند. بنا بر این، مدخل واژگانی روابط بین حالات و مفاهیم مرتبط به آن‌ها را کنترل می‌کند. یک نمونه از مدخل واژگانی دارای یک تا چند حالات مختلف می‌باشد و ممکن است دارای یک تا چندین مفهوم مختلف باشد. در کلاس مدخل واژگانی امکان تعریف زیر کلاس‌ها وجود ندارد.

۵-۲-۵ کلاس حالت

کلاس حالت، یک حالت انتزاعی است که یک واژه قاموسی، گونه‌های مختلف ساختواژه یک واژه قاموسی یا یک واژه را در بر می‌گیرد. کلاس حالت یک یا چند حالت مختلف املائی از حالت انتزاعی، و نیز رده‌های داده‌ای که صفات حالت کلمه را شرح می‌دهد، مدیریت می‌کند (مثال: بن‌واژه، تلفظ و هجابندی). در کلاس حالت امکان تعریف زیر کلاس فراهم است.

۵-۲-۶ کلاس بازنمایی حالت

کلاس بازنمایی حالت، کلاسی است که یک حالت املائی صحیح از یک حالت را ارائه می‌دهد. زمانی که بیشتر از یک حالت املائی صحیح برای یک حالت وجود داشته باشد، کلاس نمایش حالت دارای یک رشته یونیکد که نمایانگر یک حالت است و هم چنین، در صورت نیاز، دارای جفت‌های صفت - مقدار از روی ویژگی‌های منحصر به فرد می‌باشد که زبان، خط و املائی خاصی را توصیف می‌کنند.

۵-۲-۷ کلاس بازنمایی

کلاس بازنمایی، یک کلاس انتزاعی است که به نمایش یک رشته یونیکد می‌پردازد و هم چنین، در صورت نیاز، جفت‌های صفت - مقدار منحصر به فرد را که به تشریح یک زبان، خط و املائی خاصی می‌پردازند، نمایش می‌دهد. در کلاس بازنمایی امکان تعریف زیر کلاس فراهم است.

۵-۲-۸ کلاس مفهوم

کلاس مفهوم، کلاسی است که یک معنی از مدخل واژگانی را ارائه می‌دهد. در این کلاس امکان تعریف زیر کلاس فراهم است. کلاس مفهوم دارای مفاهیم سلسله مراتبی است که در آن یک مفهوم ممکن است تخصصی‌تر از مفاهیم دیگر همان مدخل واژگانی باشد.

۵-۲-۹ کلاس تعریف

کلاس تعریف، کلاسی است که شرح روایتی از یک مفهوم را ارائه می‌دهد. کلاس تعریف برای کاربران انسانی به نمایش درمی‌آید تا تفهیم معنای یک مدخل واژگانی را برای کاربر تسهیل بخشد، در حالی که برای برنامه‌های کامپیوتری قابل پردازش نیست. یک نمونه‌ای از مفهوم می‌تواند دارای صفر یا چند تعریف باشد.

هر نمونه‌ای از تعریف می‌تواند با صفر یا چند نمونه از بازنمایی متن همراه باشد تا تعریف متن را در بیشتر از یک زبان یا خط مدیریت کند. شرح روایتی می‌تواند بهتر از یک نمونه مدخل واژگانی توسط زبان و یا خط متفاوتی بیان گردد.

۵-۲-۱۰ کلاس گزاره

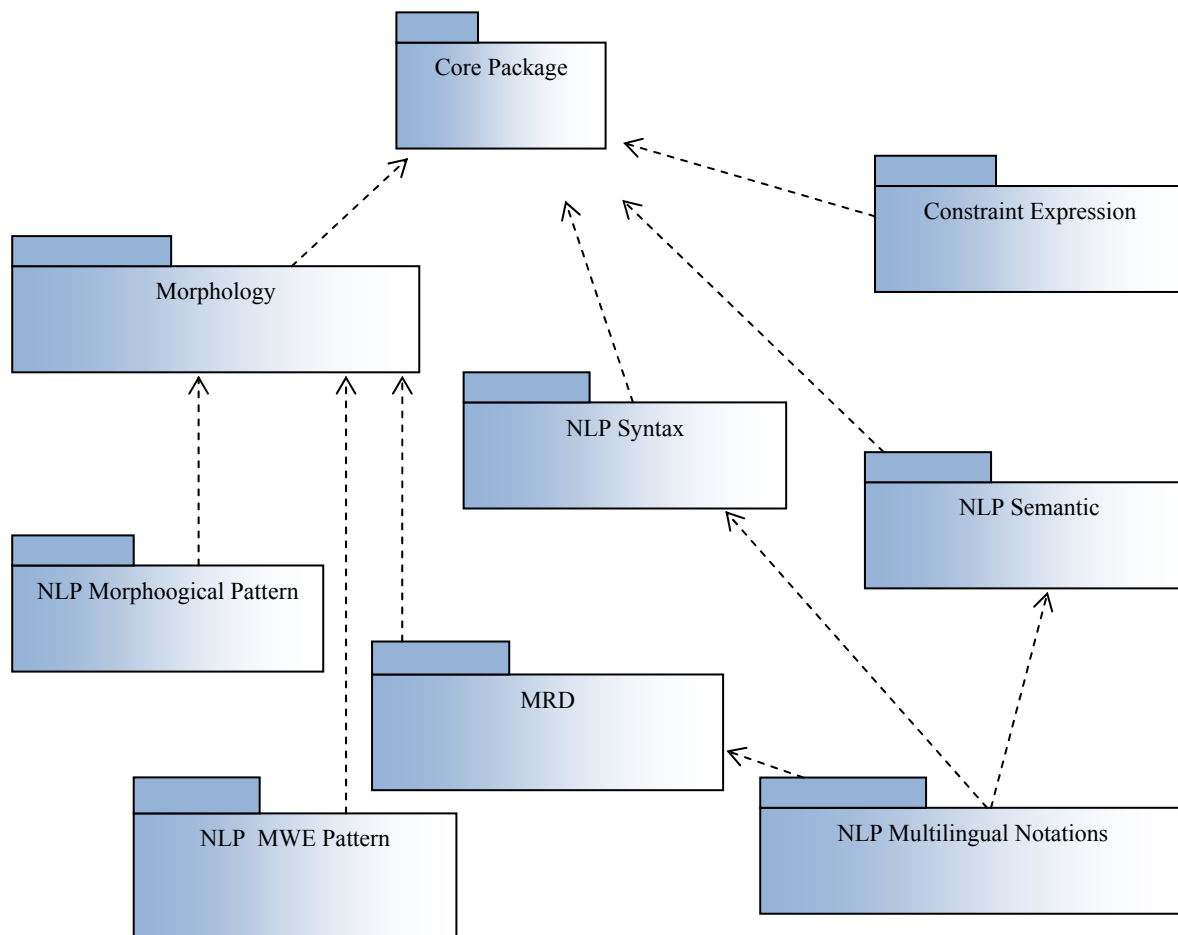
کلاس گزاره، کلاسی است که یک شرح روایتی را ارائه داده و تعریف را تصحیح و یا تکمیل می‌کند. یک نمونه تعریف می‌تواند دارای صفر یا چند نمونه گزاره باشد. یادآوری - یک مثال کامل برای این کلاس در پیوست ح آمده است.

۵-۲-۱۱ کلاس بازنمایی متن

کلاس بازنمایی متن، کلاسی است که یک محتوای متنی از یک تعریف یا گزاره را ارائه می‌دهد. زمانی که بیشتر از یک حالت املائی صحیح برای یک کلمه وجود داشته باشد، کلاس نمایش متن دارای یک رشته یونیکد نمایش‌دهنده یک محتوای متنی است، هم چنین در صورت نیاز، دارای جفت‌های صفت - مقدار از روی ویژگی‌های منحصر به فرد می‌باشد که زبان، خط و املائی خاصی را توصیف می‌کنند.

۵-۳ استفاده از بسط LMF

تمامی بسط‌ها با هسته مرکزی LMF از این نظر که هر بسط متصل به یک زیر مجموعه از کلاس‌های هسته مرکزی می‌باشد، مطابقت دارند. یک بسط نمی‌تواند برای بازنمایی داده‌های واژگانی مستقل از مجموعه هسته مرکزی مورد استفاده قرارگیرد. به نسبت نوع داده‌های زبان‌شناسی، یک بسط می‌تواند به بسط‌های دیگر وابسته باشد. از دیدگاه UML، یک بسط در واقع یک بسته UML است. بخش‌های متصل به بسط‌های مختلف در شکل دو شرح داده شده است.



شکل ۲- وابستگی‌های بین هسته‌های مرکزی و بسط‌های LMF

بسط‌های اضافی ممکن است به مرور زمان اضافه شوند. یک بسط جدید ممکن است مبنی بر خود مجموعه مرکزی LMF باشد و یا بر مبنای بسط موجود هسته مرکزی و یا ترکیبی از ساز و کارهای بسط هسته مرکزی و بسط‌های موجود باشد.

سازوکارهای بسط عبارتند از:

الف - ایجاد زیرکلاس‌ها بر اساس اصول مدل‌سازی UML؛

ب - افزودن کلاس‌های جدید؛

پ - محدودیت‌های نوع و درجه اهمیت تناظرها؛

ت - ویژگی‌های نقاط الحاقی متفاوت برای تناظرها؛

ث - گزینش‌های رده‌بندی داده‌ای^۱ (DCSs)؛

بسط‌های فعلی LMF در پیوست‌های الف، ب، پ، ت، ث، ج، چ، ح، خ، د، ذ، ر، ز، ژ، س و ش این استاندارد شرح داده شده است. پیوست‌های الف، پ، ث، چ، خ، ذ، ز و س حالت یکپارچه این استاندارد را تشکیل

1-Cardinality and type of associations

2-Data category selections

می‌دهد. پدیدآورندگان فرهنگ لغات بهتر است زیر مجموعه‌های این بسط‌های احتمالی را که با نیازهایشان مرتبط است، انتخاب کنند.

۴-۵ روش‌های انتخاب رده‌های داده‌ای LMF

۱-۴-۵ ویژگی‌های LMF

مدل‌های UML مانند LMF توسط صفات UML توصیف شده‌اند که اطلاعاتی راجع به ویژگی‌های خاص یا ویژگی‌های مرتبط با مدل را فراهم می‌کنند. تمامی صفات LMF در رده‌های داده‌ای پیچیده‌ای هستند. برای هر کلاس مفروض، تمامی صفات از یکدیگر متفاوتند. هر مقدار از هر صفت می‌تواند هم یک رده داده‌ای ساده و یا یک رشته یونیکد باشد. هر صفت فقط می‌تواند دارای یک مقدار باشد.

۲-۴-۵ ثبت رده داده‌ای (DCR)

ثبت رده داده‌ای، مجموعه‌ای از ویژگی‌های رده داده‌ای می‌باشد که در استاندارد ISO 12620 شرح داده شده است. طراحان هر فرهنگ لغت LMF باید هنگام ایجاد رده داده‌ای صفات به DCR اتکا کنند.

۳-۴-۵ انتخاب رده داده‌ای (DCS)

در مفهوم کلی، انتخاب رده داده‌ها شامل تمامی رده‌های داده‌ای مورد استفاده در یک دامنه مفروض در زمینه منابع زبانی می‌باشد. DCS هم چنین می‌تواند فهرستی از رده‌های داده‌ای را که در یک فرهنگ لغت LMF مفروض مورد استفاده قرار می‌گیرند، تهیه کرده، شرح دهد. DCS محدودیت‌های چگونگی نگاشت رده‌های داده‌ای را در کلاس‌های خاص مرتبط شرح می‌دهد.

۴-۴-۵ رده‌های داده‌ای معین شده توسط کاربران

پدیدآورندگان فرهنگ لغات، می‌توانند فهرستی از رده‌های داده‌ای جدیدی را برای پوشش دادن مفاهیم رده داده‌ای که مورد نیاز بوده و قابل دسترس در DCR نمی‌باشند، تعریف کنند. این مجموعه مکمل رده‌های داده‌ای باید مطابق با استاندارد ISO12620 باشد.

۵-۴-۵ مقایسه فرهنگ لغات

زمانی که دو فرهنگ لغت مطابق LMF، مبنی بر دو DCS متفاوت باشند، مقایسه DCS در هر فرهنگ لغت، چارچوبی را برای شناسایی این که چه نوع اطلاعاتی می‌توانند بین قالب‌ها رد و بدل شوند و یا این که چه چیزی در طول مدت تبدیل^۱ حذف خواهد شد، فراهم می‌سازد. زمانی که LMF برای شرح یک منبع موجود، مورد استفاده قرار می‌گیرد، لازم است که منبع موجود به رده‌های داده‌ای متناظر آن‌ها در DCR نگاشت شود.

۵-۵ فرایند LMF

LMF باید مطابق مراحل زیر مورد استفاده قرار گیرد:

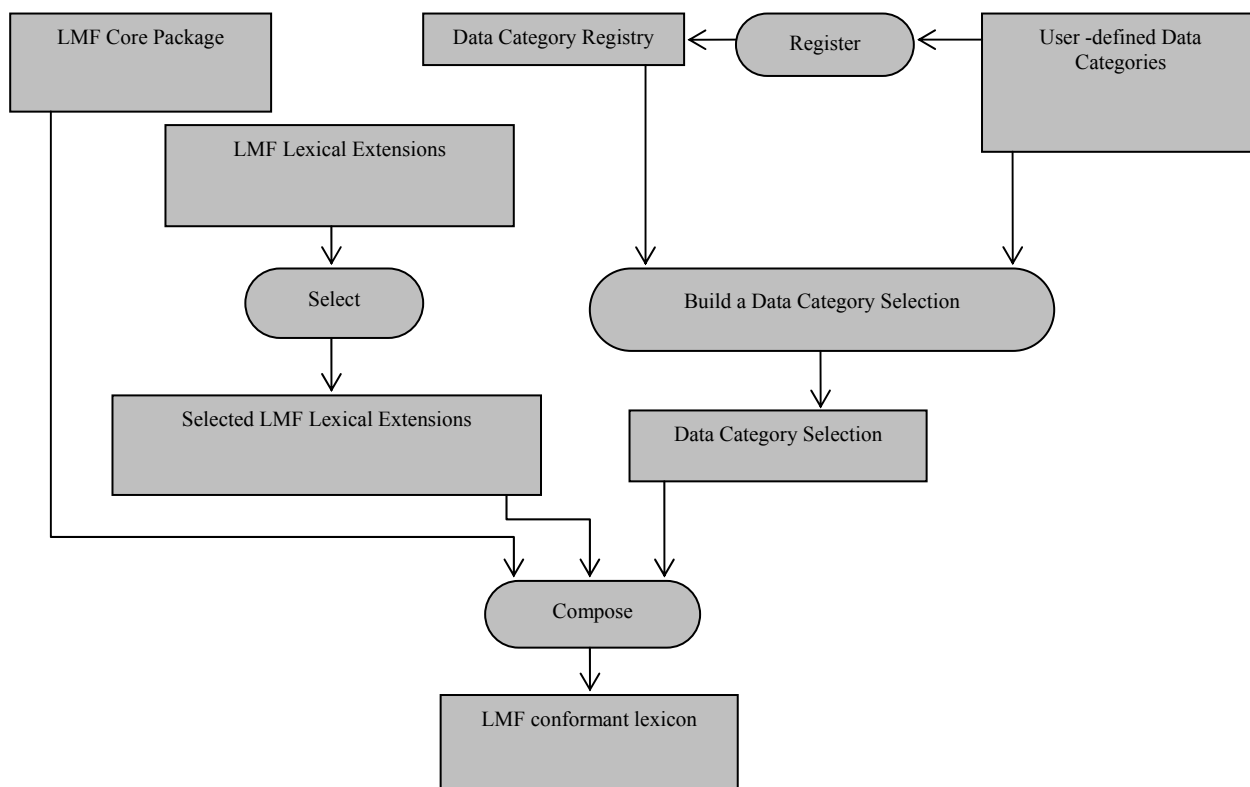
الف - مرحله اول

تعریف یک فرهنگ لغت مطابق با LMF

ب - مرحله دوم

پر کردن این فرهنگ لغت.

یک فرهنگ لغت مطابق با LMF به عنوان ترکیبی از یک مجموعه مرکزی LMF، صفر یا چند بسط واژگانی و همچنین بسته‌ای از رده‌های داده‌ها، تعریف شده است. ترکیب تمامی این موارد در نمودار فعالیت UML در شکل سه شرح داده شده است.



شکل ۳ - فرآیند LMF

پیوست الف (الزامی) بسط ساختواژه

الف-۱ اهداف

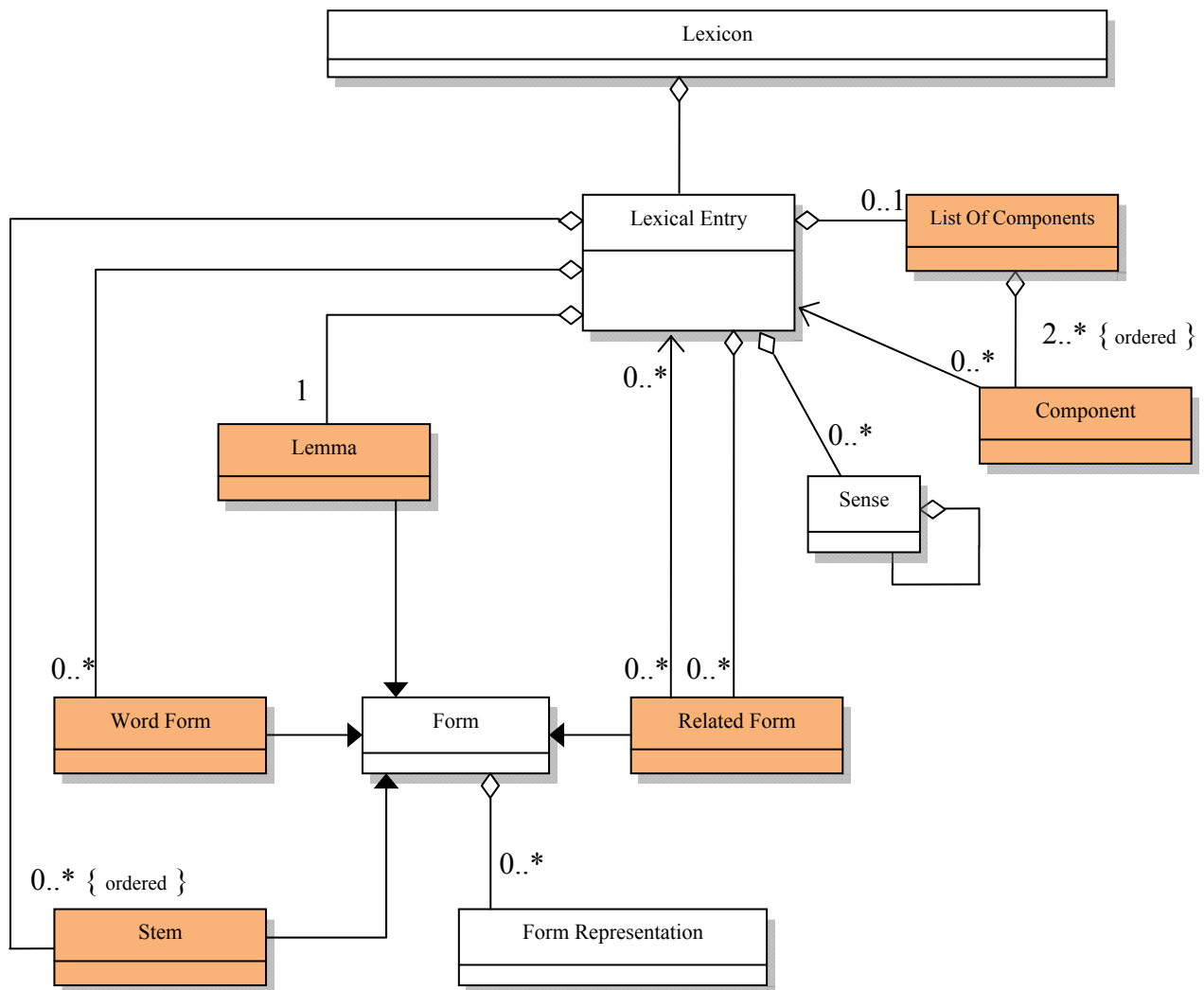
هدف بسط ساختواژه عبارت است از فراهم کردن ساز و کارهایی برای حمایت از توسعه فرهنگ‌های لغاتی که شامل وصف بسطی^۱ برای ساختواژه مدخل‌های واژگانی هستند.

مثال - در زبان تصریفی، "وصف بسطی" به این مفهوم است که تمامی حالات تصریفی به طور واضح در یک نمونه فرهنگ لغت شرح داده می‌شوند.

یادآوری - ساز و کارهای وصف بسطی ساختواژه در پیوست ذ (در الگوهای ساختواژه) مشخص شده است.

الف-۲ نمودار کلاس

شکل الف-۱ به شرح ساختار بسط ساختواژه پرداخته است.



شکل الف-۱- مدل کلاس ساختواژه

1-Extensional description

الف-۳ شرح مدل ساختواژه

مدل ساختواژه دو رده از زیرکلاس حالت را مدیریت می‌کند: زیرکلاس‌های حالت که نمایانگر مجموعه‌ای از گونه‌های مختلف قواعدی که یک واژه قاموسی انتزاعی را می‌سازند، و زیرکلاس‌های حالتی که می‌توانند به یک حالت در نمونه‌ای از یک مدخل واژگانی دیگری مرتبط باشند. کلاس‌های اول شامل بن‌واژه، حالت کلمه و ستاک می‌باشند. کلاس‌های دوم شامل حالت مرتبط می‌باشد. مدخل واژگانی به اجزای کلام محدود می‌شود.

الف-۳-۱ زیرکلاس‌های حالت

الف-۳-۱-۱ زیر کلاس بن‌واژه

زیر کلاس بن‌واژه، یک زیرکلاس از حالت می‌باشد که نمایانگر حالت کلمه است که به طور قراردادی برای تعیین مدخل واژگانی انتخاب می‌شود. کلاس بن‌واژه در یک تناظر کلی یک به یک^۱ با مدخل واژگانی می‌باشد به طوری که این مدخل واژگانی تکثر^۲ ماهوی کلاس حالت را تحت‌الشعاع قرار می‌دهد. بن‌واژه معمولاً با یکی از حالات صرفی ستاک یا MWE مانند ترکیب، عبارت اصطلاحی^۳ هم‌ارز می‌باشد. قرارداد انتخاب بن‌واژه ممکن است با زبان، خانواده زبان و یا گزینه ویرایشی تغییر کند.

الف-۳-۱-۲ کلاس حالت کلمه

کلاس حالت کلمه، زیر کلاس حالت بوده، نمایانگر حالتی است که یک واژه قاموسی می‌تواند هنگام استفاده در یک جمله و یا یک عبارت به خود بگیرد. بنابراین، کلاس حالت کلمه می‌تواند واژه‌های قاموسی ساده، مرکب و اصطلاحات چندکلمه‌ای را مدیریت کند.

الف-۳-۱-۳ کلاس ستاک

کلاس ستاک، زیر کلاسی از حالت است که نمایانگر یک واژه می‌باشد. تناظر کلی موجود بین یک مدخل واژگانی و یک ستاک مرتب می‌شود. بنابراین کلاس ستاک بخش‌های زیر واژه قاموسی را مدیریت می‌کند.

الف-۳-۱-۴ کلاس حالت مرتبط

کلاس حالت مرتبط، زیر کلاس حالت است که نمایانگر حالت کلمه یا واژی است که می‌تواند با مدخل واژگانی در یک یا چندین حالت (برای مثال اشتقاق، بن^۴) مرتبط باشد. حالت مرتبط می‌تواند دارای انواع مختلفی باشد. بی‌گمان حالت مرتبط، با کلاس مفهوم در مدخل واژگانی مرتبط است.

الف-۳-۲ کلاس فهرست مولفه‌ها

کلاس فهرست مولفه‌ها، کلاسی است که جنبه کلی یک عبارت چندکلمه‌ای را نمایش می‌دهد. این کلاس در یک رابطه کلی برابر صفر و یا برابر با مجموع روابط با کلاس مدخل واژگانی می‌باشد. هر نمونه‌ای از فهرست مولفه‌ها بایستی دست کم دارای دو مولفه باشد.

1-One to one aggregate association

2-Multiplicity

3-Idiomatic phrase

4-Root

ساز و کار موجود هم چنین می‌تواند به صورت برگشتی به کار رود، به عبارت دیگر یک عبارت چندکلمه‌ای که خود متشکل از عبارت‌های چندکلمه‌ای دیگر است. کلاس فهرست مولفه‌ها در بسته‌های الگوی ساختارهای و الگوی MWE نیز مورد استفاده قرار می‌گیرد.

الف-۳-۳

کلاس مولفه

کلاس مولفه، کلاسی است نمایانگر مرجعی برای مدخل واژگانی که برای هر مولفه واژگانی در یک کلاس فهرست مولفه‌ها گروه‌بندی شده است.

پیوست ب
(اطلاعاتی)
مثال‌های ساختواژه

ب-۱ مقدمه

این بسط مثال‌هایی از چگونگی توسعه مدل‌هایی برای فرهنگ‌های لغات ساختواژه MRD و NLP ارائه می‌کند.

ب-۲ مثال آرایش کلاس

کلاس‌ها ممکن است با صفات زیر آرایش شوند:

تفسیر	مثالی برای صفات	نام کلاس
یونیکد به عنوان مقادیر استفاده کنید. / حالت نوشتاری / و / حالت آواشناختی / از رشته‌های	حالت نوشتاری حالت آواشناختی تغییر جغرافیایی شما	بن‌واژه
زمانی که / حالت نوشتاری / به صورت "kitten" باشد، / تیره‌گذاری / به صورت "kit ten" خواهد بود. / عدد دستوری / احتمالاً به صورت / جمع ^۲ / خواهد بود.	حالت نوشتاری حالت آواشناختی تیره‌گذاری ^۱ عدد دستوری جنسیت دستوری زمان دستوری شخص	حالت کلمه
	حالت نوشتاری حالت آواشناختی نوع	حالت مرتبط
		حالت مولفه
		فهرست مولفه‌ها

ب-۳ مثالی از توضیح واژه قاموسی

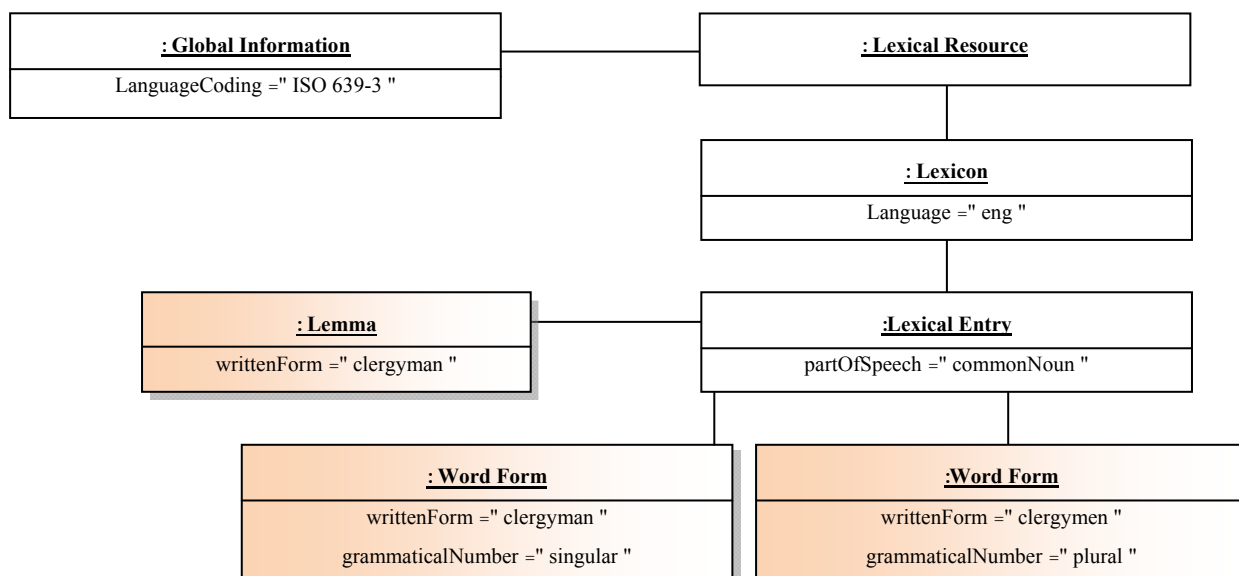
ب-۳-۱ مثالی از ساختواژه ساده

در مثال زیر مدخل واژگانی با یک بن‌واژه *clergyman* و دو حالت تصریفی *clergyman* و *clergymen* با یکدیگر مرتبط شده‌اند. کدگذاری زبان برای کل منبع واژگانی، با استفاده از استاندارد ISO 639-3 (که در شکل ب-۱ شرح داده شده است)^۳ تنظیم شده است.

1-Hyphenation

2-Plural

۳- برای این که فهم این شکل آسان‌تر شود، محتوی کادرهای سایه‌دار برای نمونه‌های مربوط به گروه‌هایی که در هسته کنونی تعریف شده‌اند، مورد استفاده قرار می‌گیرند. محتوی کادرهای مربوط به نمونه‌های گروه‌هایی که در هسته دیگر تعریف شده، سایه ندارند.



شکل ب ۱- نمونه‌ای از نمودار یک مثال

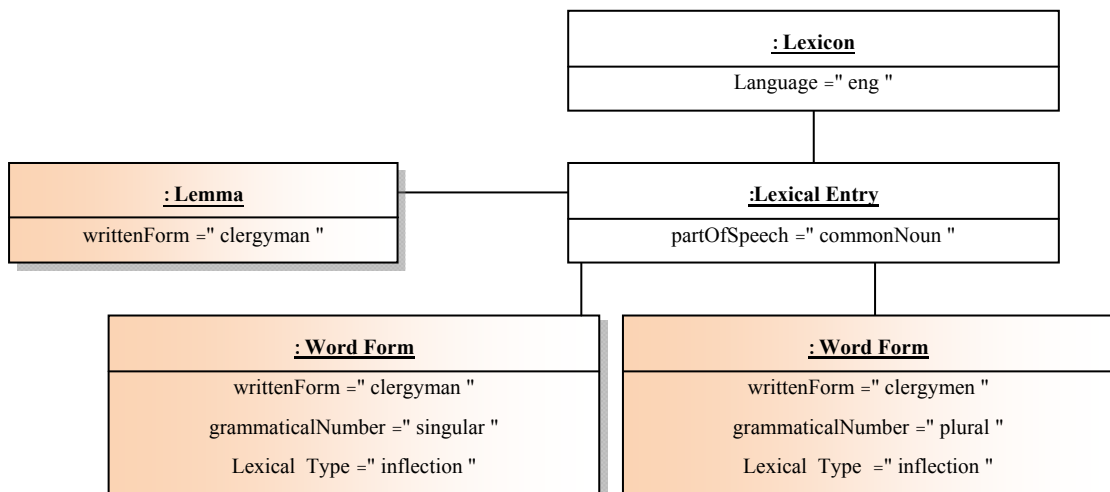
همان داده‌ها در قطعه XML به شرح زیر بیان شود:

```

<LexicalResource dtdVersion="16">
  <GlobalInformation>
    <feat att="languageCoding" val="ISO 639-3"/>
  </GlobalInformation>
  <Lexicon>
    <feat att="language" val="eng"/>
    <LexicalEntry>
      <feat att="partOfSpeech" val="commonNoun"/>
      <Lemma>
        <feat att="writtenForm" val="clergyman"/>
      </Lemma>
      <WordForm>
        <feat att="writtenForm" val="clergyman"/>
        <feat att="grammaticalNumber" val="singular"/>
      </WordForm>
      <WordForm>
        <feat att="writtenForm" val="clergymen"/>
        <feat att="grammaticalNumber" val="plural"/>
      </WordForm>
    </LexicalEntry>
  </Lexicon>
</LexicalResource>
  
```

هم چنین می‌توان نوع حالت کلمه را با افزودن یک صفت خاص نوع واژگانی^۱ همان‌طوری که در نمودار نمونه زیر دیده می‌شود، تعیین کرد. به شکل ب-۲ مراجعه کنید.

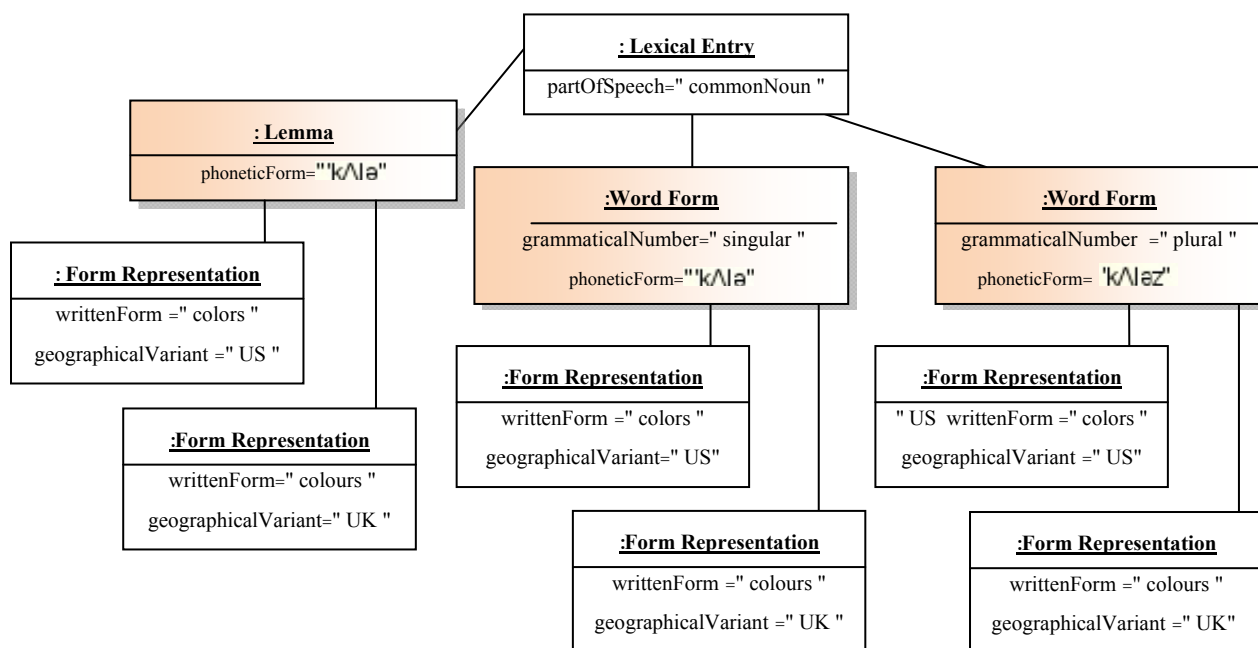
1- lexicalType



شکل ب- ۲- مثال کاملاً تخصصی از حالت کلمه

ب-۳-۲ مثالی برای گونه‌های مختلف منطقه‌ای^۱

گونه‌های مختلف منطقه‌ای می‌توانند در زبان انگلیسی با استفاده از کلاس نمایش حالت و بایک صفت مشترک حالت آواشناختی، همان‌طور که در شکل ب-۳ نشان داده شده است، مدل‌سازی شوند.

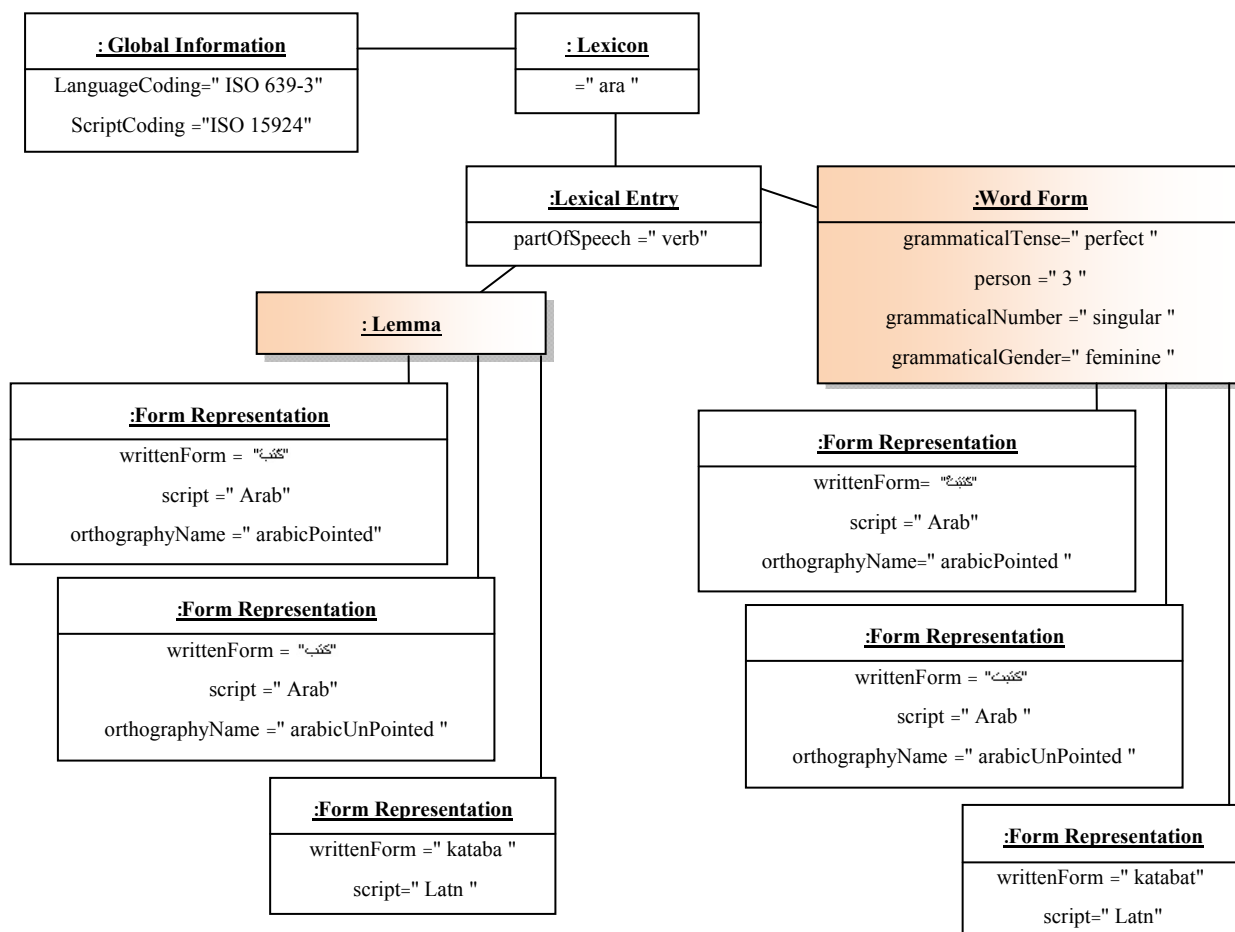


شکل ب- ۳- مثال گونه‌های مختلف منطقه‌ای با استفاده از بازنمایی حالت

1-Regional

ب-۳-۳ مثالی از چند دست خط و املا صحیح

در مثال زیر مدخل واژگانی با یک بن‌واژه‌ای مرتبط شده است که به سه روش متفاوت قادر به نمایش حالت کلمه می‌باشد. مدخل واژگانی همان طوری که در شکل ب-۴ دیده می‌شود، با یک حالت تصریفی که آن نیز به سه روش متفاوت قادر به نمایش حالت کلمه می‌باشد، مرتبط شده است.



شکل ب-۴-۴-۳ مثالی از چند خط و املا صحیح

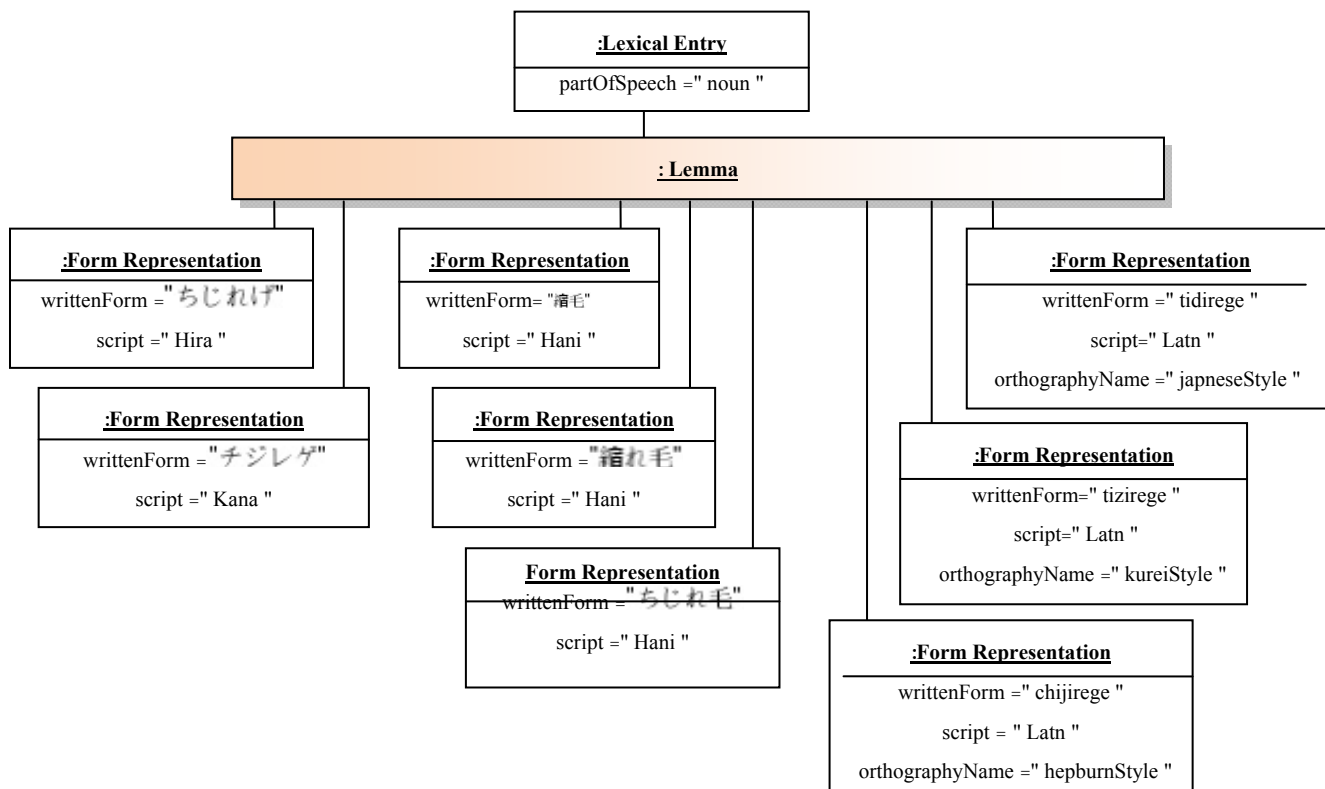
شایان ذکر است که این شیوه تنها روش انتخاب ممکن در زبان عربی نیست. شیوه دیگر، شرح حالات خط حرکت‌گذاری شده^۱ زبان عربی در فرهنگ لغت و ایجاد یک ساز و کار خارجی برای محاسبه خودکار حالات خط بدون حرکت^۲ زبان عربی و نویسه‌گردانی‌ها می‌باشد. در این مورد، به نمونه‌های بازنمایی حالت نیازی نمی‌باشد.

ب-۳-۴ مثالی برای خط‌ها، املاها و گونه‌های متعدد

تعداد نمونه‌های بازنمایی حالت ممکن است در زبان ژاپنی پر اهمیت باشد جایی که چهار نوع سیستم نوشتاری همزمان در این زبان اعمال شده و با هم ترکیب می‌شوند. هیراگانا^۳، کاتاکانا^۴،

- 1-Arabic pointed script forms
- 2-Unpointed
- 3-Hiragana
- 4-Katakana

کانژی^۱ و رمانیزه کردن^۲ آن‌ها. مجموعه‌ای از گونه‌های مختلف واژه با نام خط یکسان در مثال زیر با بازنمایی curly hair با یکدیگر ترکیب شده‌اند، به شکل ب-۵ مراجعه کنید.



شکل ب - ۵ - مثالی برای املاها و گونه‌های متعدد خطها

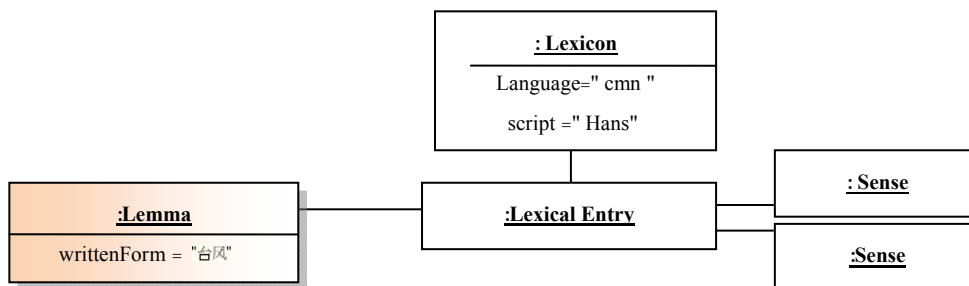
ب-۳-۵ مثال در زبان چینی

زبان چینی یک زبان گسسته می‌باشد. با مرور زمان سیستم نوشتاری زبان چینی تغییر یافته است. حرکت به سوی ساده‌سازی سیستم نوشتاری زبان چینی در سال ۱۸۹۰ شروع شد و در سال ۱۹۵۰ گسترش و ادامه یافته است. شیوه ساده‌سازی دربرگیرنده کاهش تعداد حرکات^۳ مورد استفاده حروف می‌باشد. در حال حاضر دو گونه مختلف واژه مورد استفاده قرار می‌گیرد. بر طبق استاندارد ISO 15924، کد خط برای گونه ساده شده *Hans* و برای گونه سنتی کلمه *Hant* می‌باشد.

مثال زیر، شرایطی را نشان می‌دهد که چگونه دو حالت سنتی مختلف برابر با یک حالت ساده شده امروزی می‌باشند. اگر کاربر بخواهد شرح حالات ساده شده را مختصرتر کند، یک بن‌واژه مستقل برای این عمل کافی می‌باشد، چنان چه نمودار زیر برای رفتار روی صحنه^۴ و تایفون^۵ نشان می‌دهد. اطلاعات زبان و خط برای

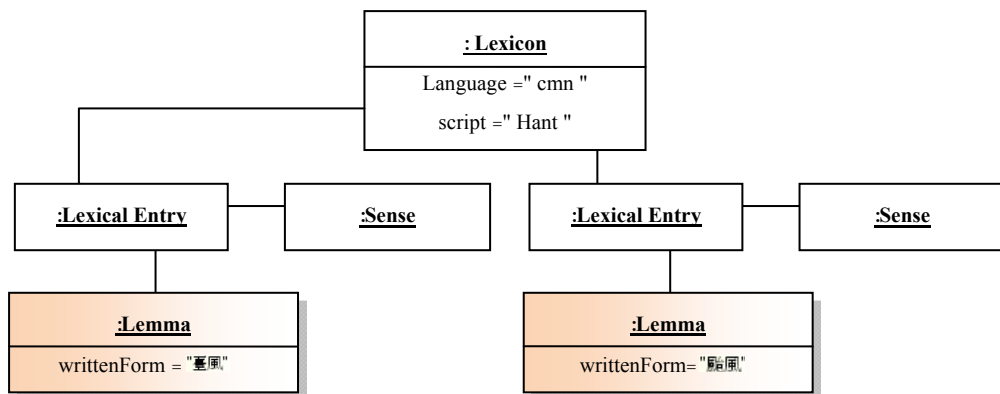
-
- 1-Kanji
 - 2-Romanization
 - 3-Strokes
 - 4- Behavior on the stage
 - 5-Typhoon

تمامی مدخل‌های واژگانی یکسان می‌باشند، از این رو، این صفات در نمونه فرهنگ لغت، در شکل ب-۶، جای می‌گیرند.



شکل ب-۶ - مثالی برای سیستم نوشتاری ساده شده چینی

اما اگر کاربر بخواهد حالات سنتی را شرح دهد، دو مدخل واژگانی مورد نیاز است چرا که دو حالت سنتی مشخص وجود دارد و معنی هر یک از این واژه‌های قاموسی با هم متفاوت است. به شکل ب-۷ رجوع کنید.

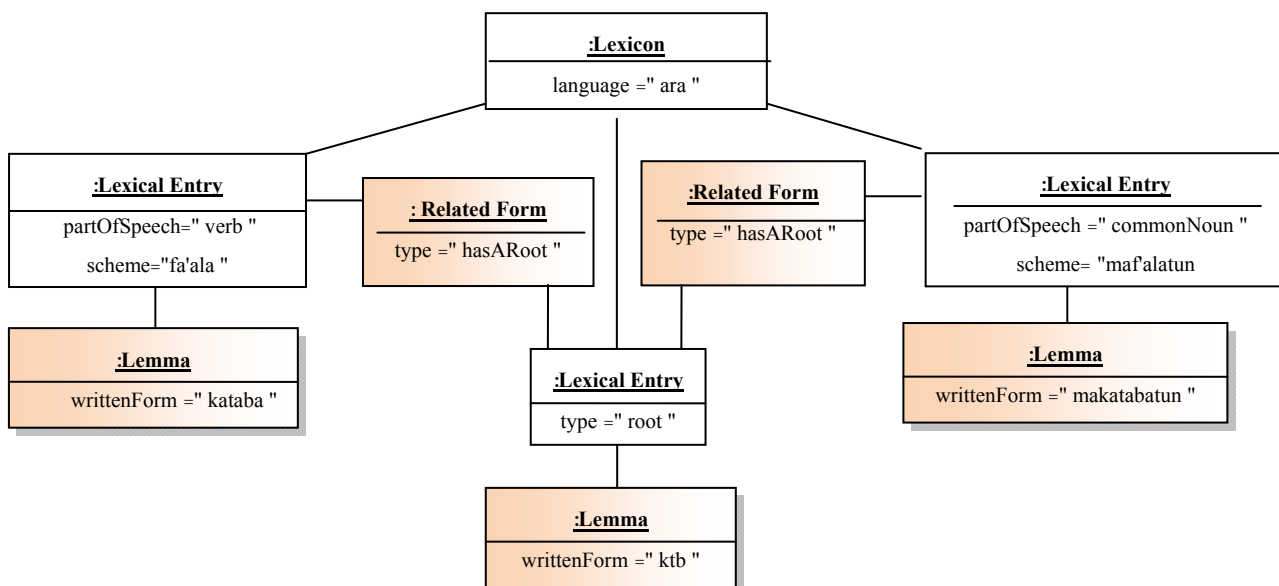


شکل ب-۷ - مثالی برای سیستم نوشتاری سنتی

شایان ذکر است که اگر کاربر بخواهد دو حالت سنتی و ساده شده امروزی را در همان فرهنگ لغت ترکیب کند صفت خط نمی‌تواند برای نمونه فرهنگ لغت تنظیم شود و بایستی برای هر نمونه بازنمایی حالت، همانند مثال قبلی برای زبان ژاپنی، تنظیم شود.

ب-۳-۶ مثال مدیریت بن زبان عربی

در این مثال، بن زبان عربی در نمونه مدخل واژگانی خاص مشترکی نمایش داده شده است. فعل *kataba* و اسم *maktabatun* هر دو به نمونه مدخل واژگانی *ktb* مرتبط می‌باشند. به شکل ب-۸ رجوع شود.



شکل ب - ۸ - مثالی برای مدیریت بن زبان عربی

پیوست پ

(الزامی)

بسط فرهنگ لغت قابل خواندن توسط ماشین

پ-۱ اهداف کلی

بسط فرهنگ لغت قابل خواندن توسط ماشین (MRD) یک بسته فرامدلی را برای بازنمایی داده‌های ذخیره شده در فرهنگ‌های لغات ماشین ایجاد می‌کند. این بسط، فرهنگ لغت الکترونیکی قابل خواندن توسط ماشین را هم برای استفاده انسان و هم پردازش ماشینی پشتیبانی می‌کند. از آن جایی که بسط MRD براساس هسته مرکزی LMF و بسط ساختوازی پایه‌ریزی شده است، امکان تبادل داده‌ها میان بسط مزبور و سایر بسط‌های LMF وجود دارد.

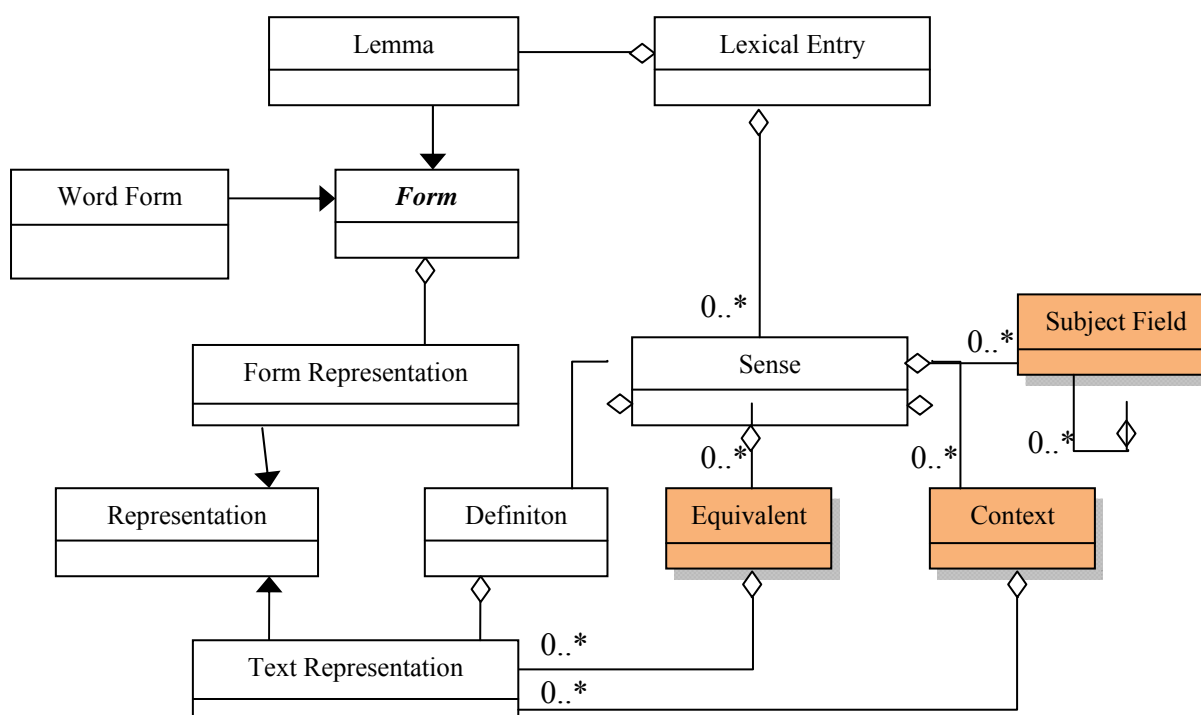
پ-۲ مجموعه فرهنگ لغت قابل خواندن توسط کامپیوتر

پ-۲-۱ اهداف

اهداف بسته MRD عبارت است از پشتیبانی از فرهنگ لغت دو زبانه و تک‌زبانه، برای مترجم‌های انسانی و هم چنین پشتیبانی از سیستم‌های سازمانی و پوشش زبان‌های متعدد و خانواده‌های زبانی، پشتیبانی از آماده‌سازی داده‌های واژگانی برای استفاده در سیستم‌های NLP و پشتیبانی مستقیم از سیستم‌های NLP (برای مثال: داده‌های واژگانی برای استخراج موجودیت‌های نامدار^۱).

پ-۲-۲ نمودار کلاس

بسط MRD مطابق با شکل پ-۱ سازمان‌دهی شده است.



شکل پ-۱ - مدل کلاس MRD

پ-۲-۳ شرح فرامدل MRD

فرامدل MRD بر اساس بسط ساختواژه NLP با اصلاحات زیر بنا شده است.

پ-۲-۳-۱ بسته MRD همه محدودیت‌های کلاس حالت مرتبط را کاهش می‌دهد (برای مثال، کلاس حالت مرتبط می‌تواند گونه‌بندی شود یا زیر کلاس‌های گستره کامل حالات کلمه مرتبط با مدخل واژگانی مانند مترادف‌ها، متضادها و کلمات اختصار را بپذیرد).

پ-۲-۳-۲ بسته کلاس‌های بیشتری را به کلاس مفهوم اضافه می‌کند. کلاس‌های معینی از MRD، همانند فرهنگ لغات دو زبانه، عموماً نیازمند نمونه کلاس مفهوم می‌باشند. کلاس‌های دیگر MRD، مانند فرهنگ لغات املائی، به نمونه کلاس مفهوم نیاز ندارند.

پ-۲-۴ کلاس معادل

در یک MRD دو زبانه، کلاس معادل نمایانگر ترجمه معادل یک حالت کلمه مدیریت شده توسط کلاس بن‌واژه است. کلاس معادل در تناظر صفر تا گروهی با کلاس مفهوم است، که به توسعه‌دهندگان فرهنگ لغات اجازه می‌دهد تا کلاس معادل را از فرهنگ لغت تک‌زبانه حذف کنند.

پ-۲-۵ کلاس متن

کلاس متن یک رشته متنی را نمایش می‌دهد که بستر مناسبی را برای استفاده حالت کلمه مدیریت شده به وسیله بن‌واژه، ایجاد می‌کند. کلاس زمینه در تناظر صفر تا گروهی زیاد با کلاس مفهوم است و ممکن است با صفر تا چند کلاس بازنمایی متن متناظر شوند که بازنمایی ترجمه معادل را در بیشتر از یک خط یا املائی صحیح را مدیریت می‌کند. یادآوری - ممکن است در متن از حالت تصریفی بن‌واژه استفاده شود.

پ-۲-۶ کلاس حوزه^۱ موضوع

حوزه موضوع، کلاسی است که نمایانگر رشته متنی است که اطلاعات حوزه یا وضعیت را فراهم می‌آورد. کلاس حوزه موضوع در تناظر صفر تا چند با کلاس مفهوم قرار دارد. کلاس حوزه موضوع مفاهیم سلسله مراتبی را منظور می‌کند که در آن یک نمونه حوزه موضوع ممکن است تخصصی‌تر از نمونه حوزه موضوع دیگر در همان مدخل واژگانی باشد.

پ-۲-۷ کلاس بازنمایی متن

بازنمایی متن زیر کلاسی از کلاس بازنمایی حالت می‌باشد. کلاس بازنمایی متن با زیر کلاس‌های کلاس مفهوم و نه با مدخل واژگانی مرتبط می‌شود. نمونه کلاس بازنمایی متن شامل املائی صحیح خاص و هم چنین یک یا چند رده‌های داده‌ای که صفات املائی صحیح را شرح می‌دهند، می‌باشند. یادآوری - نمونه‌های بازنمایی متن می‌توانند زبان‌ها و خط‌های مختلفی را داخل حوزه کلاس گروهی نمایش دهد.

پیوست ت (اطلاعاتی)

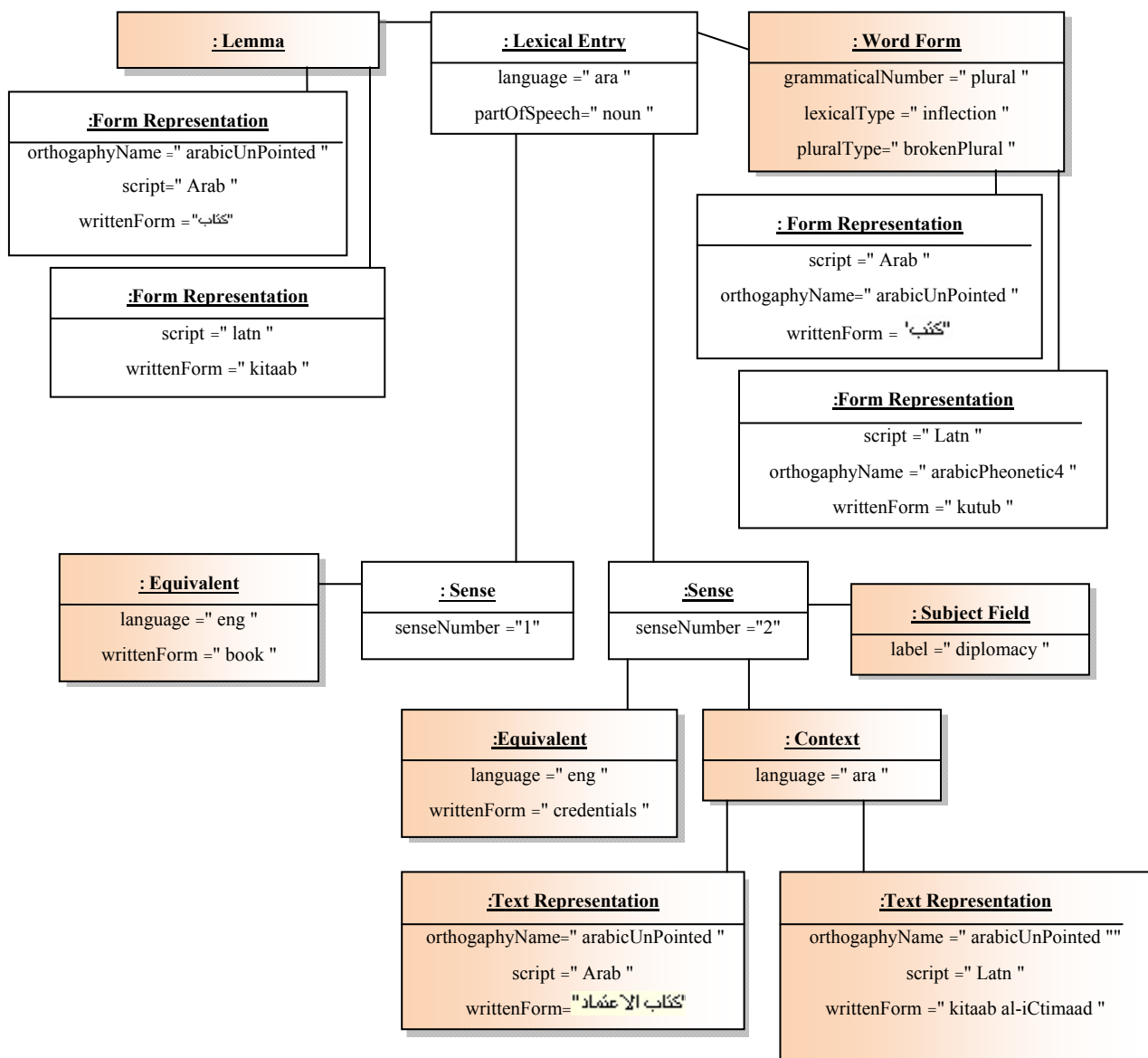
مثال‌هایی برای فرهنگ لغت قابل خواندن توسط ماشین

ت-۱ مثال MRD

ت-۱-۱ مثال MRD دوزبانه با بازنمایی‌های دو زبانه

مثال MRD دو زبانه در شکل ت-۱، یک مدخل را که شامل واژه عربی "kitaab" و دو معادل در زبان انگلیسی "book" (رایج‌ترین معنای کلمه) و "credentials" می‌باشد را، نشان می‌دهد. آوانویسی، اطلاعات بیشتری را در باره تلفظ کلمات و متن‌ها که قابل استخراج از خط زبان عربی هستند، برای کاربر فراهم می‌آورد. در این مثال، کلاس حالت کلمه، اطلاعاتی را در باره حالت و تلفظ جمع مکسر زبان عربی، که یک حالت تصریفی بی‌قاعده است، ارائه می‌دهد. تصمیم به قرار دادن کلاس بازنمایی حالت، یک گزینه ویرایشی بوده و از اهداف توسعه‌دهندگان فرهنگ لغات می‌باشد. اگر هدف، تولید یک MRD عربی - انگلیسی که فقط شامل خط عربی برای حالت‌های کلمات عربی بود، در این صورت قراردادن کلاس بازنمایی حالت الزامی نداشت، به شکل ت-۱^۱ رجوع شود.

۱- برای این که فهم این شکل آسان‌تر شود، محتوی کادرهای سایه‌دار برای نمونه‌های مربوط به گروه‌هایی که در بسته کنونی تعریف شده‌اند، مورد استفاده قرار می‌گیرند. محتوی کادرهای مربوط به نمونه‌های گروه‌هایی که در بسته دیگر تعریف شده، سایه ندارند.



شکل ت - ۱ - مثالی از یک MED دوزبانه

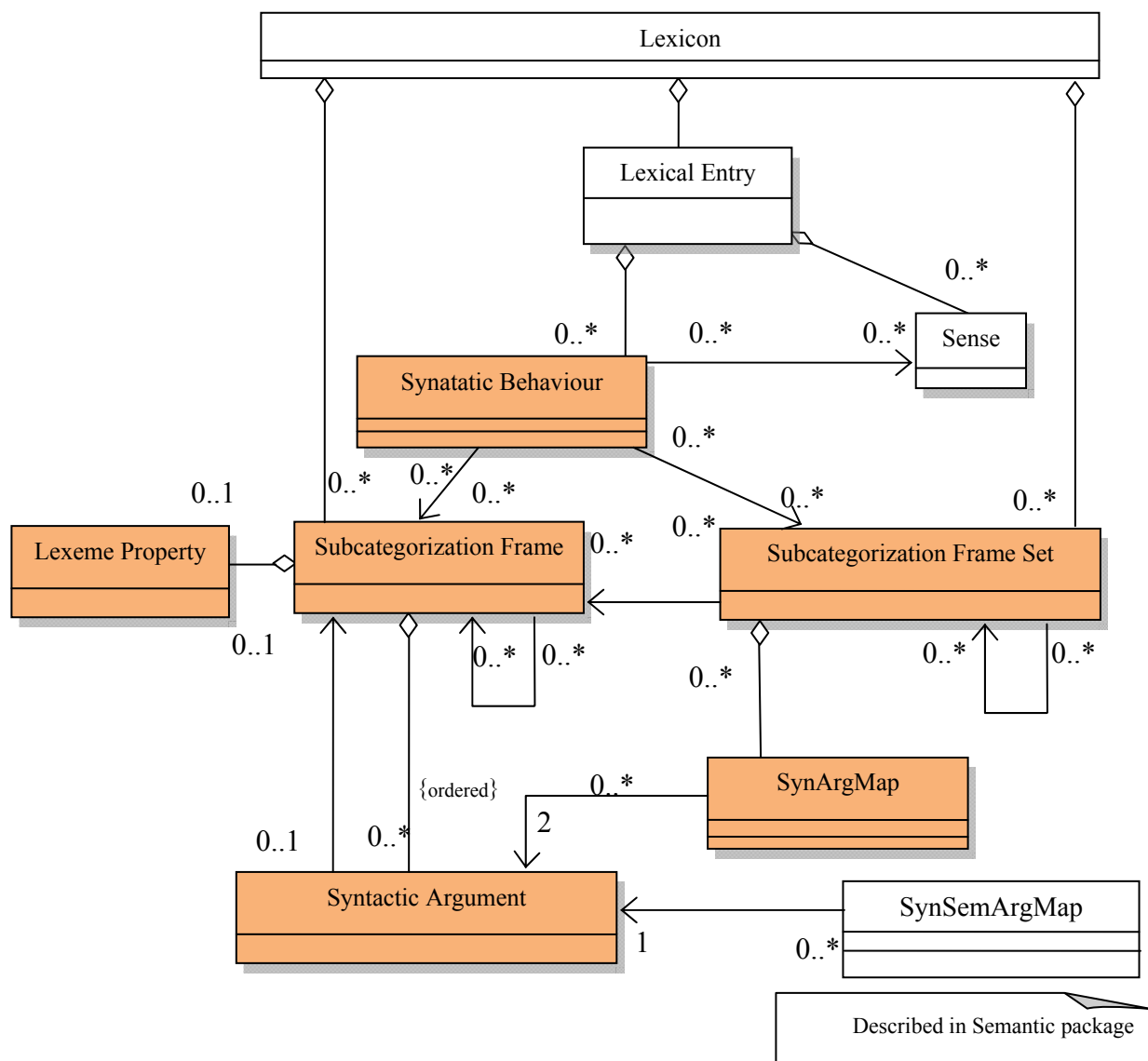
پیوست ث
(الزامی)
بسط نحوی NLP

ث-۱ اهداف

هدف این پیوست شرح ویژگی‌های یک واژه قاموسی هنگام ترکیب با واژه‌های قاموسی دیگر در یک جمله، می‌باشد. زمانی که یک مدخل واژگانی در فرهنگ لغتی ثبت می‌شود، ویژگی‌های نحوی، شرح نحوی آن را تشکیل می‌دهند.

این پیوست ویژگی‌های نحوی خاص واژه‌های قاموسی را شرح می‌دهد و به شرح دستور زبان کلی نمی‌پردازد.
ث-۲ نمودار کلاس

بسط نحوی NLP مطابق با شکل ث-۱ سازمان‌دهی شده است.



شکل ث-۱ - مدل نحوی

ث-۳-۱ کلاس عملکرد^۱ نحوی

کلاس عملکرد نحوی، کلاسی است که نمایانگر یکی از عملکردهای احتمالی یک واژه قاموسی می‌باشد. نمونه عملکرد نحوی به نمونه مدخل واژگانی و هم چنین به طور اختیاری به نمونه مفهوم پیوسته است. حضور در یک فرهنگ لغت مفروض از نمونه عملکرد نحوی برای یک مدخل واژگانی به این مفهوم می‌باشد که این واژه قاموسی این عملکرد را در زبان فرهنگ لغت می‌تواند داشته باشد.

شرح نحوی اختیاری می‌باشد، یعنی شرح ساختواژه و معنایی بدون شرح نحوی ممکن است. نمونه‌های مدخل واژگانی، عملکرد نحوی و مفهوم، رئوس یک مثلث نمایانگر ساختواژه، نحو و معنا را تشکیل می‌دهند. شرح جزئیات عملکرد نحوی مدخل واژگانی توسط نمونه چارچوب زیر رده‌بندی تعریف شده است.

ث-۳-۲ کلاس چارچوب زیر رده‌بندی

کلاس چارچوب زیر رده‌بندی، کلاسی است نمایانگر ساختار نحوی. یک نمونه چارچوب زیر رده‌بندی با تمامی نمونه‌های مدخل واژگانی که عملکرد نحوی یکسانی را در همان زبان دارند، مشترک است. چارچوب زیر رده‌بندی می‌تواند روابط و ویژگی‌هایش را توسط پیونده انعکاسی^۲ از چارچوب زیر رده‌بندی عام‌تر به ارث ببرد. بنابراین، ترکیب ساختار سلسله مراتبی نمونه‌های چارچوب زیر رده‌بندی امکان پذیر خواهد بود.

ث-۳-۳ کلاس ویژگی واژه قاموسی

کلاس ویژگی واژه قاموسی، کلاسی است نمایانگر گره مرکزی^۳ چارچوب زیر رده‌بندی و هم چنین کلاسی است که به نمونه مدخل واژگانی کنونی ارجاع دارد. نمونه ویژگی واژه قاموسی متصل به نمونه چارچوب زیر رده‌بندی با تمامی واژه‌های قاموسی که عملکرد نحوی یکسانی دارند، در اشتراک است.

ث-۳-۴ کلاس آرگمان نحوی

کلاس آرگمان نحوی، کلاسی است نمایانگر آرگمان مربوط به چارچوب زیر رده‌بندی مفروض. آرگمان نحوی می‌تواند به طور بازگشتی با نمونه چارچوب زیر رده‌بندی به منظور شرح کامل آرگمان‌های مرکب و پیچیده در ارتباط باشد. آرگمان‌های نحوی ارتباط با آرگمان معنایی را به وسیله نمونه نگاشت^۴ آرگمان نحوی معنایی^۴ میسر می‌سازند.

ث-۳-۵ کلاس مجموعه چارچوب زیر رده‌بندی

کلاس مجموعه چارچوب زیر رده‌بندی، کلاسی است که دسته‌ای از ساختارهای نحوی و همچنین روابط احتمالی بین این ساختارها را نشان می‌دهد. دسته چارچوب زیر رده‌بندی می‌تواند روابط و ویژگی‌هایش را توسط پیونده انعکاسی از دسته چارچوب زیر رده‌بندی بخصوص دیگری بگیرد. بنابراین یکپارچه کردن ساختار سلسله مراتب نمونه‌های دسته چارچوب زیر رده‌بندی امکان‌پذیر خواهد بود.

1-Behaviour
2-Reflexive link
3-Central node
4-SynSemArgMap

مجموعه چارچوب زیر رده‌بندی، ساختارهای نحوی مختلفی را که به طور مکرر برای دسته‌های معین واژه‌های قاموسی ظاهر می‌شوند، گروه‌بندی می‌کند. هدف، تجزیه‌کردن^۱ شرح‌های نحوی و حفظ حداقل نمونه‌های عملکرد نحوی در فرهنگ لغت می‌باشد.

ث-۳-۶ کلاس نگاشت آرگمان نحوی

کلاس نگاشت آرگمان نحوی، کلاسی است نمایانگر روابطی که نمونه‌های مختلف آرگمان نحوی مربوط به نمونه یکسان دسته چارچوب زیر رده‌بندی را نگاشت می‌کند.

1-Factorize

پیوست ج

(اطلاعاتی)

مثال‌های نحو NLP

ج-۱ مثال آرایش کلاس

کلاس‌ها ممکن است با ویژگی‌های زیر آرایش شوند:

نام کلاس	مثالی برای ویژگی‌ها	توضیح
عملکرد نحوی	شناسه ^۱ برچسب ^۲	
چارچوب زیر رده‌بندی	شناسه برچسب تفسیر	
ویژگی واژه قاموسی	اجزای کلام حالت صدا فعل کمکی موقعیت	رده موقعیت / داده / ممکن است موقعیت واژه قاموسی را در جمله، با توجه به آرگمان‌های نحوی تعیین می‌کند.
آرگمان نحوی	عملکرد نحوی ساختمان نحوی معرف ^۳ برچسب محدودیت ^۴ مثال	رده عملکرد نحوی / داده ممکن است مقادیری همچون /فاعل/ یا /مفعول/ دارد. /ساختمان نحوی/ ممکن است مقادیری همچون /NP/ و یا /PP/ به ترتیب برای عبارت اسمی و عبارت قیدی ^۵ می‌باشد. /معرف/ ممکن است حرف اضافه‌ای را که برای معرفی /ساختمان نحوی/ نیاز است را تعیین می‌کند.
دسته چارچوب زیر گروه‌بندی	شناسه برچسب تفسیر	
نگاشت آرگمان نحوی	تفسیر	

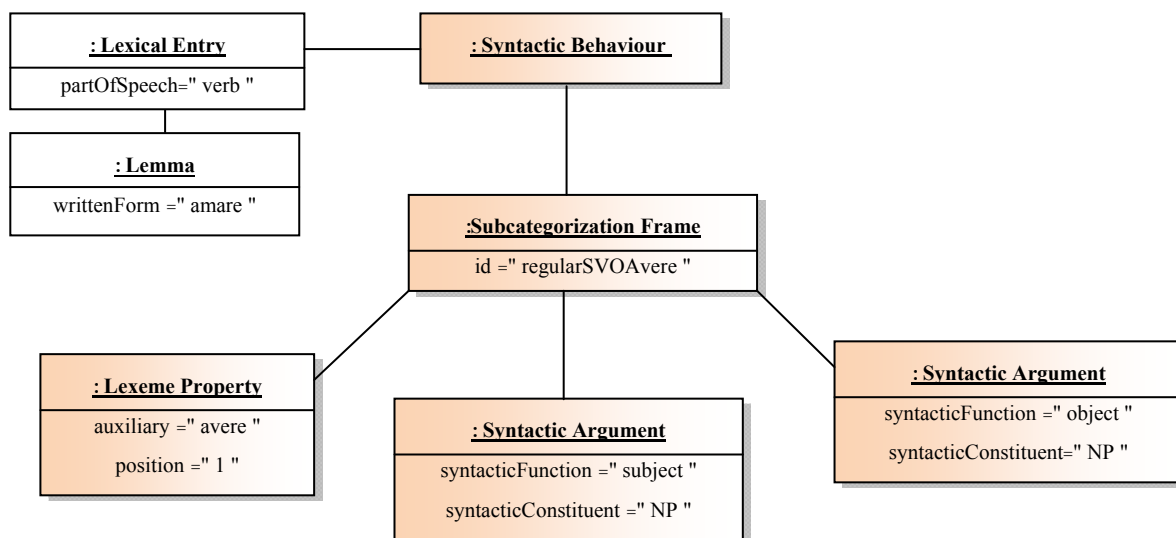
ج-۲ مثال‌های شرح واژه قاموسی

ج-۲-۱ مثالی در زبان ایتالیایی

مثال نشان‌داده شده در شکل ج-۱ از فرهنگ لغت Parole/CLIPS گرفته شده است. در این مثال فقط از ساختارهای نحوی استفاده شده است، و اطلاعات معنایی شرح داده نشده است. ساختار نحوی شرح داده

1-Id
2-Lable
3-Introducer
4-Restriction
5-Prepositional Phrase

شده، یک ساختار نسبتاً ساده در زبان ایتالیایی می‌باشد، جایی که هم فاعل و هم مفعول مستقیم هر دو دارای ویژگی رده داده ساده/عبارت اسمی/ می‌باشند. نمونه ویژگی واژه قاموسی، فعلی را شرح می‌دهد که دارای فعل کمکی *avere* است. یک مثال ساده از یک چنین ساختاری *Gianni ama Maria* می‌باشد، به شکل ج-۱^۱ مراجعه کنید.



شکل ج-۱ - نمودار نمونه در زبان ایتالیایی

همان داده‌ها در قطعه XML به شرح زیر بیان شده است:

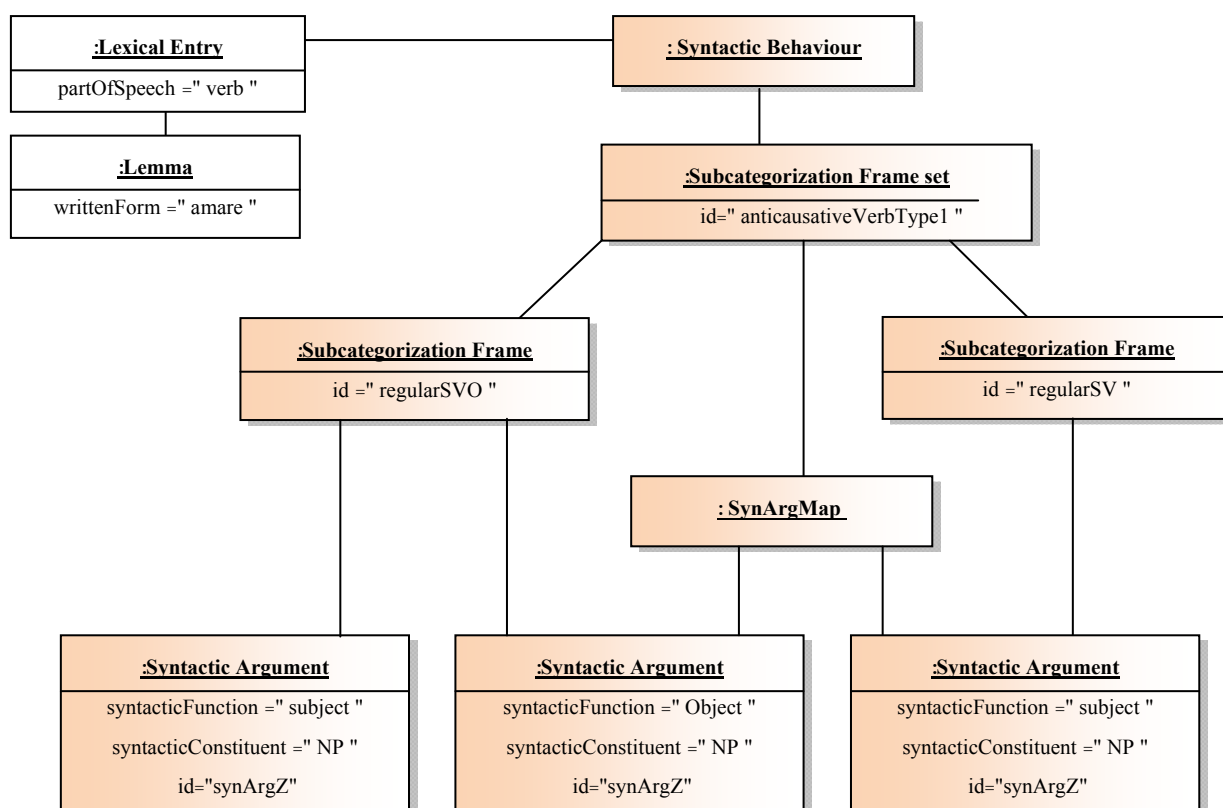
```

<LexicalEntry>
  <feat att="partOfSpeech" val="verb"/>
  <Lemma>
    <feat att="writtenForm" val="amare"/>
  </Lemma>
  <SyntacticBehaviour subcategorizationFrames="regularSVOAvere"/>
</LexicalEntry>
<SubcategorizationFrame id="regularSVOAvere">
  <LexemeProperty>
    <feat att="auxiliary" val="avere"/>
    <feat att="position" val="1"/>
  </LexemeProperty>
  <SyntacticArgument>
    <feat att="syntacticFunction" val="subject"/>
    <feat att="syntacticConstituent" val="NP"/>
  </SyntacticArgument>
  <SyntacticArgument>
    <feat att="syntacticFunction" val="object"/>
    <feat att="syntacticConstituent" val="NP"/>
  </SyntacticArgument>
</SubcategorizationFrame>
  
```

۱- برای این که فهم این شکل آسان‌تر شود، محتوی کادرهای سایه‌دار برای نمونه‌های مربوط به مقوله‌هایی که در بسته کنونی تعریف شده‌اند، مورد استفاده قرار می‌گیرند. محتوی کادرهای مربوط به نمونه‌های مقوله‌هایی که در بسته دیگر تعریف شده، سایه ندارند.

ج-۲-۲ مثالی در زبان انگلیسی

در زبان انگلیسی، استفاده از فقط یک دسته چارچوب زیر رده‌بندی برای فعل‌های ضدکمی^۱ ممکن می‌باشد. برای مثال فعل *boil* در جمله‌های *he boils a kettle of water* و *the kettle boils*، به این صورت، این فعل تنها به وسیله یک عملکرد نحوی، به جای دو عملکرد، شرح داده شده است. بنابراین همان طوری که در شکل ج-۲ مشاهده می‌شود، فقط به یک نمونه از مجموعه چارچوب زیر رده‌بندی نیاز است.



شکل ج - ۲ - نمودار نمونه در زبان انگلیسی

1-Anticausative

همان داده‌ها در قطعه XML به شرح زیر بیان شده است:

```
<LexicalEntry>
  <feat att="partOfSpeech" val="verb"/>
  <Lemma>
    <feat att="writtenForm" val="boil"/>
  </Lemma>
  <SyntacticBehaviour subcategorizationFrameSets="anticausativeVerbType1"/>
</LexicalEntry>
<SubcategorizationFrameSet id="anticausativeVerbType1"
  subcategorizationFrames="regularSVO regularSV">
  <SynArgMap arg1="synArgY" arg2="synArgZ"/>
</SubcategorizationFrameSet>
<SubcategorizationFrame id="regularSVO">
  <SyntacticArgument id="synArgX">
    <feat att="syntacticFunction" val="subject"/>
    <feat att="syntacticConstituent" val="NP"/>
  </SyntacticArgument>
  <SyntacticArgument id="synArgY">
    <feat att="syntacticFunction" val="object"/>
    <feat att="syntacticConstituent" val="NP"/>
  </SyntacticArgument>
</SubcategorizationFrame>
<SubcategorizationFrame id="regularSV">
  <SyntacticArgument id="synArgZ">
    <feat att="syntacticFunction" val="subject"/>
    <feat att="syntacticConstituent" val="NP"/>
  </SyntacticArgument>
</SubcategorizationFrame>
```

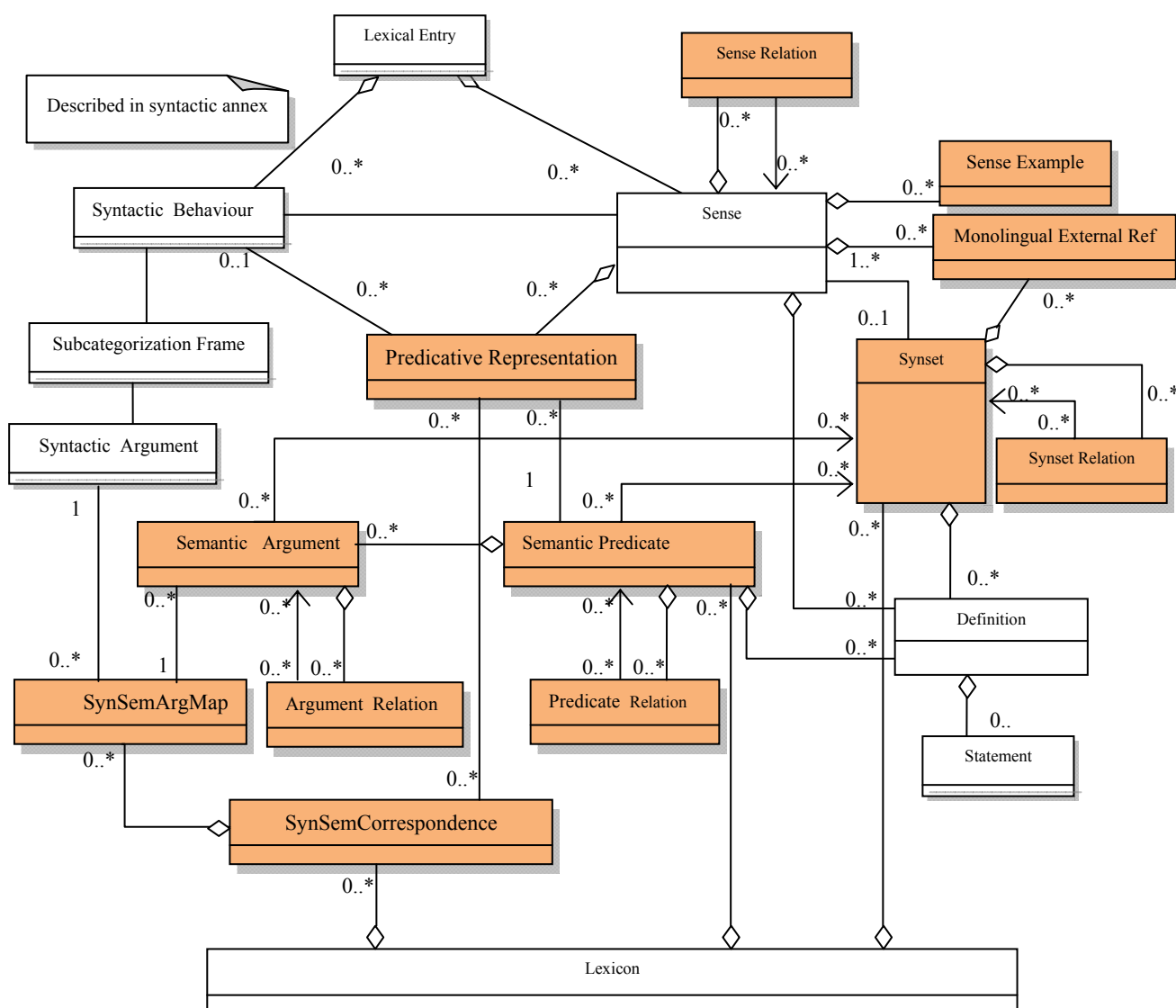
**پیوست چ
(الزامی)
بسط معنایی NLP**

چ ۱-اهداف

هدف این پیوست شرح یک مفهوم و ارتباط آن با مفاهیم دیگر متعلق به همان زبان است. به خاطر برهم کنش‌های ظریف و پیچیده بین نحو و معنا در بیشتر زبان‌ها، این پیوست ارتباط با نحو را فراهم می‌آورد. رابطه مفاهیم متعلق به زبان‌های مختلف با استفاده از پیوست نشانه‌گذاری چند زبانه شرح داده خواهد شد.

چ ۲-نمودار کلاس

بسط معنایی NLP مطابق با شکل چ-۱ سازمان‌دهی شده است.



شکل چ ۱- مدل معنایی

چ-۳ ارتباط با هسته مرکزی

کلاس مفهوم در هسته مرکزی مشخص شده است. این کلاس مفهوم به کلاس مدخل واژگانی اضافه شده است. بنابراین، یک نمونه مفهوم در میان دو نمونه مدخل واژگانی مختلف به اشتراک گذاشته نمی‌شود.

چ-۴ شرح مدل معنایی

چ-۴-۱ کلاس دسته نحوی^۱

کلاس دسته نحوی، کلاسی است نمایانگر دسته‌ای از معانی مشترک در همان زبان. دسته نحوی مترادف‌های تشکیل‌دهنده دسته را به یکدیگر مرتبط می‌سازد. یک نمونه دسته نحوی مفاهیم نمونه‌های مداخل واژگانی مختلفی را با اجزای کلام یکسان به یکدیگر مرتبط می‌سازد.

مثال - در نرم افزار WordNet نسخه 2.1، دسته نحوی "12100067" معنای *oak* و *oak tree* را که دو مترادف در نظر گرفته می‌شوند، در یکدیگر گروه‌بندی می‌کند.

چ-۴-۲ کلاس رابطه دسته نحوی

کلاس رابطه دسته نحوی، کلاسی است نمایانگر رابطه جهت‌دار^۲ بین نمونه‌های دسته نحوی.

چ-۴-۳ کلاس رابطه مفهوم

کلاس رابطه مفهوم، کلاسی است نمایانگر رابطه جهت‌دار بین نمونه‌های مفاهیم.

چ-۴-۴ کلاس مثال مفهوم

کلاس مثال مفهوم، کلاسی است که برای توضیح معنای بخصوص یک نمونه مفهوم مورد استفاده قرار می‌گیرد. یک مفهوم می‌تواند دارای صفر یا چند مثال باشد.

مثال - برای یک مدخل واژگانی عبارت چند کلمه‌ای "سازمان مردم نهاد (سمن)"^۳ "یک مثال مفهوم می‌تواند "سازمان عمومی عفو بین‌الملل"^۴ باشد.

چ-۴-۵ کلاس گزاره معنایی

کلاس گزاره معنایی، کلاسی است نمایانگر یک معنای انتزاعی به همراه تناظر آن با کلاس آرگمان معنایی. یک نمونه گزاره معنایی برای بازنمایی معنای رایج بین مفاهیم مختلف که لزوماً کاملاً مترادف نمی‌باشند، مورد استفاده قرار می‌گیرد. این مفاهیم ممکن است با نمونه‌های مدخل واژگانی که اجزای کلام متفاوتی دارند، در ارتباط باشند. نمونه گزاره معنایی به وسیله یک رابطه با یک یا چند نمونه دسته نحوی گونه‌بندی می‌شود. نمونه گزاره معنایی به نمونه فرهنگ لغت مفروض متعلق است.

یادآوری - در یک نمونه مدخل واژگانی برای فعل *to buy* به مفهوم "خریدن چیزی با پرداخت پول" نمونه گزاره معنایی توسط دو آرگمان معنایی تعریف می‌شود: یکی برای شخصی که می‌خرد و دیگری برای چیزی که خریده می‌شود. نمونه مدخل واژگانی دیگری می‌تواند برای *buyer* ثبت شود و به همان گزاره پیوند داده شود.

چ-۴-۶ کلاس بازنمایی گزاره‌ای

کلاس بازنمایی گزاره‌ای، کلاسی است نمایانگر ارتباط بین مفهوم و کلاس‌های گزاره معنایی.

1-Synset

2-Oriented relationship

3-Non-governmental organization (NGO)

4-Amnesty International

مثال - در مثال قبلی (چ-۴-۵)، ارتباط بین مفهوم فعل (یعنی *to buy*) و گزاره، به وسیله *master* نشانه‌گذاری می‌شود. ارتباط بین مفهوم اسم (یعنی *buyer*) و گزاره برای مثال به صورت *agentiveNominalization* نشانه‌گذاری می‌شود.

چ-۴-۷ کلاس آرگمان معنایی

کلاس آرگمان معنایی، کلاسی است نمایانگر آرگمان گزاره معنایی مفروض. یک نمونه آرگمان معنایی توسط یک رابطه با یک یا چند نمونه دسته نحوی شناسایی می‌شود.

مثال - در مثال مذکور، در زیر عبارت کلاس گزاره معنایی (چ-۴-۵)، گزاره می‌تواند دارای دو نمونه آرگمان معنایی باشد: یکی برای کسی که می‌خرد و دیگری برای چیزی که خریده می‌شود.

چ-۴-۸ کلاس رابطه آرگمان

کلاس رابطه آرگمان، کلاسی است نمایانگر یک رابطه جهت‌دار بین نمونه‌های آرگمان معنایی مربوط به همان نمونه گزاره‌ای.

چ-۴-۹ کلاس نگاشت آرگمان معنایی نحوی

کلاس نگاشت آرگمان معنایی نحوی، کلاسی است نمایانگر روابط بین یک آرگمان معنایی و یک آرگمان نحوی.

چ-۴-۱۰ کلاس تطابق معنایی نحوی^۱

کلاس تطابق معنایی نحوی، کلاسی است نمایانگر دسته‌ای از نمونه‌های نگاشت آرگمان معنایی و نحوی بین نمونه چارچوب زیر رده‌بندی مفروض.

چ-۴-۱۱ کلاس رابطه گزاره‌ای

کلاس رابطه گزاره‌ای، کلاسی است نمایانگر یک رابطه جهت‌دار بین نمونه‌های گزاره معنایی.

چ-۴-۱۲ کلاس مرجع خارجی تک زبانه

کلاس مرجع خارجی تک زبانه، کلاسی است نمایانگر رابطه بین یک نمونه مفهوم و یا دسته نحوی و یک سیستم خارجی.

یادآوری - دستورالعمل‌های این کلاس در پیوست درج شده است.

پیوست ح
(اطلاعاتی)
مثال‌های معنایی NLP

ح-۱-۱ مثال آرایش کلاس

کلاس‌ها ممکن است با ویژگی‌های زیر آرایش شوند:

تفسیر	مثالی برای ویژگی‌ها	نام کلاس
	تاریخ‌گذاری سبک فراوانی ^۱ جاننداری ^۲	مفهوم
کلاس رابطه مفهوم، کلاس چند منظوره‌ای است که برای نمایش تضادی و رابطه گروهی/اختصاصی و یا بخشی از رابطه، به کار می‌رود.	برچسب	رابطه مفهومی
برای نمونه، فرهنگ لغت در زبان بامبارا ^۳ (Bamanankan, bam) شامل مثال‌هایی با املاهای صحیح رایج آن زبان و هم چنین مثال‌هایی به همراه تلفظ برای مبتدی‌ها جهت درک و تلفظ صحیح مثال، می‌باشد.	متن منبع زبان	مثال مفهومی
	برچسب تعریف	گزاره معنایی
برای نمونه، یک اشتقاق معنایی بین مفهوم یک اسم و مفهوم یک فعل می‌تواند به یک گزاره مشترک متصل باشد. در این شرایط، بازنمایی گزاره‌ای مفهوم اسم به صورت /verbNominalization/ شناسایی می‌شود.	نوع تفسیر	نمایش گزاره‌ای
	نقش معنایی محدودیت	آرگمان معنایی
		رابطه آرگمانی
	برچسب	نوع معنایی
		نگاشت آرگمان معنایی نحوی
		تطابق معنایی و نحوی
	برچسب نوع	رابطه گزاره‌ای
	برچسب منبع	دسته نحوی

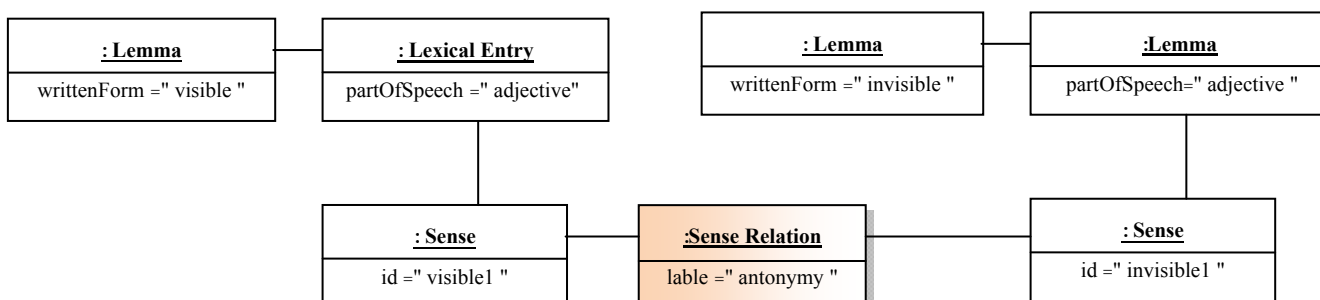
1-Frequency
2-Animacy
3-Bambara

نام کلاس	مثالی برای ویژگی‌ها	تفسیر
رابطه دسته نحوی	برچسب نوع	
مرجع خارجی تک زبانه	سیستم خارجی مرجع خارجی	هدف پیوست معنایی، ایجاد یک سیستم سازمانی دانش مرکب و پیچیده که می‌بایستی به عنوان یک یا چند سیستم خارجی طراحی شده برای آن هدف ساختاردهی شوند، نمی‌باشد. برخلاف این <i>/externalSystem/</i> و <i>/externalReference/</i> به ترتیب برای رجوع به نام‌های سیستم خارجی و هم چنین برای رجوع به گره مربوط این سیستم خارجی مفروض ایجاد شده‌اند.

ح-۱-۲ مثال‌های شرح واژه قاموسی

ح-۱-۲-۱ یک مثال ساده

مثال انگلیسی زیر دو صفت را ارائه می‌دهد: *visible* و *invisible* که به منظور شرح بیشتر، جزو مدخل‌های واژگانی تک‌مفهومی^۱ در نظر گرفته می‌شوند. این دو واژه قاموسی در سطح معنایی به وسیله نمونه ارتباط معنایی به یکدیگر ارتباط داده شده‌اند تا نشان دهند که *visible* متضاد *invisible* می‌باشد، به شکل ح-۱^۲ رجوع شود.



شکل ح-۱- نمودار نمونه برای یک مثال ساده

1-Monosemous

۲- برای این که فهم این شکل آسان‌تر شود، محتوی کادرهای سایه‌دار برای نمونه‌های مربوط به مقوله‌هایی که در بسته کنونی تعریف شده‌اند، مورد استفاده قرار می‌گیرند. محتوی کادرهای مربوط به نمونه‌های مقوله‌هایی که در بسته دیگر تعریف شده، سایه ندارند.

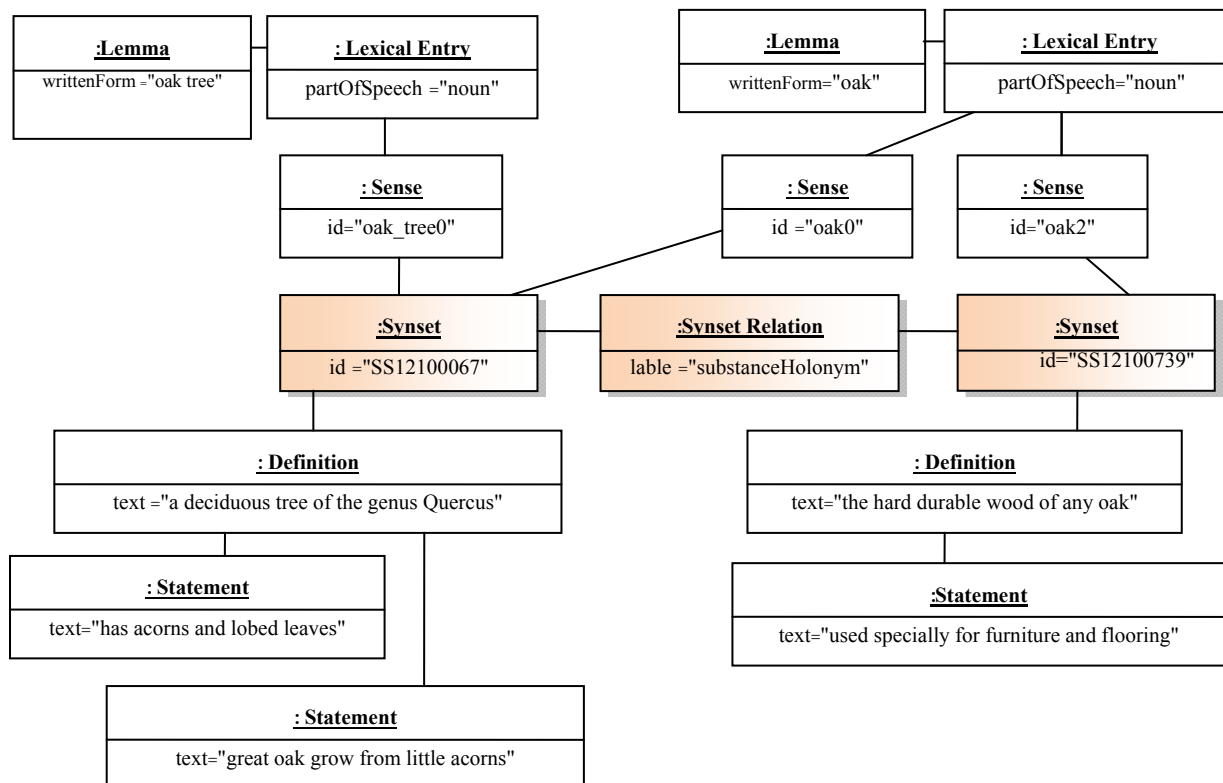
همان داده‌ها در قطعه XML به شرح زیر بیان شده است:

```
<LexicalEntry>
  <feat att="partOfSpeech" val="adjective"/>
  <Lemma>
    <feat att="writtenForm" val="visible"/>
  </Lemma>
  <Sense id="visible1">
    <SenseRelation targets="invisible1">
      <feat att="label" val="antonym"/>
    </SenseRelation>
  </Sense>
</LexicalEntry>
<LexicalEntry>
  <feat att="partOfSpeech" val="adjective"/>
  <Lemma>
    <feat att="writtenForm" val="invisible"/>
  </Lemma>
  <Sense id="invisible1"/>
</LexicalEntry>
```

شایان ذکر است که هیچ نیازی به برچسب XML در مسیر بازگشت از "invisible1" به "visible1" نمی‌باشد زیرا، این اطلاعات قبلاً از طرف "visible1" فراهم شده‌اند.

ح-۱-۲-۲ مثالی از نرم‌افزار Princeton WordNet نسخه 2.1

مثال انگلیسی زیر، بیشتر روی نمونه‌های دسته نحوی تمرکز نموده است. این مثال از نرم‌افزار WordNet نسخه 2.1 گرفته شده است و دو نمونه دسته نحوی را ارائه می‌دهد: یکی برای *oak* به عنوان یک درخت و دیگری برای *oak* به عنوان چوب درخت. هر شناسه واژگانی در نرم‌افزار WordNet برای شناسایی یک نمونه مفهوم به کار می‌رود. هر دسته نحوی شامل یک شرح وصفی چند قسمتی می‌باشد. قسمت اول همیشه یک تعریف و قسمت‌های دیگر گزاره‌هایی غالباً با محتوایی ناهمگن هستند. برای مثال، دسته نحوی "12100067" برای *oak* تعریف "درخت برگ‌ریز که یکی از انواع درخت بلوط می‌باشد" ارائه شده است. بخش دوم یک شرح روایتی از ویژگی‌های *oak* می‌باشد: "دارای میوه‌های بلوط بوده و برگ‌هایش را می‌ریزد". بخش سوم یک ضرب‌المثل می‌باشد: "درختان بلوط بزرگ از میوه‌های کوچک درختان بلوط به وجود می‌آیند". در این استاندارد بخش اول یک نمونه تعریف و دو بخش دیگر، دو نمونه گزاره می‌باشند. دو نمونه دسته نحوی توسط نمونه ارتباط دسته نحوی که توسط *substanceHolonym* نشانه‌گذاری می‌شوند، به یکدیگر پیوند داده شده‌اند، به شکل ح-۲ رجوع شود.



شکل ح-۲- یک نمونه نمودار برای مثال گرفته شده از نرم افزار WordNet

همان داده‌ها در قطعه XML به شرح زیر بیان شده است:

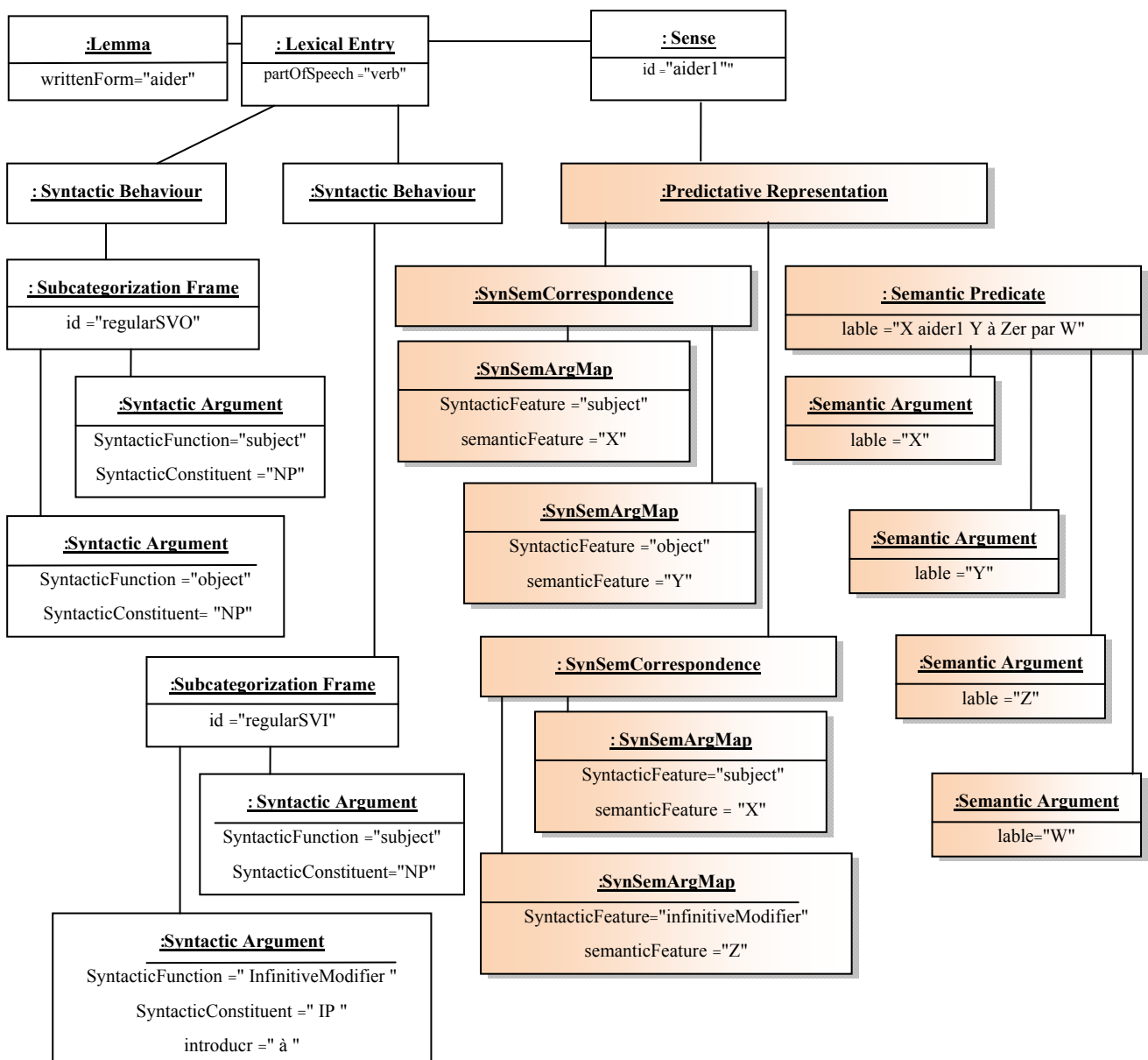
```
<LexicalEntry>
  <feat att="partOfSpeech" val="noun"/>
  <Lemma>
    <feat att="writtenForm" val="oak tree"/>
  </Lemma>
  <Sense id="oak_tree0" synset="SS12100067"/>
</LexicalEntry>
<LexicalEntry>
  <feat att="partOfSpeech" val="noun"/>
  <Lemma>
    <feat att="writtenForm" val="oak"/>
  </Lemma>
  <Sense id="oak0" synset="SS12100067"/>
  <Sense id="oak2" synset="SS12100739"/>
</LexicalEntry>
<Synset id="12100067">
  <Definition>
    <feat att="text" val="a deciduous tree of the genus Quercus"/>
    <Statement>
      <feat att="text" val="has acorns and lobed leaves"/>
    </Statement>
    <Statement>
      <feat att="text" val="great oaks grow from little acorns"/>
    </Statement>
  </Definition>
  <SynsetRelation targets="SS12100739"
    <feat att="label" val="substanceHolonym"/>
  </SynsetRelation>
</Synset>
<Synset id="SS12100739">
  <Definition>
    <feat att="text" val="the hard durable wood of any oak"/>
    <Statement>
      <feat att="text" val="used especially for furniture and flooring"/>
    </Statement>
  </Definition>
</Synset>
```

ح-۱-۲-۳ مثالی از فرهنگ لغت^۱ (DEC)

مثال فرانسوی زیر روی نمونه‌های گزاره معنایی و ارتباط بین بازنمایی‌های معنایی و نحوی تمرکز می‌کند. این مثال، نحو مفهوم *Aider1* گرفته شده از *Dictionnaire Explicatif et Combinatoire* را ارائه می‌دهد. "Aider1" به آرگمان معنایی مرتبط شده است: "X aide Y à Z-er par W" که در "il vous aidera par Z" به آرگمان معنایی مرتبط شده است: "son intervention à surmonter cette épreuve" (در معنای: "X به Y کمک می‌کند تا Y توسط W به Z برسد" یعنی "او به شما کمک می‌کند که با پادرمیانی او شما در آزمون موفق شوید") می‌باشد. این نمونه

1- Dictionnaire Explicatif et Combinatoire

مدخل واژگانی، ۸ نمونه چارچوب زیر رده‌بندی را ایجاد می‌کند، اما شکل زیر فقط دو مثال اول را برای شما ارائه می‌دهد: "La Grande-Bretagne aide ses voisins" ("بریتانیای کبیر به همسایگان خود کمک می‌کند") و "La Grande Bretagne a aidé à créer l'ONU" ("بریتانیای کبیر در شکل‌گیری ایالات متحده کمک می‌کند")، و هم چنین روی پیوندهای بین بازنمایی‌های نحوی و معنایی تمرکز می‌کند. نمونه‌های چارچوب زیر رده‌بندی به گزاره معنایی متداولی که آرگمان‌های معنایی آن دارای (X, Y, Z و W) می‌باشد مرتبط هستند. این نمونه‌ها برای ارتباط با نمونه‌های آرگمان نحوی بخصوصی در ساختارهای متفاوت فعل، نشان داده می‌شوند. به همین دلیل نمونه‌های چارچوب زیر رده‌بندی مستقیماً با گزاره مرتبط نیستند، اما یک آرگمان نحوی مخصوصی در هر نمونه چارچوب زیر رده‌بندی به نمونه آرگمان معنایی ویژه‌ای پیوند داده شده‌اند، به شکل ح-۳ رجوع شود.



شکل ح-۲- نمودار نمونه برای مثال برگرفته از فرهنگ لغت DEC

همان داده‌ها در قطعه XML به شرح زیر بیان شده است:

```
<LexicalEntry>
  <feat att="partOfSpeech" val="verb"/>
  <Lemma>
    <feat att="writtenForm" val="aider"/>
  </Lemma>
  <Sense id="aider1">
    <PredicativeRepresentation predicate="P1" correspondences="SVO_XY SVI_XZ">
    </PredicativeRepresentation>
  </Sense>
  <SyntacticBehaviour subcategorizationFrames="regularSVO"/>
  <SyntacticBehaviour subcategorizationFrames="regularSVI"/>
</LexicalEntry>
<SynSemCorrespondence id="SVO_XY">
  <SynSemArgMap>
    <feat att="syntacticFeature" val="subject"/>
    <feat att="semanticFeature" val="X"/>
  </SynSemArgMap>
  <SynSemArgMap>
    <feat att="syntacticFeature" val="object"/>
    <feat att="semanticFeature" val="Y"/>
  </SynSemArgMap>
</SynSemCorrespondence>
<SynSemCorrespondence id="SVI_XZ">
  <SynSemArgMap>
    <feat att="syntacticFeature" val="subject"/>
    <feat att="semanticFeature" val="X"/>
  </SynSemArgMap>
  <SynSemArgMap>
    <feat att="syntacticFeature" val=" infinitiveModifier "/>
    <feat att="semanticFeature" val="Z"/>
  </SynSemArgMap>
</SynSemCorrespondence>
<SubcategorizationFrame id="regularSVO">
  <SyntacticArgument>
    <feat att="syntacticFunction" val="subject"/>
    <feat att="syntacticConstituent" val="NP"/>
  </SyntacticArgument>
  < SyntacticArgument>
    <feat att="syntacticFunction" val="object"/>
    <feat att="syntacticConstituent" val="NP"/>
  </SyntacticArgument>
</SubcategorizationFrame>
<SubcategorizationFrame id="regularSVI">
  <SyntacticArgument>
    <feat att="syntacticFunction" val="subject"/>
    <feat att="syntacticConstituent" val="NP"/>
  </SyntacticArgument>
  <SyntacticArgument>
    <feat att="syntacticFunction" val="infinitiveModifier"/>
    <feat att="syntacticConstituent" val="IP"/>
    <feat att="introducer" val="à"/>
  </SyntacticArgument>
</SubcategorizationFrame>
```

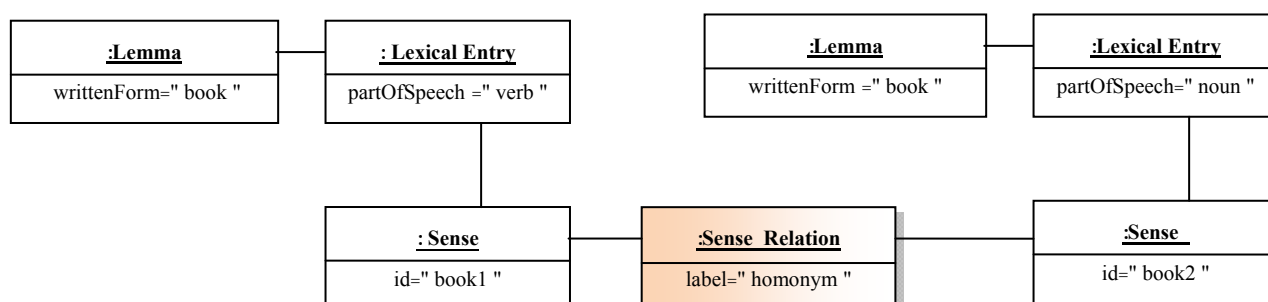
```

</SyntacticArgument>
</SubcategorizationFrame>
<SemanticPredicate id="P1">
  <feat att="label" val="X aider1 Y à Z par W"/>
  <SemanticArgument>
    <feat att="label" val="X"/>
  </SemanticArgument>
  <SemanticArgument>
    <feat att="label" val="Y"/>
  </SemanticArgument>
  <SemanticArgument>
    <feat att="label" val="Z"/>
  </SemanticArgument>
  <SemanticArgument>
    <feat att="label" val="W"/>
  </SemanticArgument>
</SemanticPredicate>

```

ح-۱-۲-۴ مثال هم‌آوایی

مثال زیر دو مدخل واژگانی مختلفی را با حالت نوشتاری یکسان ارائه می‌دهد. نمونه‌های مدخل واژگانی بدین صورت ایجاد شده‌اند: یکی برای فعل *to book* و دیگری برای اسم *book*. هر کدام از آن‌ها به نمونه بن‌واژه‌های متفاوتی مرتبط شده‌اند. هر نمونه مدخل واژگانی توسط شرح‌دهندگان مرتبط با صرف واژه‌های قاموسی کامل شده‌اند و این شرح‌دهندگان هر کدام کاملاً متفاوت‌تر از دیگری هستند. نمونه‌های مدخل واژگانی در سطح معنایی به وسیله نمونه ارتباط معنایی که برچسب *homonym* دارند به یکدیگر متصل شده‌اند.



شکل ح-۴ - مثال هم‌آوایی

مفهوم هم‌آوایی به عنوان یک رابطه نشان داده شده است، نه به عنوان یک ساختار. به عبارت دیگر، نمونه بن‌واژه بین مداخل واژگانی از لحاظ مفهومی مشترک نیست. این بدان معنی نیست که یک اجرای فیزیکی رشته‌های حروف موجود در پایگاه داده را لزوماً دو برابر می‌کند، اما این سیستم به عنوان یک سیستم فشرده‌سازی داده در نظر گرفته می‌شود، نه به عنوان یک سیستم مفهومی. شایان ذکر است معیاری که تصمیم‌گیری در مورد داشتن یک یا دو واژه قاموسی را کنترل می‌کند، در این استاندارد مشخص نشده است: تصمیم‌گیری برای جداکردن یا به اشتراک گذاشتن اطلاعات واژگانی به مدیریت فرهنگ لغت بستگی دارد.

پیوست خ

(الزامی)

بسط نشانه‌گذاری‌های چند زبانه NLP

خ-۱ اهداف

هدف این پیوست شرح بازنمایی معادل‌هایی برای نمونه‌های مفهوم و یا عملکرد نحوی بین یا درمیان دو یا چند زبان می‌باشد.

خ-۲ نبودن و یا بودن نشانه‌گذاری‌های چند زبانه در یک فرهنگ لغت

مدل چند زبانه می‌تواند برای پایگاه داده‌های واژگانی که دو یا چند زبان را شرح می‌دهند (یعنی مراجع دو زبانه در مقایسه با مراجع چند زبانه) استفاده شود. در یک پایگاه داده تک‌زبانه نیازی به استفاده از نشانه‌گذاری‌های چند زبانه نمی‌باشد.

خ-۳ نمودار کلاس

بسط نشانه‌گذاری‌های چند زبانه مطابق با شکل خ-۱ سازمان‌دهی شده‌اند.

خ-۴ گزینش‌ها

خ-۴-۱ عمومی

ساده‌ترین پیکره‌بندی فرهنگ لغت دو زبانه‌ای است که در آن فقط یک پیوند مستقل برای بازنمایی معادل یک مفهوم مفروض از یک زبان به زبان دیگر استفاده می‌شود، اما اجرای واقعی همان طوری که در بندهای خ-۴-۲ تا خ-۴-۶ آورده شده است، دست کم در بیش از پنج پیکره‌بندی پیچیده آشکار می‌شود.

خ-۴-۲ تنوع‌سازی^۱ و خنثی‌سازی^۲

در شرایط معینی، نگاشت ساده یک‌به‌یک بین معادل‌های آشکاری در دو یا چند زبان به خوبی عمل نمی‌کند، زیرا دامنه‌مفهومی^۳ ارائه شده توسط واژه‌های قاموسی و عبارات در زبان‌های مختلف غالباً یکسان نیستند. مثال - در زبان آلمانی کلمه *Lack* دامنه گسترده‌ای از مفهوم‌های کلی بیان شده در زبان انگلیسی را با واژه‌های قاموسی بخصوصی پوشش می‌دهد: *shellac* ، *varnish* ، *lacquer* ، *paint* و غیره. در این مورد، جهت‌نمایی از زبان آلمانی به انگلیسی مستلزم تنوع‌سازی می‌باشد و از زبان انگلیسی به آلمانی مستلزم یک نوع خنثی‌سازی می‌باشد.

خ-۴-۳ تعداد پیوندها

اگر چه شیوه هم‌ارزی یک‌به‌یک برای هر دو زبان عملی می‌باشد اما برای تعداد وسیعی از زبان‌ها این شیوه توجیه‌ناپذیر است، چرا که تعداد پیوندها افزایش یافته و غیرقابل مدیریت خواهند بود. علاوه بر این، لزوماً نمی‌توان فرض کرد که اگر تحت شرایط خاصی مفهوم $L_1-A=L_2-A$ و $L_1-A=L_3-A$ برقرار باشد مفهوم $L_2-A=L_3-A$ برقرار است. علاوه بر این که منطقی معمول بر این امر اشاره دارد. هرچه تعداد زبان‌ها و تعداد پیوندها بیشتر می‌شوند، شانس این که پیوندهای جانبی بین زبان‌های مختلف دارای اشتباهات بیشتری شوند، نیز افزایش می‌یابد.

1-Diversification
2-Neutralization
3-Conceptual scope

خ-۴-۴ انتقال یا محور دو زبانه^۱

ترجمه NLP محور بر مبنای دو رویکرد است: استفاده از یک محور دو زبانه که براساس تحلیل معنایی عمل می‌کند، و انتقال، که براساس تجزیه^۲ کامپیوتری نحو متن مبدا بنا نهاده شده است. رویکرد محوری از طریق کلاس محورمفهوم^۳ و رویکرد انتقال از طریق کلاس محورانتقال^۴ اجرا می‌شوند.

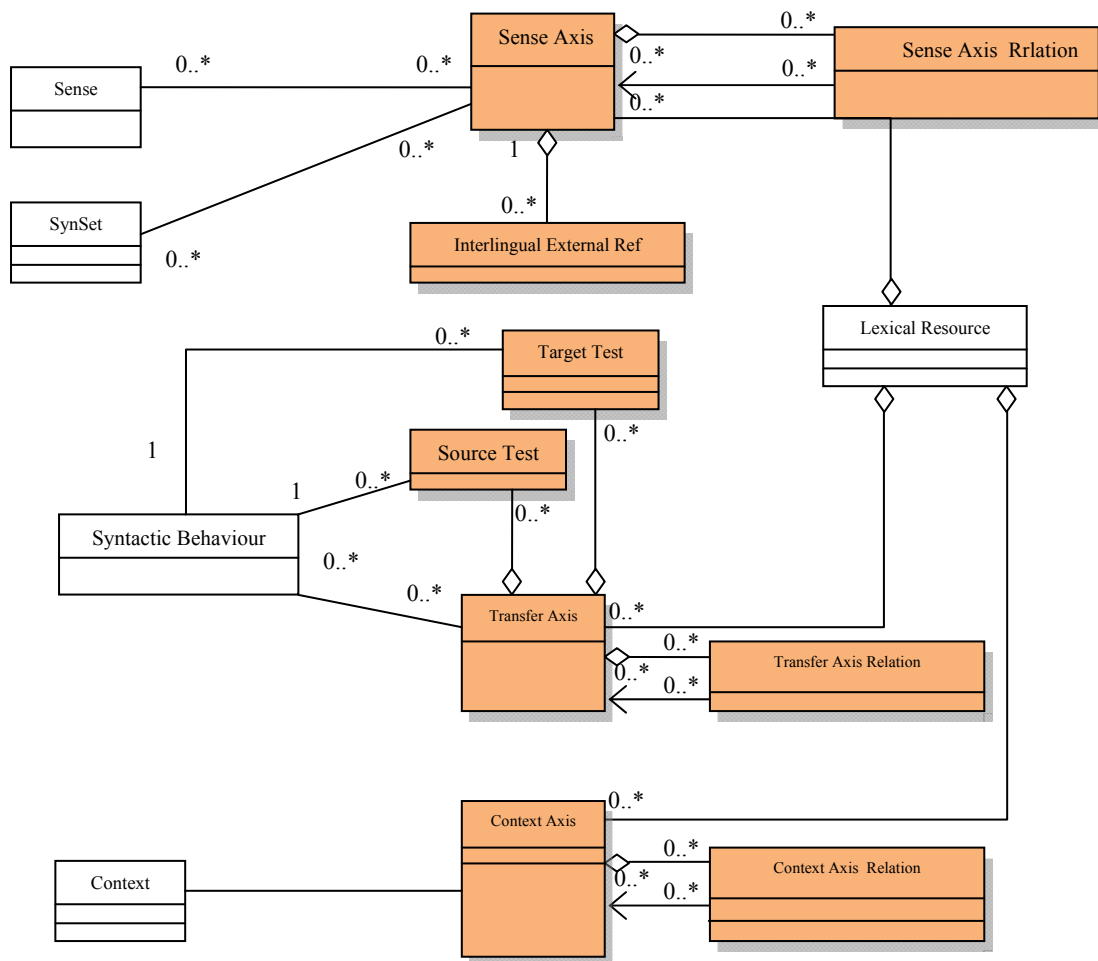
خ-۴-۵ بازنمایی زبان‌های مشابه

زبان‌هایی که به یکدیگر خیلی نزدیک بوده و در الگوهای مهم و اساسی مشترک می‌باشند، به طور مناسبی با استفاده از نمونه‌های محورمفهوم مشترک (به ترتیب نمونه‌های محور انتقال) و همچنین به همراه تعداد محدودی از نمونه‌های محور مفهوم غیرمشترک (به ترتیب نمونه‌های محور انتقال) برای بازنمایی تفاوت‌های موجود بین زبان‌ها، قابل ارائه هستند.

خ-۴-۶ دستورالعمل و آزمون‌ها

برخی فرهنگ لغات چند زبانه کاملاً به شکل اخباری^۵ هستند به طوری که هر ترجمه آن با هدف میان‌زبانی ارائه می‌شوند. در عوض، برخی دیگر کاملاً اسلوب‌گرا^۶ می‌باشند، به طوری که در آن فرهنگ لغت، هر ترجمه به آزمون‌های منطقی که در سطح‌های زبانی مبدا و یا مقصد اعمال می‌شوند محدود می‌شود.

1-Interlingual pivot
2-Parsing
3-SenseAxis class
4-TransferAxis class
5- Declarative
6-Procedural



شکل خ-۱- مدل نشانه گذاری های چند زبانه

خ-۵ شرح مدل نشانه گذاری های چند زبانه

این مدل بر اساس مفهوم محورها پایه ریزی شده است که نمونه های مفهوم، عملکرد نحوی و مثال مفهوم^۱ وابسته به چند زبان مختلف را به یکدیگر مرتبط می سازد. طراحان فرهنگ لغت می توانند چندین محور مختلف را به طور مستقل، مستقیم و یا غیرمستقیم، در میان و یا بین چندین زبان مختلف ساختاردهی نمایند. یک ارتباط مستقیم توسط یک محور مستقل اجرا می شود. یک ارتباط غیرمستقیم با استفاده از محورهای مختلف و توسط یک یا چندین رابطه به اجرا در می آید. این الگو بر پایه سه کلاس اصلی بنا نهاده شده است: محور مفهوم، محور انتقال و محور متن.

خ-۵-۱ کلاس محور مفهوم

کلاس محور مفهوم، کلاسی است که رابطه بین چندین مفهوم مختلف به هم مرتبط را در زبان های مختلف نشان می دهد و دیدگاهی را بر پایه محور میان زبانی به اجرا در می آورد. هدف، شرح ترجمه واژه های قاموسی

¹Sense Example

از یک زبان به زبان دیگر می‌باشد. یک محور مفهومی می‌تواند به طور انتخابی، به یک سیستم نمایش دانش خارجی که در آن معادل مناسبی را می‌تواند پیدا کند، رجوع کند.

خ-۵-۲ کلاس ارتباط محور مفهوم

کلاس ارتباط محور مفهوم، کلاسی است نمایانگر ارتباط بین دو نمونه مختلف محور مفهوم.

خ-۵-۳ کلاس مرجع خارجی دو زبانه

کلاس مرجع مرجع خارجی دو زبانه، کلاسی است نمایانگر ارتباط بین یک نمونه محور مفهومی و یک سیستم خارجی.

خ-۵-۴ کلاس محور انتقال

کلاس محور انتقال، کلاسی است نمایانگر انتقال چند زبانی. یک نمونه محور انتقالی، نمونه‌های عملکرد نحوی مختلف وابسته به چندین زبان مختلف را به یکدیگر مرتبط می‌سازد.

خ-۵-۵ کلاس ارتباط محور انتقال

کلاس ارتباط محور انتقال، کلاسی است نمایانگر ارتباط بین دو نمونه محور انتقالی.

خ-۵-۶ کلاس آزمون مبدا^۱

کلاس آزمون مبدا، کلاسی است نمایانگر شرایطی که ترجمه را در رابطه با کاربرد آن در زبان مبدا تحت تاثیر قرار می‌دهد.

خ-۵-۷ کلاس آزمون مقصد

کلاس آزمون مقصد، کلاسی است نمایانگر شرایطی که ترجمه را در رابطه با کاربرد آن در زبان مقصد تحت تاثیر قرار می‌دهد.

خ-۵-۸ کلاس محور متن^۲

کلاس محور متن، کلاسی است نمایانگر مثال‌های ترجمه‌ای که قبلاً ترجمه شده‌اند، به طوری که این مثال‌ها دارای نظایر مشابهی در قطعه متن مفروض می‌باشند.

خ-۵-۹ کلاس ارتباط محور متن

کلاس ارتباط محور متن، کلاسی است نمایانگر ارتباط بین دو نمونه محور متن.

1-Source Test class
2-Context Axis class

پیوست د
(اطلاعاتی)

مثال‌های نشانه‌گذاری‌های چند زبانه NLP

د-۱ مثال آرایش کلاس

کلاس‌ها ممکن است با ویژگی‌های زیر آرایش شوند:

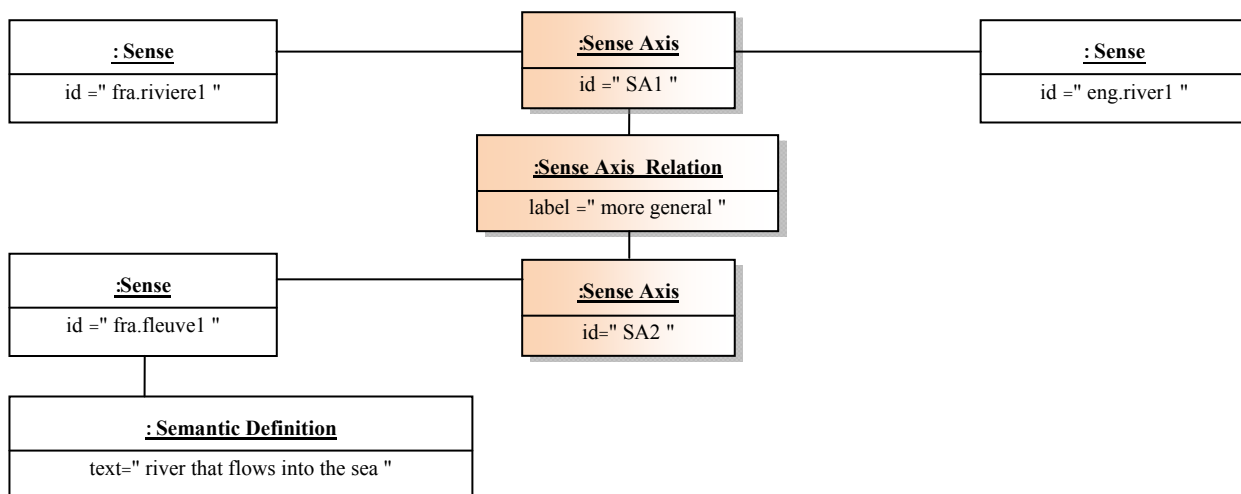
نام کلاس	مثالی برای ویژگی‌ها	تفسیر
محور مفهوم	برچسب شناسه	شایان ذکر است که هیچ محدودیتی در نوع واژه‌های قاموسی که توسط نمونه محور مفهوم به یکدیگر مرتبط می‌شوند، وجود ندارد. برای نمونه، یک واژه قاموسی در یک زبان می‌تواند در زبانی دیگر دارای معادل چند کلمه‌ای باشد.
ارتباط محور مفهوم	برچسب دیدگاه	برچسب قادر به کدگذاری روابط دو زبانه ساده می‌باشد، همانند کلمه تخصصی <i>flueve</i> در مقایسه با کلمات <i>rivière</i> و <i>river</i> . با این وجود، کدگذاری سیستم سازمانی دانش پیچیده و مرکب در حیطه اهداف این شیوه نمی‌باشد.
مرجع خارجی دو زبانه	سیستم خارجی مرجع خارجی	فراهم ساختن یک سیستم سازمانی دانش پیچیده و مرکب که از نظر تئوری احتمالاً به عنوان یک یا چندین سیستم خارجی طراحی شده برای آن هدف ساختاردهی شده‌اند، جزو اهداف پیوست چند زبانه نمی‌باشد. لیکن، <i>/externalSystem/</i> و <i>/externalReference/</i> به ترتیب برای ارجاع به نام(های) سیستم خارجی و همچنین گره مربوطه بخصوص در این سیستم خارجی مفروض ایجاد شده‌اند.
محور انتقالی	برچسب شناسه	این دیدگاه بر ترجمه آرگمان‌های نحوی شامل برگردانی زبان دلالت دارد، مانند: در زبان فرانسه: "elle me manque" = در زبان انگلیسی: "I miss her". بنا به این واقعیت که مدخل واژگانی دارای یک حالت همچون فعل کمکی باشد، می‌توان معادل‌های را بین یک فعل ساده در یک زبان و یک فعلی با ساختار پیچیده در زبان دیگر را ارائه داد، همانند یک فعل کمکی و یا اجزای اضافی دیگری مثل یک رابطه معادلی بین زبان فرانسه: "Marie rêve" و زبان ژاپنی "Marie wa yume wo miru".
ارتباط محور انتقالی	برچسب تنوع	کلاس ارتباط محور انتقال می‌تواند جهت نمایش تفاوت‌های کوچک بین زبان‌های کاملاً مشابه به کار رود. برای نمونه، به منظور نمایش تفاوت‌های کوچک بین زبان‌های اروپایی و پرتغالی برزیلی، نمونه‌های محور انتقال میانی متفاوتی ایجاد می‌شود. کلاس ارتباط محور انتقال، برچسبی دارد که این برچسب با توجه به زبان مفروض برای تشخیص اینکه کدامیک از نمونه‌های محور انتقال را استفاده کند، کاربرد دارد.

نام کلاس	مثالی برای ویژگی‌ها	تفسیر
آزمون مبدا	متن توضیح	
آزمون مقصد	متن توضیح	
محور متن	توضیح منبع شناسه	هدف این کلاس ثبت تعداد کثیری از مجموعه چند زبانه نمی‌باشد، بلکه هدف ایجاد ارتباط بین یک نمونه مدخل‌واژگانی و یک نمونه ترجمه بخصوص می‌باشد.

د-۲ مثال‌های شرح واژه قاموسی

د-۲-۱ مثالی از هم‌ارزی در سطح مفهوم

مثال زیر چگونگی استفاده از دو نمونه محور مفهوم میانی، به منظور ایجاد یک معادل نزدیک بین کلمه *fleuve* در زبان فرانسوی و *river* در زبان انگلیسی را بررسی می‌کند. این پدیده عموماً تنوع‌سازی و خنثی‌سازی نامیده می‌شود. نمونه محور مفهوم در قسمت پایین مستقیماً به مفهوم انگلیسی وصل نشده است زیرا این نظریه در انگلیسی وجود ندارد، به شکل د-۱^۱ رجوع شود.



شکل د-۱- نمودار نمونه برای کلمه "river"

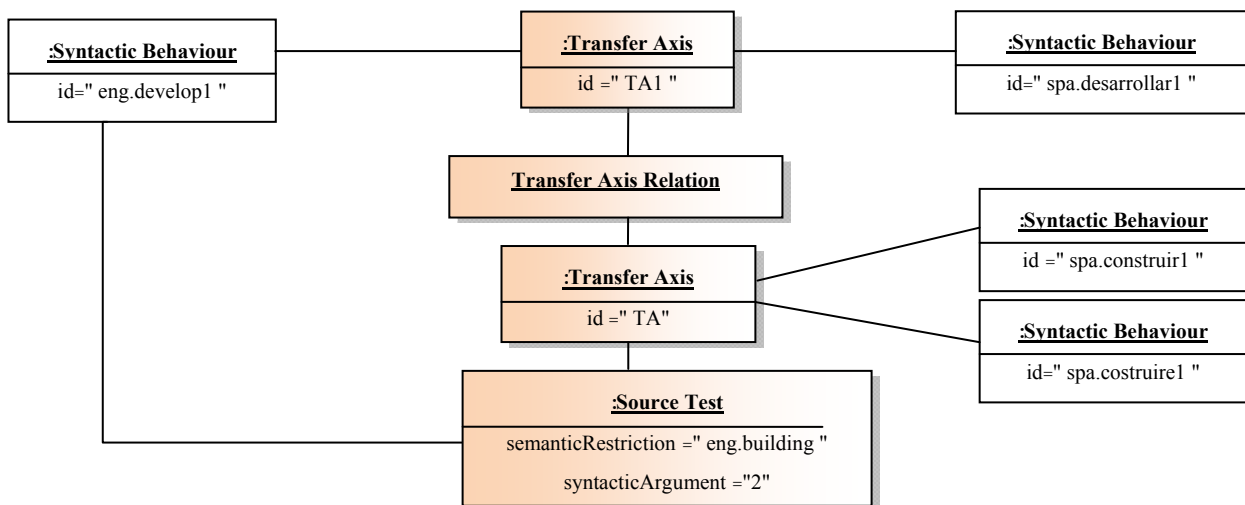
۱- برای این که فهم این شکل آسان‌تر شود، محتوی کادرهای سایه‌دار برای نمونه‌های مربوط به رده‌هایی که در بسته کنونی تعریف شده‌اند، مورد استفاده قرار می‌گیرند. محتوی کادرهای مربوط به نمونه‌های رده‌هایی که در بسته دیگر تعریف شده، سایه ندارند.

نمونه‌های الگوسازی شده در پیوست نشانه‌گذاری چند زبانه می‌توانند توسط قطعه کد XML زیر نیز بیان شوند، با این فرض که نمونه‌های مفهوم و تعریف معنایی در جای دیگری تعریف شده‌اند.

```
<SenseAxis id="SA1" senses="fra.riviere1 eng.river1">
  <SenseAxisRelation targets="SA2">
    <feat att="label" val="more general"/>
  </SenseAxisRelation>
</SenseAxis>
<SenseAxis id="SA1" senses="fra.fleuve1"/>
```

د-۲-۲ مثال از هم‌ارزی در سطح انتقال

این مثال چگونگی استفاده از نمونه ارتباط محور انتقال را برای پیوند اطلاعات مختلف در یک فرهنگ لغت انتقالی چند زبانه نشان می‌دهد. این مثال ترجمه کلمه انگلیسی *develop* را به ایتالیایی و اسپانیایی نشان می‌دهد. به یاد داشته باشید که یک مفهوم کلی‌تری کلمه *develop* را در انگلیسی به کلمه *desarrollar* در اسپانیایی مرتبط می‌سازد. هر دو زبان اسپانیایی و ایتالیایی محدودیت‌هایی دارند که می‌بایستی در زبان مبدا آزموده شوند. اگر آرگمان دوم نمونه چارچوب زیر رده‌بندی به اجزای معینی مثل *building* رجوع کند، این کلمه می‌بایستی به افعال بخصوصی ترجمه شود، به شکل د-۲ رجوع شود.



شکل د-۲-۲- نمودار نمونه برای کلمه "develop"

نمونه‌های الگوسازی شده در پیوست نشانه‌گذاری چند زبانه می‌توانند توسط قطعه کد XML زیر نیز بیان شوند، با این فرض که نمونه‌های مفهوم و تعریف معنایی در جای دیگری تعریف شده‌اند.

```
<TransferAxis id="TA1" syntacticBehaviours="eng.develop1 esp.desarrollar1">
  <TransferAxisRelation targets="TA2"/>
</TransferAxis>
<TransferAxis id="TA2" syntacticBehaviours="esp.construir1 ita.costruire1">
  <SourceTest>
    <feat att="semanticRestriction" att="eng.building"/>
    <feat att="syntacticArgument" att="2"/>
  </SourceTest>
</TransferAxis>
```

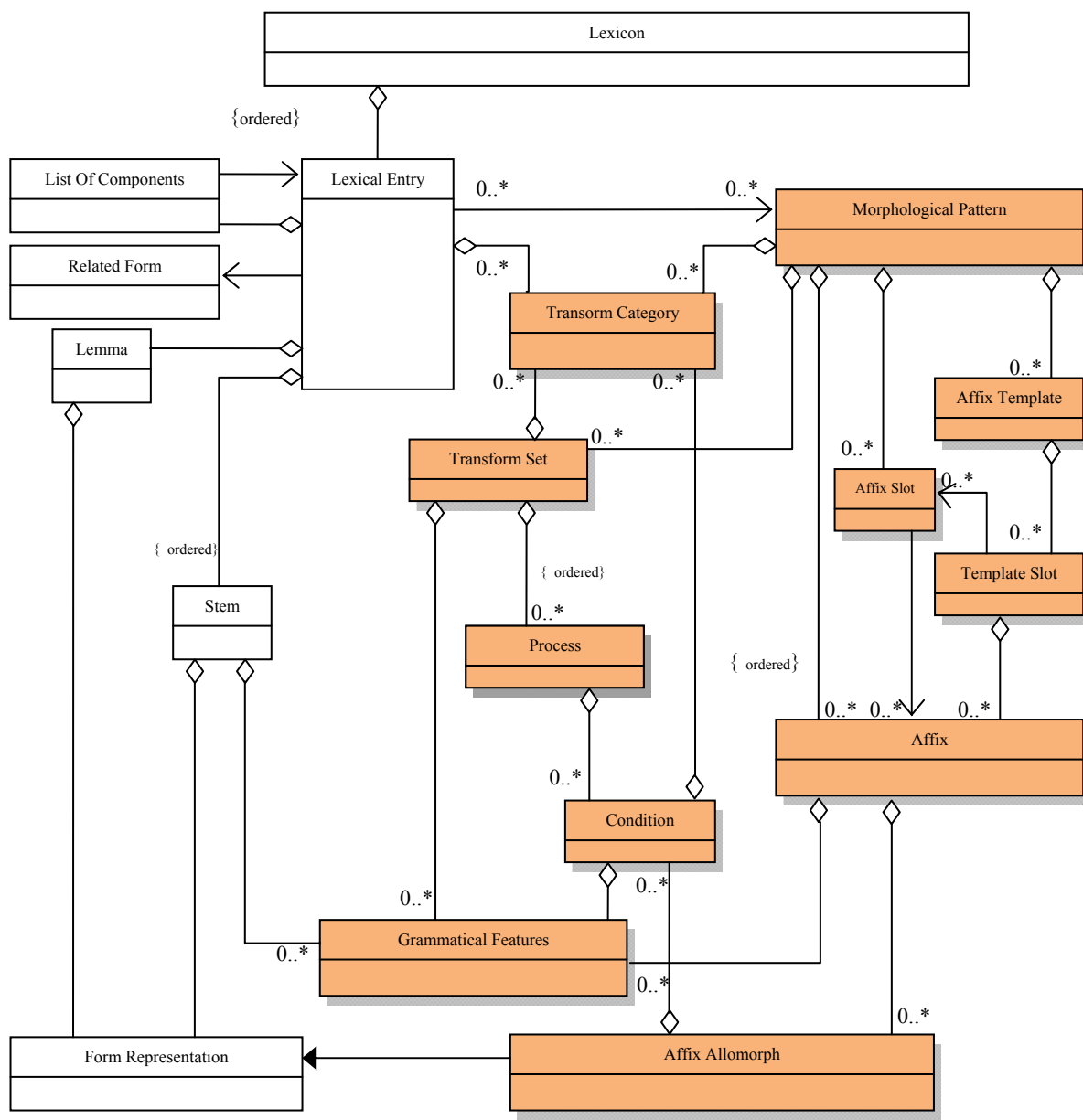
پیوست ذ
(الزامی)
بسط الگوهای صرفی NLP

ذ-۱ اهداف

هدف این بسط ارائه شرح در باره ساختواژه یک زبان مفروض می‌باشد. هدف، پشتیبانی سازماندهی و ذخیره اطلاعات واژگانی مورد نیاز برای تحلیل و تولید حالت‌های تصریفی، پیوندی، اشتقاقی و یا مرکب کلمات می‌باشد. این حالات به طور دقیق بیان نشده‌اند، اما می‌توان گفت که نمونه مدخل واژگانی به نمونه الگوی ساختواژه وابسته است. حالاتی که در مدخل واژگانی مدون شده‌اند ممکن است شامل ریشه، ستاک، و یا تک‌واژگونه‌های ستاک^۱ باشند. هم‌چنین این حالات برای هر مدخل واژگانی ویژه‌ای، منحصر به فرد می‌باشد. اطلاعات واژگانی مدون شده در ساختار الگوی صرفی ممکن است شامل حالات مشترک (برای مثال وندها) و قوانین مرتبط که منجر به پشتیبانی طرح فرهنگ لغات ساختواژی که دارای پردازش مستقل می‌باشند، باشد. یعنی این که الگوریتم‌ها برای تحلیل و تولید کلمات استفاده می‌شوند. این بسط به منظور برآورده کردن تمامی نیازهای فرهنگ لغات ساختواژی ایجاد نشده‌اند، با این حال، هسته مرکزی LMF به همراه این بسط، در واقع زیر بنایی برای گسترش بسط‌های ساختواژه دیگر فراهم می‌آورند.

1-Stem allomorphs

بسط الگوی ساختار همان طوری که در شکل ذ-۱ شرح داده شده است، سازمان دهی می شود.



شکل ذ-۱- نمونه الگوی صرفی

ذ-۳ شرح مدل الگوی ساختواژه

ذ-۳-۱ مقدمه

مدخل واژگانی و الگوی ساختواژه در یک تناظر گروهی با کلاس فرهنگ لغت قرار دارند. کلاس مدخل واژگانی، حالات مختلف کلمه و هم چنین واژهایی را که برای هر مدخل واژگانی منحصر به فرد می‌باشند، مدیریت می‌کند. برخلاف این موضوع، الگوی ساختواژه کلاس‌هایی را مدیریت می‌کند که یک شمای مشترکی را توسط مدخل واژگانی مختلف تشکیل می‌دهند.

ذ-۳-۲ کلاس الگوی ساختواژه

کلاس الگوی ساختواژه، کلاسی است نمایانگر ساختار وندها و/یا دسته‌ای از فرآیندهایی که این فرآیندها یک الگوی تغییرات ساختواژه‌ای را برای یک زبان مفروض شرح می‌دهند. الگوی ساختواژی، مدل‌سازی الگوهای تصریفی، اشتقاقی، پیوندی و مرکب را پشتیبانی می‌کند و بر طبق این الگوها انواع آن مشخص می‌شود، مثال: نوع الگو="inflectional". محدودیت‌های کلاس، انواع مرتبط و ویژگی‌های ساختواژه‌ای و انتخاب الگوهای خاص باعث می‌شوند تا کلاس الگوی ساختواژی قادر به توصیف انواع مختلف روش‌های به کار گرفته شده در گروه‌های زبانی گوناگون برای کدگذاری تغییرات ساختواژی باشد. برای مثال، الگوی ساختواژی ممکن است به منظور مدیریت الگوهای تصریفی برای زبان‌های هنداروپایی به اجزای کلام محدود شود و یا برای مدیریت فرآیند تکرار مورد استفاده قرار می‌گیرد که این فرآیند، تاکید و تشدید فعل را برای زبان تایلندی، بدون نیاز به محدودیت اجزای کلام الگوسازی می‌کند. یک الگوی ساختواژی می‌تواند یک کلاس وند را به طور غیر مستقیم از طریق کلاس جایگاه وند مدیریت کند یا می‌تواند کلاس وند را به طور مستقیم مدیریت کند، اما این دو عمل در یک مدل یکسان انجام نمی‌گیرد. الگوی^۱ ساختواژی هم چنین می‌تواند کلاس دسته تبدیل را که مدل‌سازی قوانین زبان‌شناسی و پردازش‌ها را پشتیبانی می‌کند، مدیریت کند و هم چنین می‌تواند کلاس قالب وند را که مدیریت الگوهای وند (برای مثال دسته‌های ترتیبی پسوندها و یا پیشوندها) را پشتیبانی می‌کند، مدیریت نماید. تناظرهای کلاس دسته تبدیل، پیوندهای وند و جایگاه وند و قالب وند متقابلاً مشخص نیستند. با وجود این، استفاده از کلاس قالب وند، محدودیت افزونی را برای استفاده از کلاس‌های وند و جایگاه وند ایجاد می‌کند.

ذ-۳-۳ کلاس دسته تبدیل

کلاس دسته تبدیل، کلاسی است نمایانگر تناظر بین کلاس پردازش و کلاس مشخصه‌های دستوری که دامنه الگوی مدیریت شده را تعریف می‌کند. دسته تبدیل در یک رابطه گروهی صفر به چند با الگوی ساختواژی قرار دارد.

ذ-۳-۴ کلاس پردازش

کلاس پردازش، کلاسی است نمایانگر قوانین و یا عملیات زبان‌شناسی به کار گرفته شده در یک حالت کلمه مانند وند یا ستاک یا هم چنین ترکیب حالات کلمات مثل وندها و ستاک‌ها. یک نمونه پردازش می‌تواند

گونه‌بندی شود مانند "phonological Operation" = process type و در یک ترتیب گروهی با کلاس دسته تبدیل قرار گیرد.

ذ-۳-۵ کلاس مشخصه‌های دستوری

کلاس مشخصه‌های دستوری، کلاسی است نمایانگر ترکیب نامرتب مشخصه‌های دستوری.

ذ-۳-۶ کلاس شرط

کلاس شرط، کلاسی است نمایانگر شرط‌هایی که کاربرد یک نمونه پردازش و یا نمونه تکواژگونه‌های وندی را تعیین و یا محدود می‌کند.

مثال - نمونه شرط، محیط‌های آواشناختی را شرح می‌دهد که طرز انتخاب یک وند را تعیین می‌کنند.

ذ-۳-۷ کلاس وند

کلاس وند، کلاسی است نمایانگر وندی که این وند یک حالت کلمه و یا تکواژی است که با دسته‌ای از ویژگی‌های دستوری شرح داده شده و برای تحلیل و تولید حالات کلمات الزامی می‌باشد. کلاس وند یک یا چند تکواژگونه وندی را از طریق پیوند گروهی با کلاس تکواژگونه‌های وندی مدیریت می‌کند.

ذ-۳-۸ کلاس قالب وند

کلاس قالب وند، کلاسی است که الگوی وندهای ترتیبی را برای ساختواژه تصریفی، اشتقاقی و یا پیوندی به طور غیر مستقیم از طریق کلاس جایگاه قالب مدیریت می‌کند. ویژگی‌های قالب وند، جهت‌داری^۱ وندها، شمار وندها را در یک دسته مرتب، و شروط قابل استفاده برای الگوی وندی (برای مثال، انتخاب جایگاه‌های ویژه) را شرح می‌دهد.

یادآوری - جهت محدودیت ترتیبی به زیرکلاس و زبان بستگی دارد. زبان‌های راست به چپ و چپ به راست ترتیب‌های متفاوتی دارند، هم چنین جایگاه‌های پسوندها و پیشوندها نیز دارای ترتیب‌های متفاوتی می‌باشند.

ذ-۳-۹ کلاس جایگاه قالب

کلاس جایگاه قالب، کلاسی است نمایانگر دسته‌ای از وندها که می‌توانند به موقعیت ترتیبی در کلاس قالب وند متصل شوند. جایگاه قالب می‌تواند کلاس وند را به طور غیرمستقیم از طریق کلاس جایگاه وند مدیریت نماید و یا می‌تواند کلاس وند را به طور مستقیم مدیریت کند، ولی این دو عمل به شیوه‌های یکسانی صورت نمی‌پذیرند. ویژگی‌های جایگاه وند در واقع نوع وند (مثل پسوند و وند ترکیبی)، ردیف وند را در یک دسته ترتیبی، شمار وندها را در یک دسته ترتیبی و هم چنین شروط قابل استفاده برای وند (عملکردهای ساختواژی مشترک با وندهایی که جایگاه را اشغال می‌کنند) را شرح می‌دهد.

ذ-۳-۱۰ کلاس جایگاه وند

کلاس جایگاه وند، کلاسی است که به دسته‌ای از وندهایی که به وضعیت یکسان مرتبط با یک ستاک از طریق کلاس جایگاه قالب متصل شده‌اند، اشاره دارد. دسته وندها زیر دسته وندهایی را که توسط الگوی ساختواژه به طور مستقیم و یا غیر مستقیم مدیریت می‌شوند را نمایش می‌دهد. یک وند توسط یک یا چند شی از کلاس جایگاه وند مورد اشاره قرار می‌گیرند.

یادآوری - مدیریت وندها از طریق یک پیوند جهت‌دار از کلاس جایگاه وند نیاز به ارائه نمونه شی‌های وند حشو را در زمان اجرای LMF در الگوهای داده‌ها، کاهش می‌دهد.

ذ-۳-۱۱ کلاس تکواژگونه وندی

کلاس تکواژگونه وندی، در واقع تعمیم کلاس نمایش حالتی است که تکواژگونه‌های حالات مرسوم وندی را در تمامی خط‌ها و نمایش‌ها نشان می‌دهد. یک تکواژگونه وندی با نمونه‌های کلاس شرط در ارتباط می‌باشد که این نمونه‌ها محیط آواشناختی و یا شرایط دیگری (مثل مرز تکواژ ستاک) را که موجب تولید تکواژگونه‌ها می‌شوند را شرح می‌دهند.

ذ-۳-۱۲ کلاس گروه تبدیل

کلاس گروه تبدیل، کلاسی است نمایانگر ویژگی‌هایی که این ویژگی‌ها مجموعه ویژگی‌های مورد نیاز برای مدیریت تغییرات ساختواژه را محدود کرده و یا شرح می‌دهند.
مثال - کلاس‌های صرفی در زبان اسپانیایی و یا گروه‌های ستاک در زبان عربی.

پیوست ر (اطلاعاتی)

مثال‌های الگوهای صرفی NLP

ر-۱ بودن و یا نبودن الگوهای ساختوازی در یک فرهنگ لغت

همه فرهنگ‌های لغات از رویکرد الگوی ساختوازه استفاده نمی‌کنند. از نظر تئوری، می‌توان همه حالات را در یک فرهنگ لغت فهرست‌بندی نمود، اما استفاده از الگوهای ساختوازه مزیت‌های مهم زیر را به همراه خواهد داشت:

- شرح زبان‌هایی با ساختوازه پیچیده بدون توسل^۱ به مستندسازی حجیم و غیرقابل مدیریت.
- دانش زبان‌شناسی که چگونگی پیوند یک بن‌وازه را با حالات تصریفی، پیوندی، ترکیب و اشتقاقی شرح می‌دهد، به جای پرداختن به تعداد گسترده‌ای از اجزای زبان، تمرکز روی جزء نمونه بخصوصی را ترجیح داده است.

ر-۲ مثال آرایش کلاس

کلاس‌ها ممکن است با ویژگی‌های زیر آرایش شوند:

نام گروه	مثالی برای ویژگی‌ها	توضیح
الگوی ساختوازه	شناسه تفسیر مثال اجزای کلام نوع الگو	نمونه الگوی ساختوازه برای به اشتراک گذاشته شدن و ارجاع طراحی شده است، بنابراین دارای یک شناسه می‌باشد. نمونه الگوی ساختوازه نمی‌تواند برای دو اجزای کلام مختلف (ذ-۳-۲) به کار گرفته شود، بنابراین ثبت علامت اجزای کلام بسیار مهم می‌باشد.
دسته تبدیل	تفسیر	این کلاس برای پیوند نمونه‌ها طراحی شده است، بدین ترتیب علاوه بر تفسیر این گروه نیاز به اطلاعات زبان‌شناسی دیگری ندارد.
پردازش	عملگر ردیف‌وند ردیف‌جزء ردیف‌ستاک قاعده ^۲ مقدار رشته	مقادیر /operator/ می‌توانند برای نمونه /addLemma/ ، /addAffix/ ، یا /addComponentStem/ باشند. مقادیر قواعد، مقادیر رشته‌ای می‌باشند که می‌توانند دامنه گسترده از قواعد زبان‌شناسی را برای نمونه یک الگو مانند /CVx/ و یا یک فرمولی مانند /v[ut]1 -> [X]n/ ارائه دهند.
شرط	شناسه جایگاه توافق وند نوع تبدیل	

1-Resorting
2-Rull

نام گروه	مثالی برای ویژگی‌ها	توضیح
وند	حالت نوشتاری نوع	نوع، به عنوان مثال می‌تواند با مقادیری مانند /prefix/ و یا /suffix/ مشخص شود.
قالب وند	نوع	نوع، به عنوان مثال می‌تواند با مقادیری مانند /prefix/ و یا /suffix/ مشخص شود.
جایگاه وند	شناسه	
جایگاه قالب	برچسب موقعیت الزامی	/position/ مکانی را که یک وند قرار است در یک حالت کلمه قرار بگیرد را تعیین می‌کند.
تکواژگونه‌های وندی	حالت نوشتاری	
گروه تبدیل	شناسه تفسیر	نمونه گروه تبدیل برای به اشتراک گذاشته شدن و ارجاع طراحی شده است، از این رو دارای یک شناسه می‌باشد.
مشخصه‌های دستوری	شمار دستوری جنسیت دستوری زمان دستوری شخص	

ر-۳ مثال‌های شرح واژه قاموسی

ر-۳-۱ مقدمه

این مدل گسترش اجرا را برای پشتیبانی اهداف مدل‌سازی مختلفی مانند، مدل بخش و ترتیب^۱، مدل بخش و پردازش^۲، رویکرد مبتنی بر بن‌واژه، و رویکرد مبتنی بر ستاک ممکن می‌سازد.

این مبحث به مثال‌های زیر مربوط می‌باشد:

- مثال‌های تصریفی، که با یک پدیده ساده‌ای شروع شده و با چندین پدیده پیچیده‌تری پایان می‌یابد،

- مثال‌های پیوندی،

- مثال‌های اشتقاقی و

- مثال‌های مرکب.

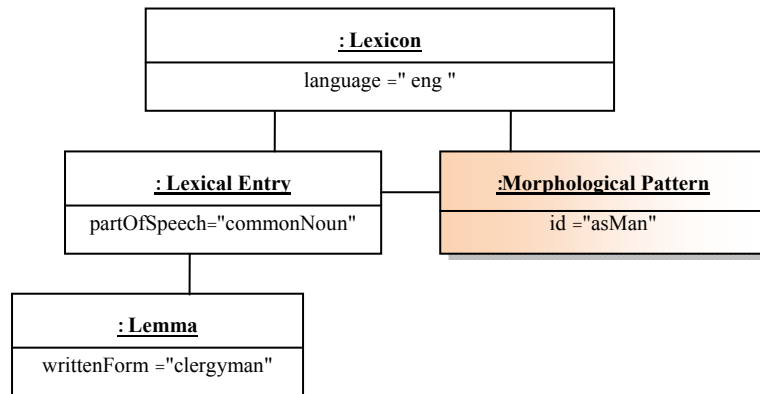
نمونه‌های ستاک، وند، مدخل واژگانی (در پیوند فهرست مولفه‌ها) و پردازش مرتب شده‌اند. در نمودارهای زیر ترتیب مد نظر نمی‌باشد. فرض بر این است که نمونه‌ها از چپ به راست و بالا به پایین خوانده می‌شوند.

ر-۳-۲ الگوی تصریفی با استفاده از نمونه الگوی واژه‌شناختی تحت تشخیص^۳

در مثال زیر، ادعا می‌شود که بن‌واژه /clergyman/ منطبق بر الگوی ساختواژه "asMan" می‌باشد. این نمونه الگوی ساختواژه، یک اسم دارد اما این اسم به طور تحلیلی در فرهنگ لغت ذکر نشده است، به شکل شماره

1-Item –and-Arrangement
2-Item-and-Process
3-Underspecified

ر-۱^۱ مراجعه شود. به عبارت دیگر، نمونه الگوی ساختواژه به عنوان یک نشانه در نظر گرفته شده است. برای مثال الگوی ساختواژه ممکن است به وسیله الگوریتم خارجی مبهم^۲ به اجرا درآید.

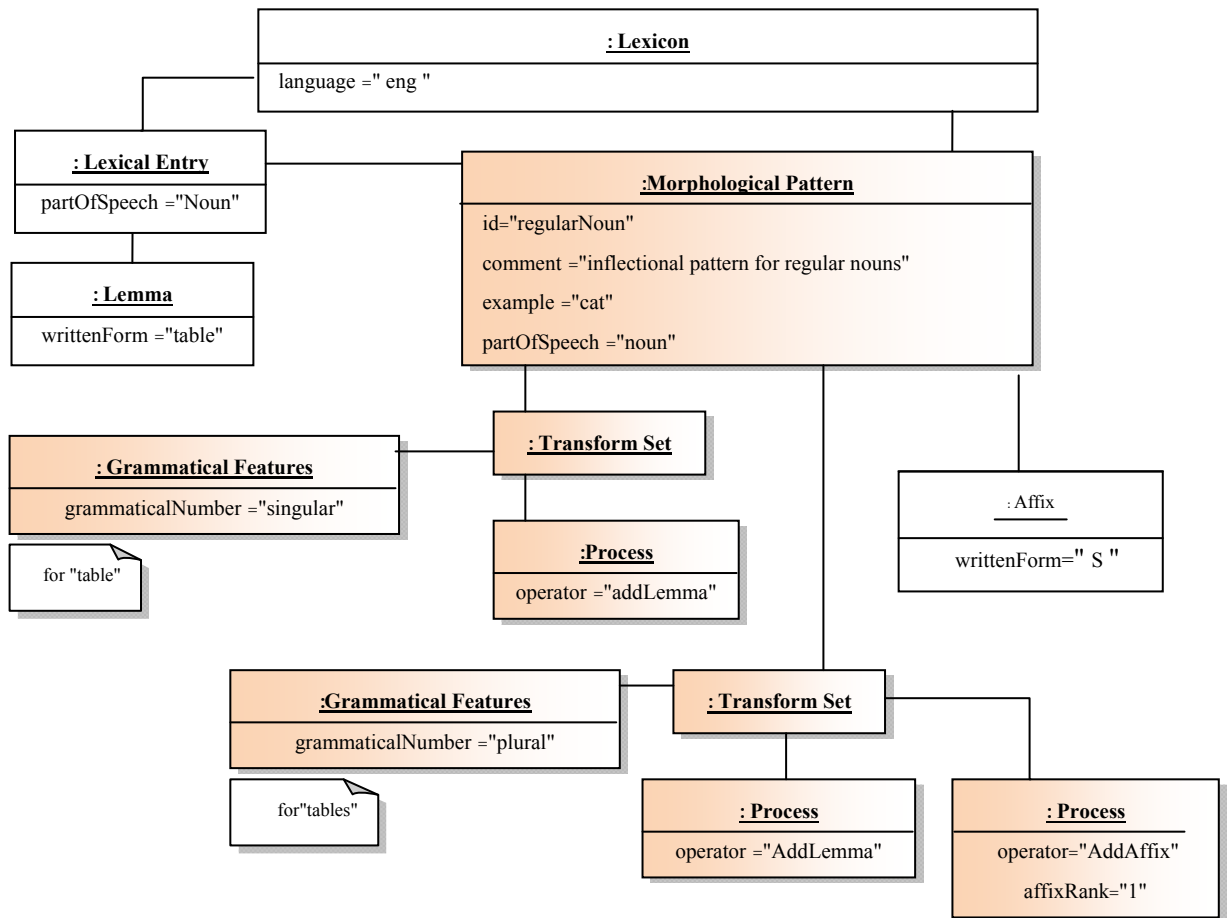


شکل ر-۱- الگوی تصریفی با استفاده از الگوی ساختواژه تحت تشخیص

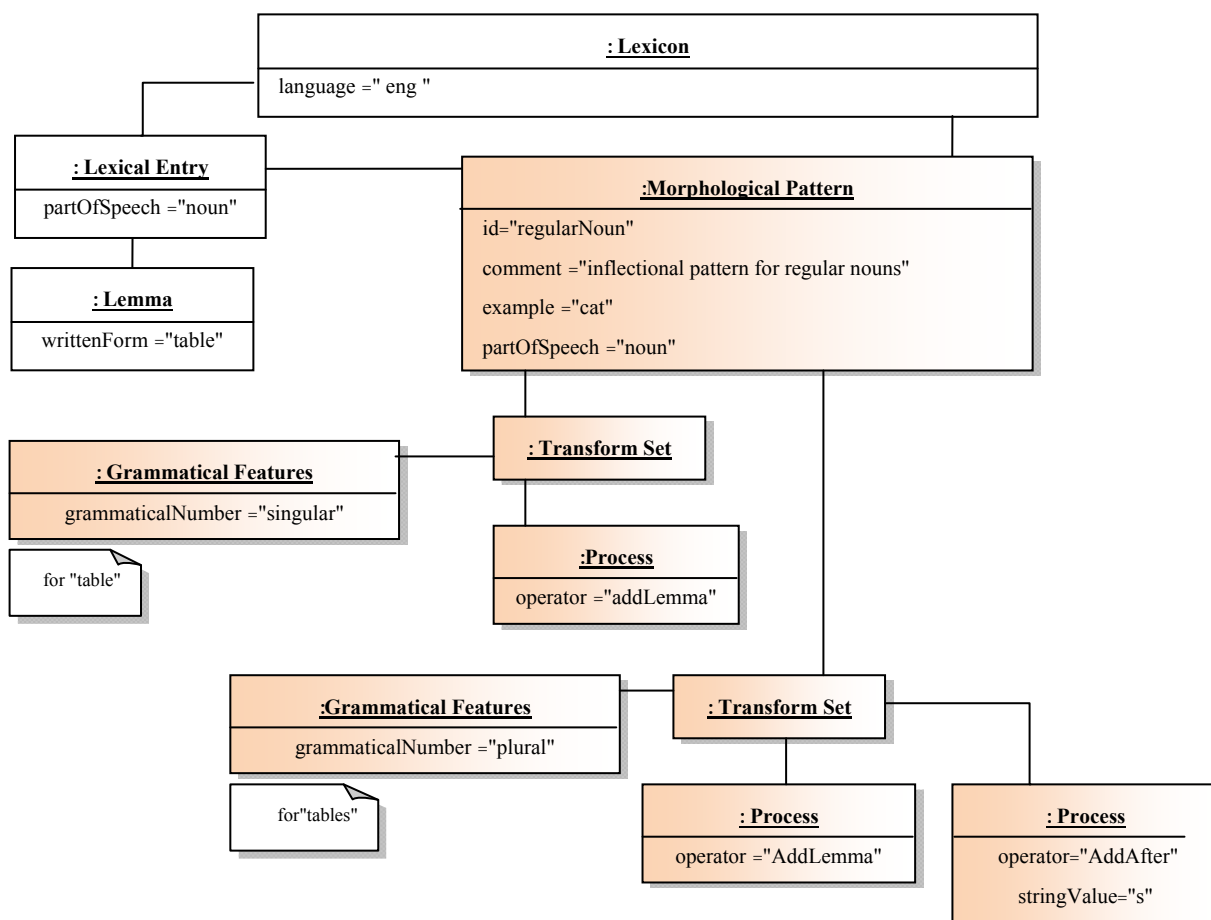
ر-۳-۳ الگوی تصریفی با استفاده از کلاس دسته تبدیل

این مثال یک الگوی تصریفی سنتی مبتنی بر بن‌واژه را با استفاده از اسم انگلیسی "table" نشان می‌دهد. این واژه قاموسی با توجه به الگوی ساختواژه "regularNoun" به شکل تصریفی در نظر گرفته شده است. شیوه‌ای که در این مثال مورد استفاده قرار گرفته است کاملاً مبتنی بر بن‌واژه است. حالت مفرد به عنوان یک بن‌واژه که "table" می‌باشد، در نظر گرفته شده است. حالت جمع به عنوان یک بن‌واژه به همراه یک وند به صورت "s" در نظر گرفته می‌شود. در نمودار فقط یک وند وجود دارد اما در شرایط پیچیده‌تری می‌تواند دارای بیش از یک وند باشد، بنابراین به منظور اتخاذ یک شیوه کلی‌تر این وندها شماره‌گذاری می‌شوند. در این مثال، شماره وند، یک می‌باشد. به شکل ر-۲ رجوع شود.

۱- برای این که فهم این شکل آسان‌تر شود، محتوی کادرهای سایه‌دار برای نمونه‌های مربوط به رده‌هایی که در بسته کنونی تعریف شده‌اند، مورد استفاده قرار می‌گیرند. محتوی کادرهای مربوط به نمونه‌های رده‌هایی که در بسته دیگر تعریف شده، سایه ندارند.



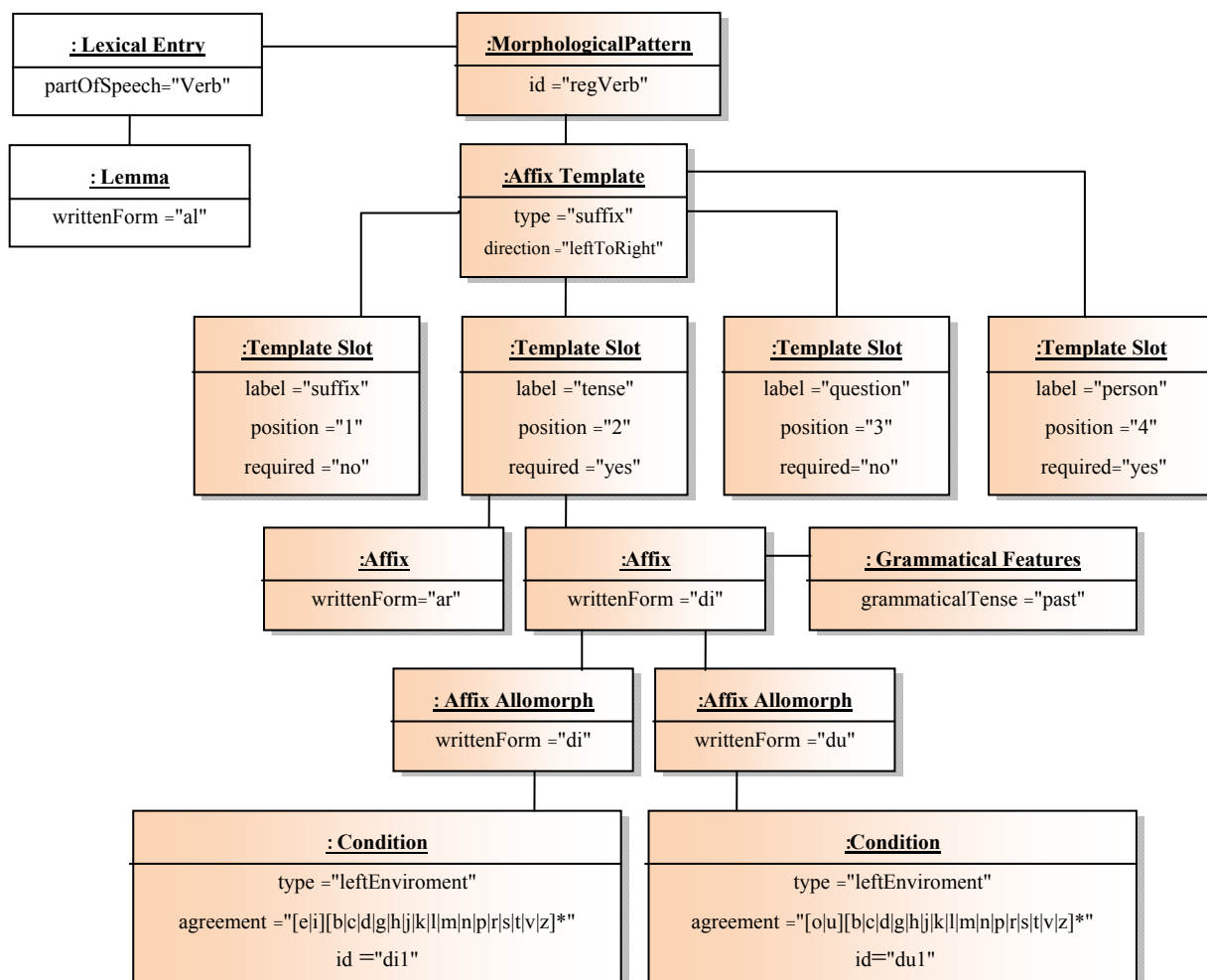
شکل ر-۲- الگوی تصریفی با استفاده از مقوله‌های دسته تبدیل و وند



شکل ر-۳- حالت تصریفی با استفاده از مقوله تبدیل

ر-۳-۴ حالت پیوندی با استفاده از کلاس قالب وند

این مثال، از یک بخش و یک دیدگاه سازمانی خاصی برای اجرای صرف در یک فعل ساده شده زبان ترکی استفاده می‌کند. در نمودار زیر، قالب‌وند، الگوی وندهای پیوندی را مدیریت می‌کند. هر نمونه جایگاه قالب، بعد متفاوت فعل را به نمایش می‌گذارد. این مثال، در واقع پسوند در فعل زمان گذشته را با استفاده از وند به منظور نمایش حالت کلی کلمه و هم چنین یک نمونه از مشخصه‌های دستوری پیوندی برای ایجاد زمان گذشته فعل نشان می‌دهد. تکواژگونه‌های وندی، حالات مختلف کلمه (تکواژگونه‌ها) را که در محیط‌های مختلف آواشناختی ایجاد می‌شوند را نشان می‌دهد. در شکل ر-۴ شی شرط، محیط‌های چپ و راست را با استفاده از اصطلاحات متداولی به عنوان مقادیرشان، نشان می‌دهد.

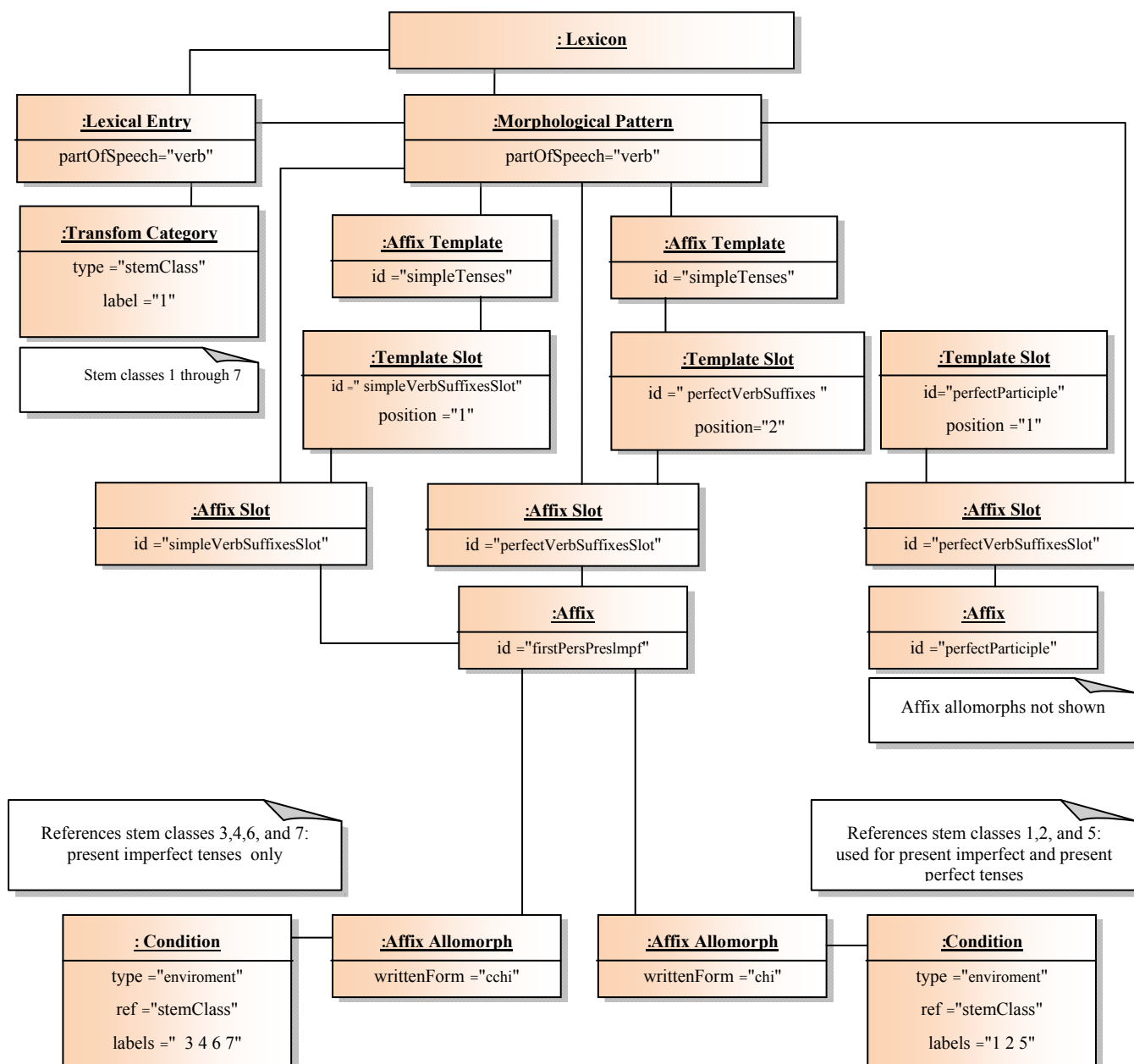


شکل ر-۴- حالت تصریفی با استفاده از کلاس دسته تبدیل

ر-۳-۵ حالت تصریفی با استفاده از کلاس قالب وند

این مثال همان طوری که در شکل ر-۵ نشان داده می‌شود، از یک بخش و دیدگاه سازمانی جهت اجرای صرف فعل ساده شده در زبان بنگلادشی را با استفاده از یک قالب وند جهت مدیریت یک الگو از وندهای تصریفی، استفاده می‌کند. این مدل، کاملاً از مدل پیوندی زبان ترکی (همان طور که در مثال قبل ملاحظه کردید) متفاوت می‌باشد، به طوری که در زبان ترکی، شی قالب وند، یک دسته ترتیبی از نمونه‌های جایگاه قالب را که ابعاد مختلف فعل را نشان می‌دهند، مدیریت می‌کند. در این مثال نمونه‌های قالب وندی جهت ارائه الگوهای پسوندی تصریفی مختلفی مورد استفاده قرار می‌گیرند. ساختار کلاس‌های مدیریتی مربوط به دسته‌های وندها کاملاً پیچیده هستند. برای مثال، قالب وندی که زمان‌های وجه وصفی را ارائه می‌دهد به دو نمونه جایگاه قالب اشاره دارد: یک، مدیریت پسوند اسم مفعول و دو، مدیریت دسته‌ای از پسوندهای فعل در زمان‌های کامل. در زبان بنگلادشی یک زیر مجموعه از پسوندهای فعل ساده نیز به عنوان اجزایی در زمان‌های کامل فعل در نظر گرفته می‌شوند. به منظور جلوگیری از حضور شی‌های وند حشو در یک اجرا، نمونه‌های جایگاه قالب، وندها را به طور غیر مستقیم و از طریق نمونه‌های جایگاه وند مدیریت می‌کنند. نمونه‌های جایگاه وند در بازگشت، به شی‌های وند مشترک اشاره دارند. گزینه هدف این طرح از طریق حالت

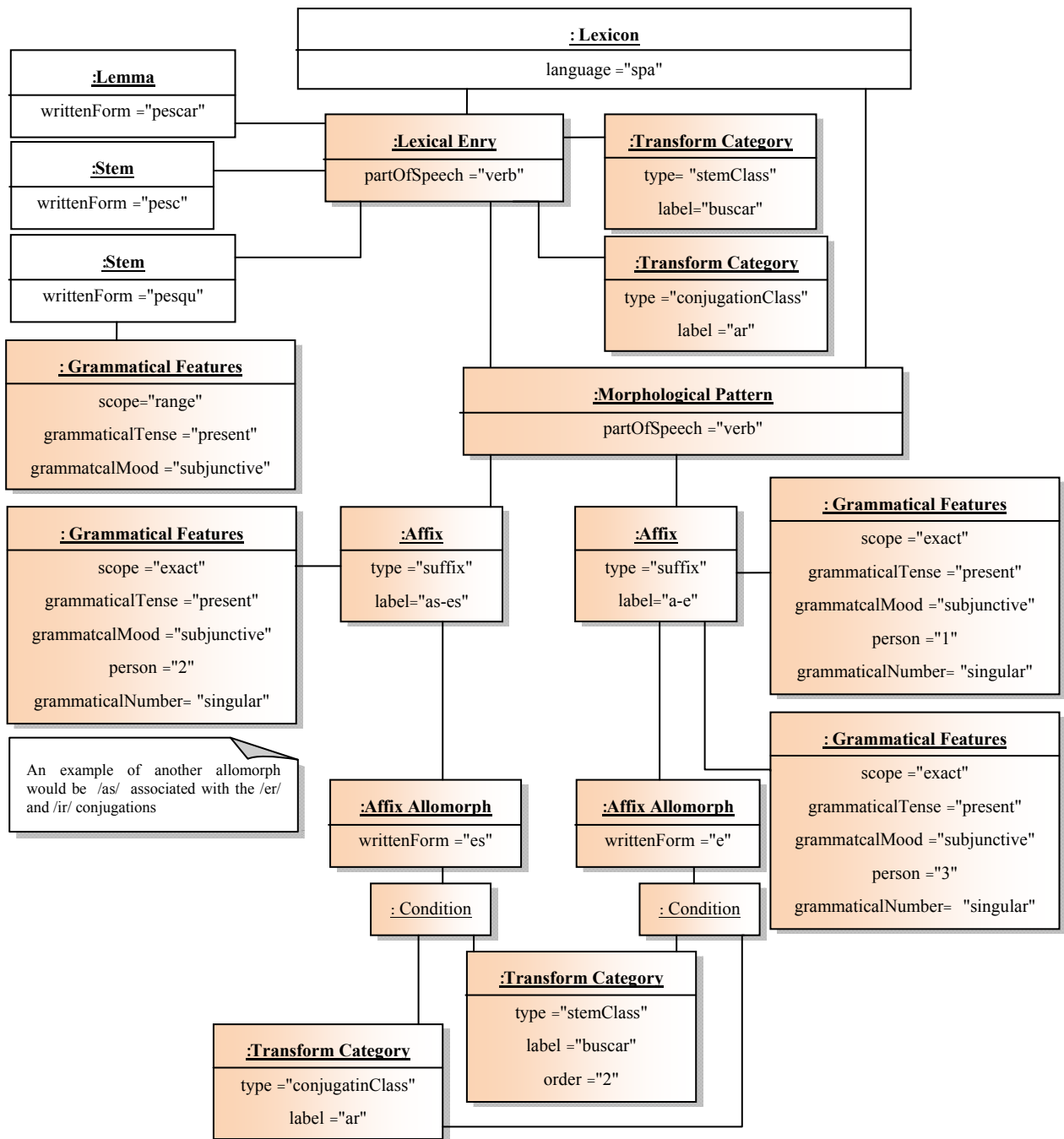
پسوند ماضی استمراری اول شخص که در نمودار زیر نشان داده می‌شود، بررسی می‌شود. این حالت پسوند برای زمان استمراری و به عنوان یک جز در زمان ماضی نقلی مورد استفاده قرار می‌گیرد. از آن جایی که تکواژگونه‌های وندی در زمان‌های مختلفی دارای محیط‌های آواشناختی مختلفی می‌باشند، لذا نمونه‌های شرط ممکن است به دسته‌هایی از نمونه‌های مشخصه‌های دستوری به عنوان محدودیت‌های مربوط، اشاره داشته باشند.



شکل ر-۵ - حالت تصریفی با استفاده از مقوله قالب وند

ر-۳-۶ حالت تصریفی با استفاده از کلاس رده تبدیل به عنوان یک شرط تکواژگونه

شکل ر-۶ زیر مربوط به یک مثال ساده از صرف فعل در زبان اسپانیایی می‌باشد. نمونه‌های رده تبدیل به عنوان یک شرط برای سامان‌دهی تکواژگونه‌های وند و هم چنین برای پیوند رده‌های تکواژگونه‌های وندی با رده‌های تکواژگونه‌های ستاک، مورد استفاده قرار می‌گیرند. مدخل واژگانی با دو نمونه از رده‌های تبدیل که نوع اول "setmClass" و نوع دوم "ConjugationClass" می‌باشد، پیوند دارد. هر نمونه وند، مستقیماً با الگوی ساختارده در ارتباط است و توسط یک یا چند نمونه مشخصه‌های دستوری شرح داده می‌شود. ساختار زبان اسپانیایی، عامل کلیدی در پشتیبانی از این رویکرد طراحی است. رویکرد مزبور در مقایسه با رویکردهای مورد استفاده برای الگوسازی زبان‌های ترکی و بنگلادشی (همان‌طور که در مثال‌های قبلی ملاحظه کردید) ساده‌تر است. مقدار آواشناختی برای نمونه تکواژگونه وندی، توسط مقدار کلاس صرف که یک زیردسته از رده تبدیل می‌باشد، تعیین می‌شود. پیوند یک تکواژگونه وند با یک ستاک خاصی، از طریق رجوع به کلاس ستاک خاصی که یک زیردسته از رده تبدیل می‌باشد و هم چنین از طریق ترتیب ستاک مربوط در آن کلاس ستاک خاص مدیریت می‌شود. در این مثال پیوند بین وند /es/ و بن /pesque/ به‌طور غیرمستقیم از طریق نمونه‌های رده تبدیل مدیریت می‌شوند. یعنی، یک یا چند نمونه مدخل واژگانی می‌تواند با هر دو کلاس صرفی /ar/ و کلاس ستاک /buscar/ ارتباط داشته باشد. در هر مورد، ترتیب دوم نمونه ستاک با تکواژ وند /es/ ارتباط دارد. این مثال هم چنین یک روش جایگزین برای پیوند تکواژگونه‌های ستاک با تکواژگونه‌های وند و هم چنین با پیوند هر کدام به نمونه مشخصه‌های دستوری را نشان می‌دهد.

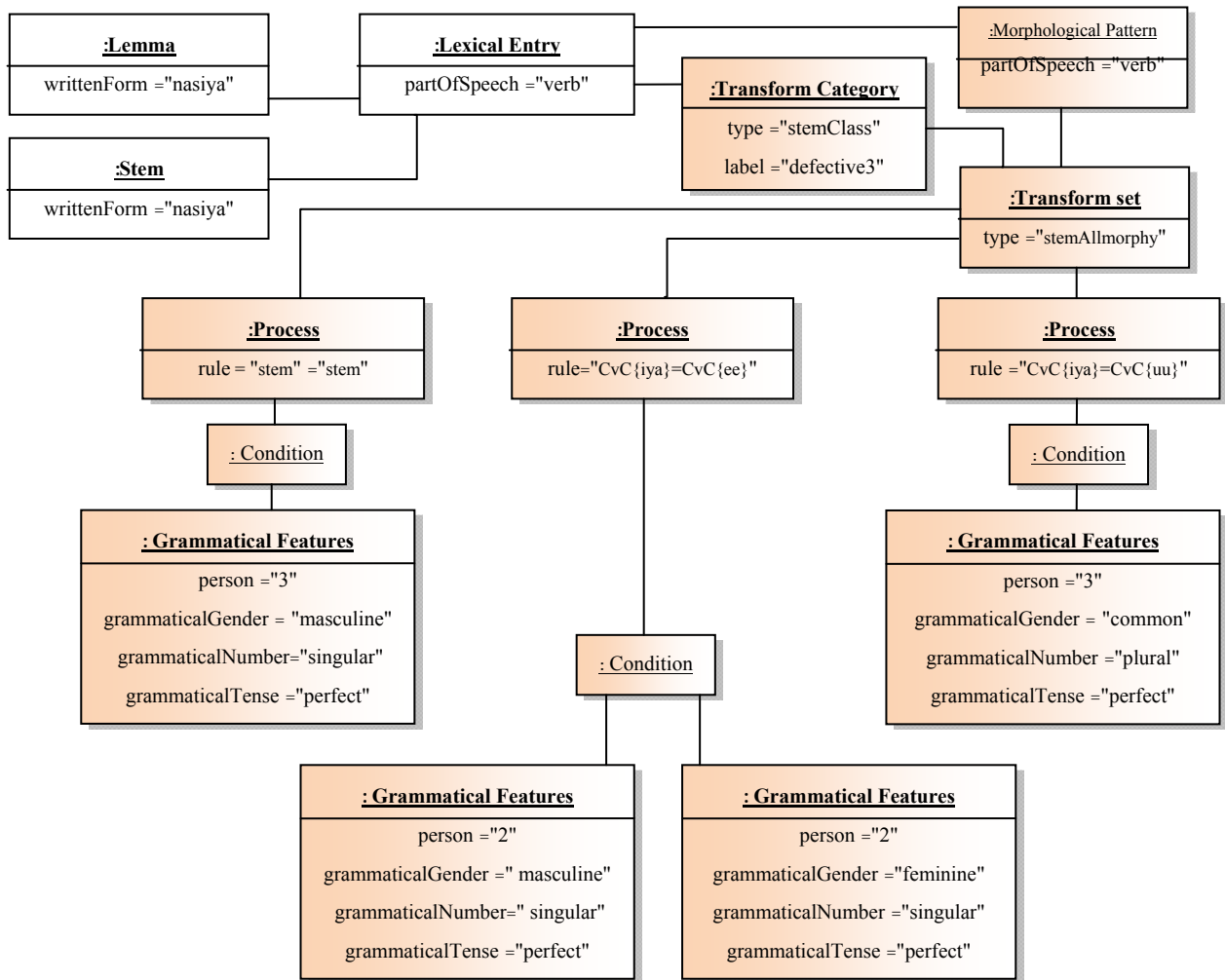


شکل ر-۶- حالت تصریفی با استفاده از مقوله رده تبدیل

ر-۳-۷ حالت تصریفی بر پایه تکواژگونه و با استفاده از رده‌های پردازش

اغلب، فرهنگ لغات حاوی مثال‌های تمامی تکواژگونه‌های ستاک مورد نیاز برای گسترش اجرای ساختارده تأکیدی نمی‌باشند. در چنین مواردی، ایجاد تکواژگونه‌های ستاک با استفاده از قوانین آواشناختی و شرایط محیطی، الزامی است. شکل زیر رویکرد خاصی را برای ایجاد تکواژگونه‌های ستاک در زبان عربی متداول امروزی را در یک آوانویسی بررسی می‌کند. در این مثال نمونه رده تبدیل به پردازشی را که هر کدام یک تکواژگونه ستاک متفاوتی را تولید می‌کنند، مدیریت می‌کنند. نمونه رده تبدیل شامل رده تبدیل با زیر رده

"stemClass" و با مقدار "defective3" می‌باشد. هر پردازشی توسط شرط بخصوصی ایجاد می‌شود، که در این مورد این شروط توسط ارجاع به دسته نمونه‌های مشخصه‌های دستوری تعیین می‌شوند. این پردازش می‌تواند برای تمامی ستاک‌های واقع در شی‌های مدخل واژگانی که شامل یک پیوند با نمونه رده تبدیل با زیردسته "stemClass" و با مقدار "defective3" است، به کار می‌رود. این مثال، همان طوری که در شکل ر-۷ ملاحظه می‌شود، می‌تواند در تلفیق با رویکردهای طراحی دیگری جهت تولید تکواژگونه‌های ستاک و هم چنین پیوند آن تکواژگونه‌های ستاک با تکواژگونه‌های وندی مرتبطی مورد استفاده قرار بگیرد.

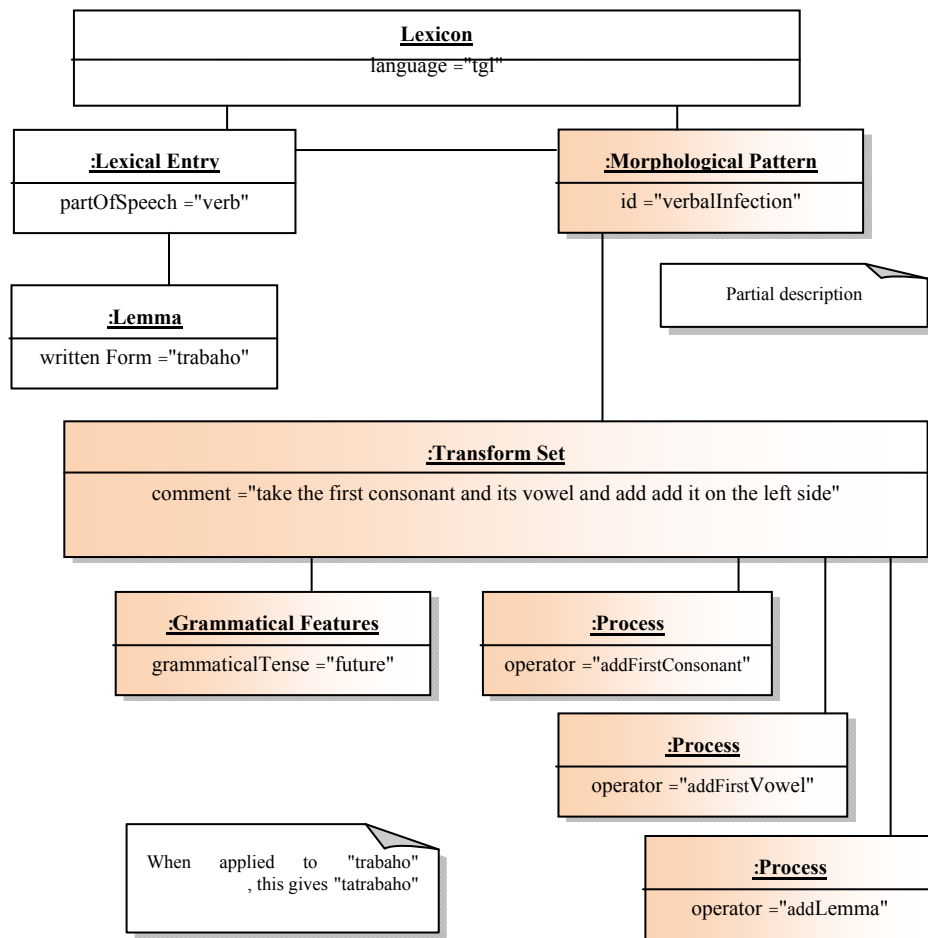


شکل ر-۷- حالت تصریفی بر پایه تکواژگونه و با استفاده از مقوله‌های پردازش

ر-۳-۸ صرف پیچیده: حالات فعلی در زبان تاگالوگ^۱

شرح فعل‌های زبان تاگالوگ توسط یک بخش محض و هم چنین رویکرد ترتیب، بسیار مشکل است. این مثال بخصوص چگونگی تشکیل زمان آینده را توسط گرفتن حرف اول صامت کلمه و اضافه کردن حرف اول مصوت و افزودن این حروف به سمت چپ بن‌واژه را نشان می‌دهد. به شکل ر-۸ رجوع شود.

1-Tagalog



شکل ر-۸- حالات فعلی در زبان تاگالوگ

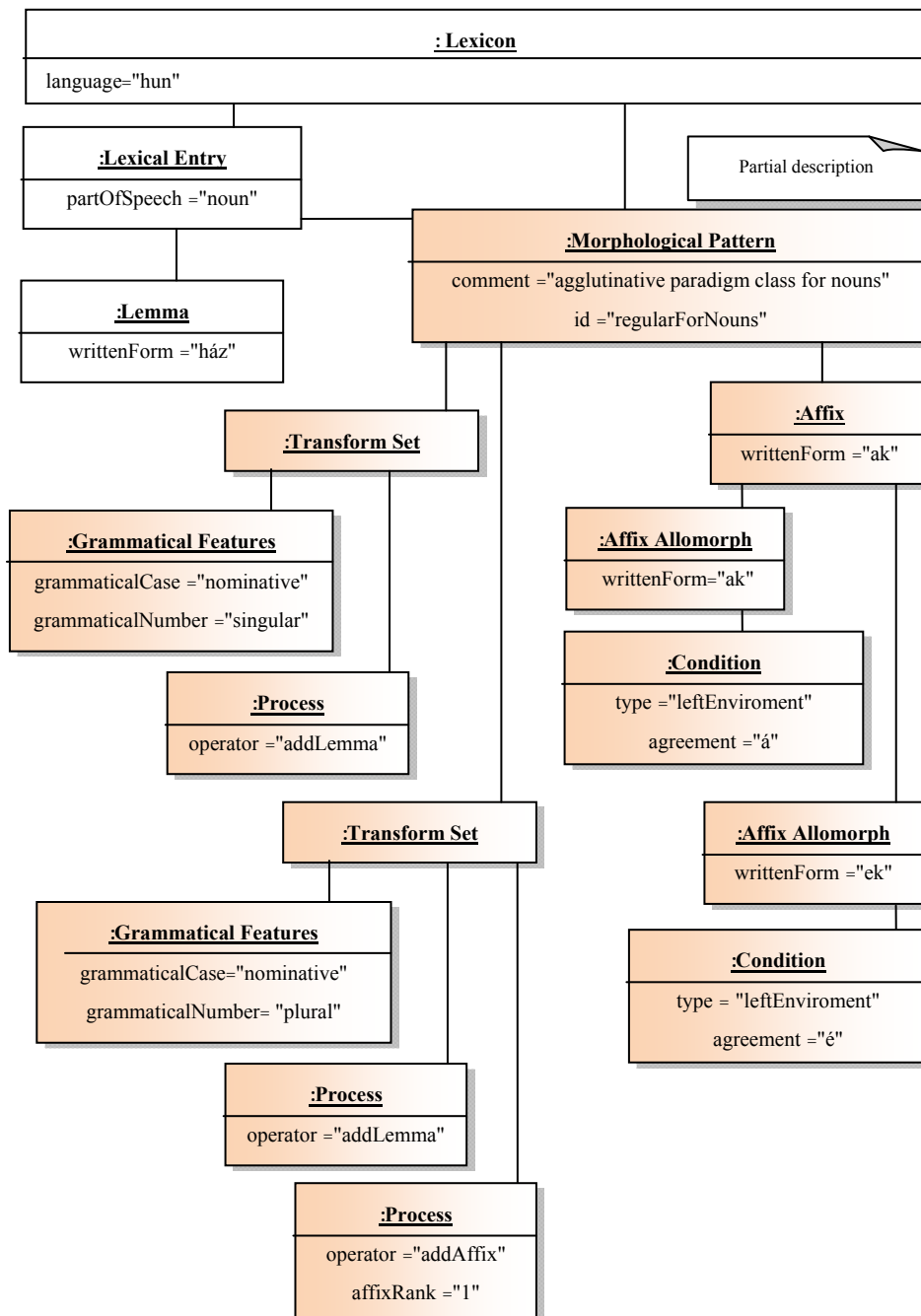
ر-۳-۹ حالت پیوندی سرواژه بنیاد

این مثال، یک روش مبتنی بر بن‌واژه را با یک الگوی حالت تاکید پیوندی چندگانه^۱ (MUF) برای تکواژگونه‌ها به اجرا درمی‌آورد. مثال‌هایی را که در زیرعبارت‌های قبل ملاحظه کردید مربوطه به زبان‌های تصریفی بودند. علی‌رغم این موضوع، مثال زیر در رابطه با زبان مجارستانی که یک زبان پیوندی است، می‌باشد. در زبان مجارستانی، دو سیستم به هم مرتبط که چندین بار تکرار خواهند شد، بر این زبان تاثیر می‌گذارند: سیستم پسوندی که یک ستاک با یک پسوند پیوند می‌خورد، و سیستم تلفیق مصوت که در آن یک وند با نگرش به توافق مصوت انتخاب می‌شود.

برای مثال: "ház" (خانه) "ház+ak" (خانه‌ها) را به دلیل "á" می‌دهد، اما "szék" (صندلی) "szék+ek" (صندلی‌ها) را به دلیل "é" می‌دهد. این سیستم تا حدودی کلی می‌باشد ولی نمی‌تواند با کل فرهنگ لغت ارتباط داده شود زیرا این سیستم دارای چندین استثناء است. این استثناءها می‌بایستی در نمونه‌های الگوی

1-Multiple underlying form

ساختواره ثبت شوند. برای مثال واژگان قاموسی معنی شده که دارای گونه‌های مختلفی مانند "hotelban" در مقابل با "hotelben" (در هتل) می‌باشند، قانون کلی را نقض می‌کنند. هماهنگی مصوت توسط نمونه شرط در پیوند با نمونه تکواژگونه‌های وندی نشان داده می‌شوند، به شکل ر-۹ رجوع شود.



شکل ر-۹- حالت پیوندی مبتنی بر بن‌واژه

شایان ذکر است با توجه به این واقعیت که در زبان مجارستانی شمار حالات به عنوان مثال برای اسم، بیش از دویست عدد می‌باشد، روش فهرست‌بندی تمامی حالات پیوندی به طور مشخصی در فرهنگ لغت (مانند پیوست الف) یک نوع مستندسازی غیر قابل مدیریتی را ایجاد می‌کند. معمولاً روشی که برپایه الگوهای ساختارده بنا نهاده شده باشد، بیشتر ترجیح داده می‌شود.

ر-۳-۱۰ حالت اشتقاقی با استفاده از تکرار

در زبان تایلندی، اشتقاق یک سیستم متداول می‌باشد، بنابراین ثبت مدخل مجزا برای حالت اشتقاقی اتلاف وقت می‌باشد. تکرار برای شرح مفهوم واژه قاموسی توسط برخی عملیات به منظور تکرار صدای سر واژه مورد استفاده قرار می‌گیرد.

انواع اشتقاق با استفاده از تکرار عبارتند از:

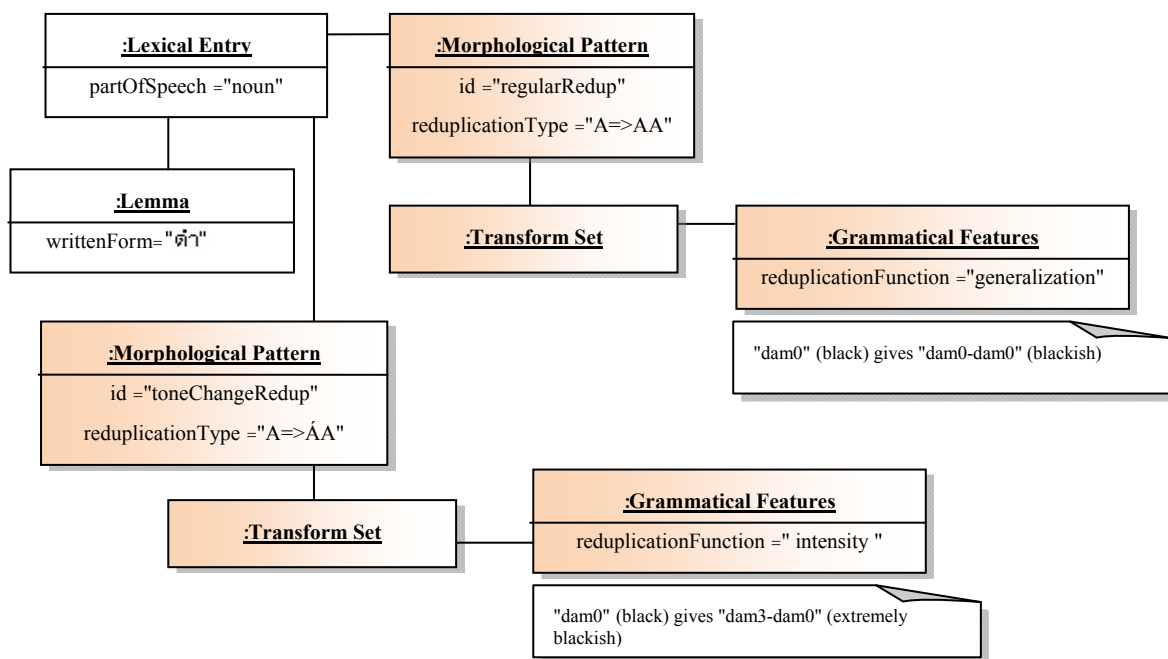
ر-۳-۱۰-۱ نوع AA. حالت مضاعف‌سازی توسط پیوند یک نماد کاراکتر "mai Yamok" (ا) برای تولید صدای مضاعف یک سرواژه ایجاد می‌شود. برای نمونه سرواژه "๓๑" (که به شکل "dam0" تلفظ شده و به معنی "سیاه" می‌باشد) می‌تواند به منظور بیان چگونگی تولید آن به شکل "๓๑๑" (که به صورت "dam0" تلفظ شده و به معنی "نسبتاً سیاه" می‌باشد) شرح داده شود.

ر-۳-۱۰-۲ نوع A'A (تغییر لحن در هجای اول)، برای نمونه "๓๑๓" (که به شکل "dam3-dam0" تلفظ شده و به معنی "کاملاً سیاه" می‌باشد) برای نشان دادن شدت مورد استفاده قرار می‌گیرد.

ر-۳-۱۰-۳ نوع AA'A (سه گانگی)، برای نمونه "กึ๓กึ๓กึ๓" (که به شکل "kin0-kin4-kin0" تلفظ شده و به معنی "خوردن مثل یک اسب" می‌باشد) برای نشان دادن شدت مورد استفاده قرار می‌گیرد.

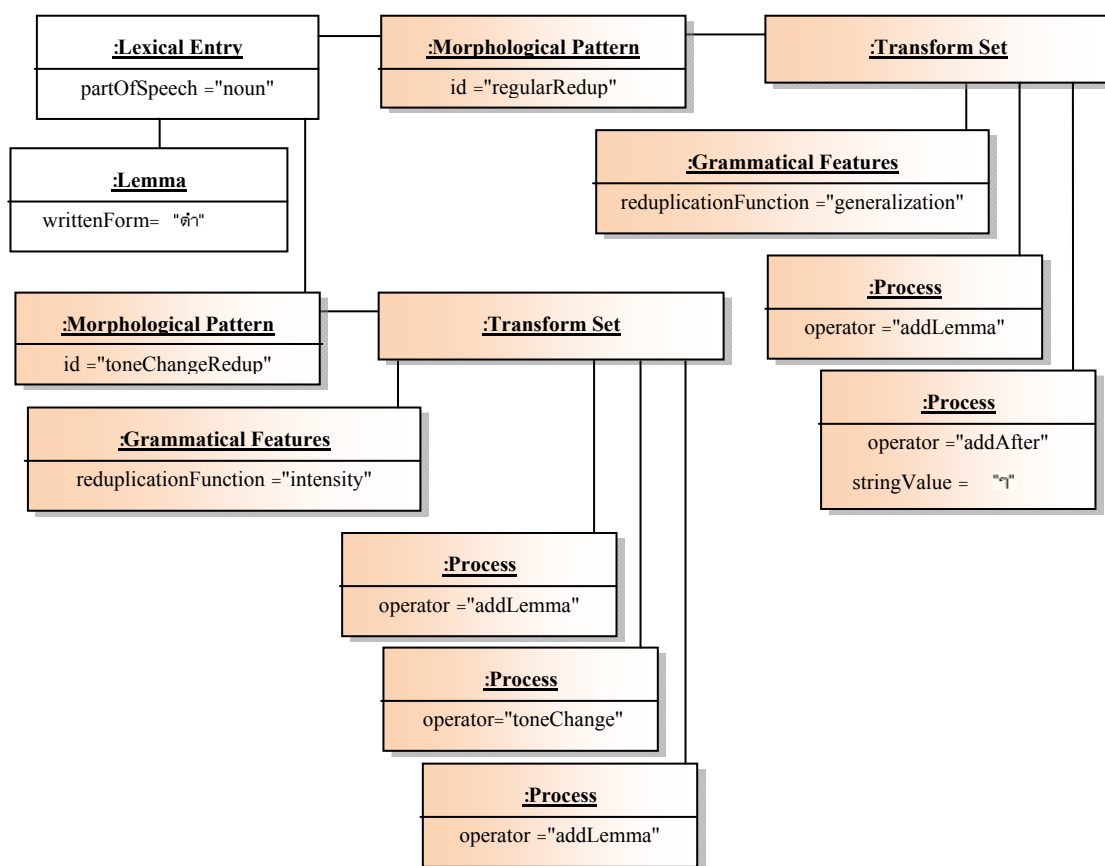
ر-۳-۱۰-۴ سیستم‌های پیچیده‌تری مانند انواع AABB یا AB'AB.

دو نمودار زیر گزینه‌های مختلفی را برای نمایش اشتقاق نشان می‌دهد. قابل ذکر است که در این مثال‌ها علی‌رغم مثال‌های دیگر، دو نمونه الگوی صرفی به نمونه مدخل واژگانی مفروض پیوست شده‌اند. رویکرد اول، همان طوری که در شکل ر-۱۰ ملاحظه می‌کنید، عمدتاً برای نشانه‌گذاری نمونه الگوی صرفی توسط ویژگی نوع تکرار می‌باشد.



شکل ر-۱۰- حالت اشتقاقی با استفاده از تکرار (رویکرد ساده)

یک نوع دیدگاه کاملا پیچیده و دقیق، همان طوری که در شکل زیر ملاحظه می کنید به وسیله نمونه های پردازش، شرح کامل تکرار می باشد.



شکل ر-۱۱- حالت اشتقاقی با استفاده از تکرار (رویکرد پیچیده)

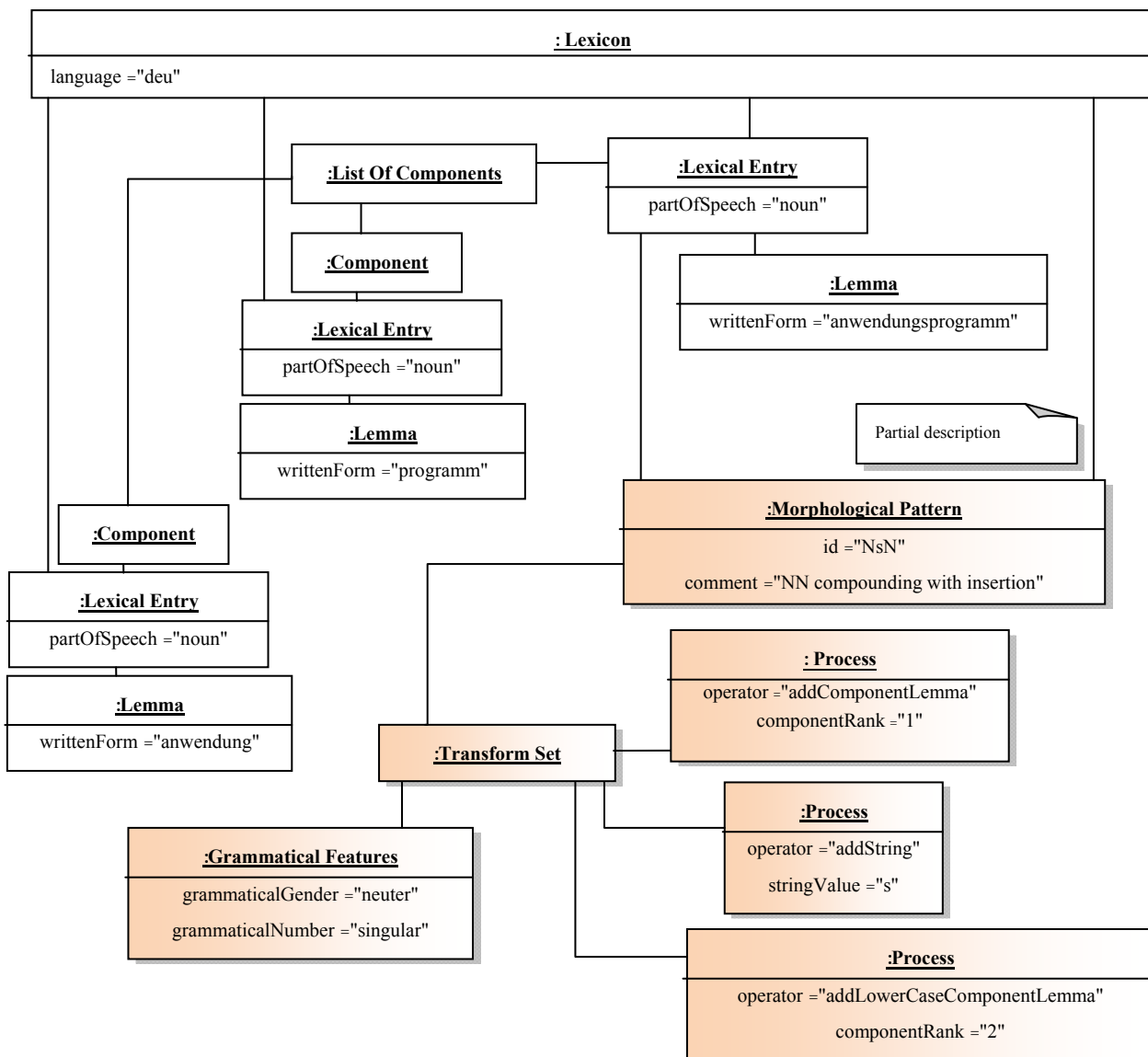
شایان ذکر است که تکرار مختص اشتقاق نمی‌باشد بلکه در زبان‌هایی چون زبان اندونزیایی به منظور شرح حالات جمع نیز وجود دارد.

ر-۳-۱۱ ترکیب مبتنی بر بن‌واژه: 'Anwendungsprogramm'

ترکیب تا حدودی از تصریف و پیوند و اشتقاق متفاوت می‌باشد. در مثال‌های اخیر، هر حالت توسط بن‌واژه (یا یک ستاک) یک مدخل با انواع مختلف عملیات مانند اضافه کردن وندها تعریف شده است. این حالت با استفاده از یک مدخل صورت می‌پذیرد.

در فرآیند ترکیب هر حالت مولفه مدخل مرکب در نمونه فهرست مولفه‌ها تعریف شده است. در این حالت، فرآیند، به عملکرد ساختواژه مجزای هر مولفه بستگی دارد.

نمودار زیر، یک مثال زبان آلمانی را از یک الگوی ساختواژه تصریفی اعمال شده بر یک نوع ترکیب که استفاده از یک عامل مجزاساز املائی صحیح را می‌طلبد، مورد بررسی قرار می‌دهد. حالات ترکیبی از دو مولفه حاضر در نمونه فهرست مولفه‌ها، استنتاج می‌شود. بن‌واژه به الگوی صرفی که چگونگی ساخت ترکیب را شرح می‌دهد، متصل می‌شود. مولفه اول انتخاب شده، یک "s" اضافه می‌شود و سپس مولفه دوم با حرف اول بزرگ تبدیل شده به حرف کوچک اضافه می‌شود. به شکل ر-۱۲ رجوع شود.

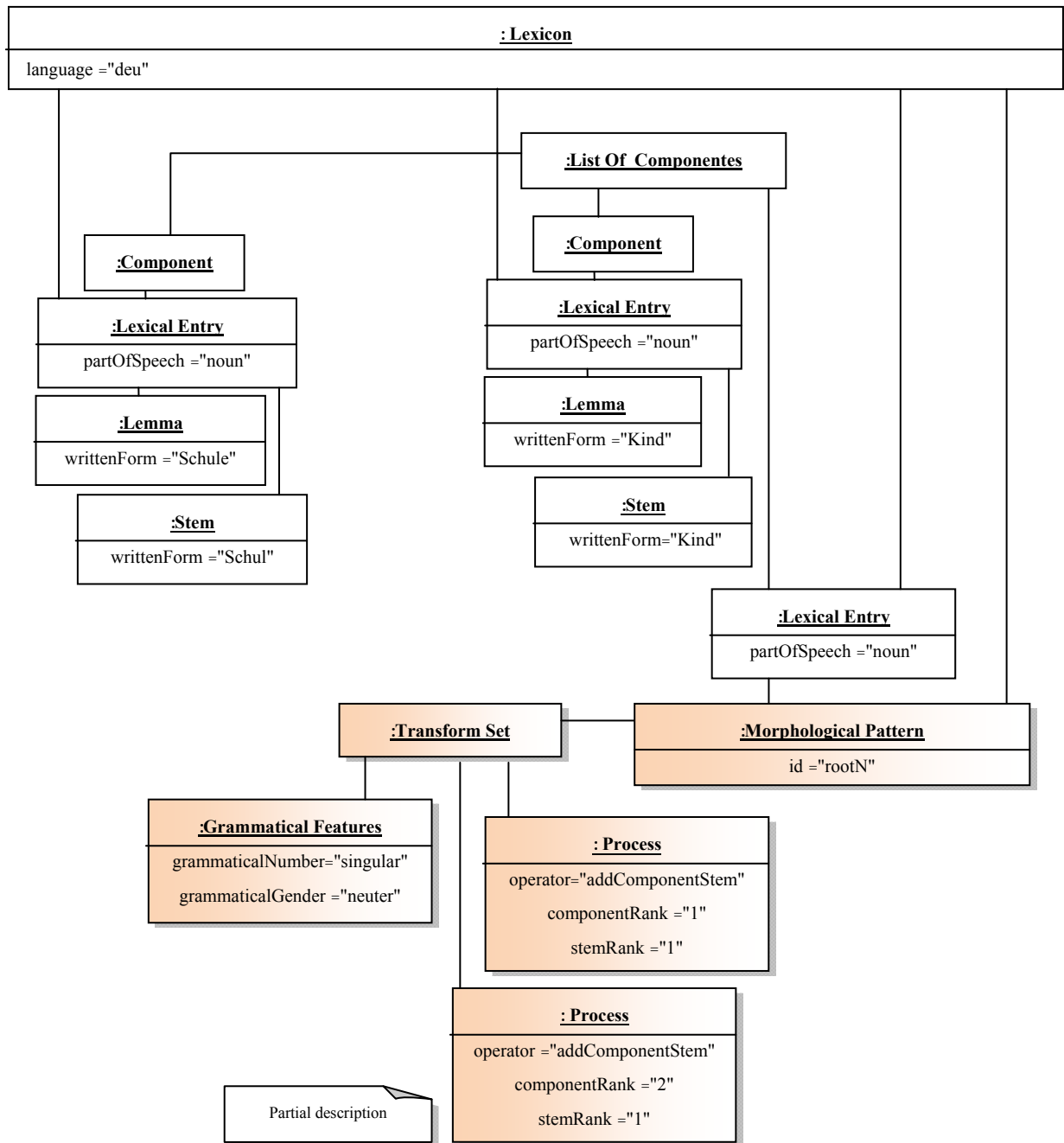


شکل ر-۱۲- مثال یک ترکیب مبتنی بر بن‌واژه

ر-۳-۱۲ ترکیب مبتنی بر ستاک : Schulkind^۱

شکل زیر، مثال دیگری از زبان آلمانی در الگوی ساختواژه اعمال شده بر ترکیبات اسم به اسم که استفاده از ستاک را می‌طلبد، مورد بررسی قرار می‌دهد. در مواردی که آخر سرواژه "e" و یا "en" می‌باشد، در عمل کوتاه‌سازی حین ترکیب، آخر بن‌واژه‌ها حذف می‌شوند. در این مثال "Schule+kind" موجب به وجود آمدن "Schulkind" می‌شود. برخلاف استفاده از عملیات حذف بر روی "Schule"، این مثال رویکردی را که "Schul" را به عنوان ستاک برای اسم "Schule" در نظر می‌گیرد را مورد بررسی قرار می‌دهد، با این فرض که افزودن‌های داخلی، حرف اول بزرگ را به حرف کوچک تبدیل می‌کند. به شکل ر-۱۳ مراجعه شود.

1- Schulkind



شکل ر-۱۳- مثال ترکیب مبتنی بر ستاک

پیوست ز (الزامی)

بسط الگوهای اصطلاحات چند کلمه‌ای NLP

ز-۱ اهداف

هدف این پیوست، بازنمایی ساختار داخلی (نیمه ثابت و یا منعطف) اصطلاحات چند کلمه‌ای در یک زبان مفروض می‌باشد.

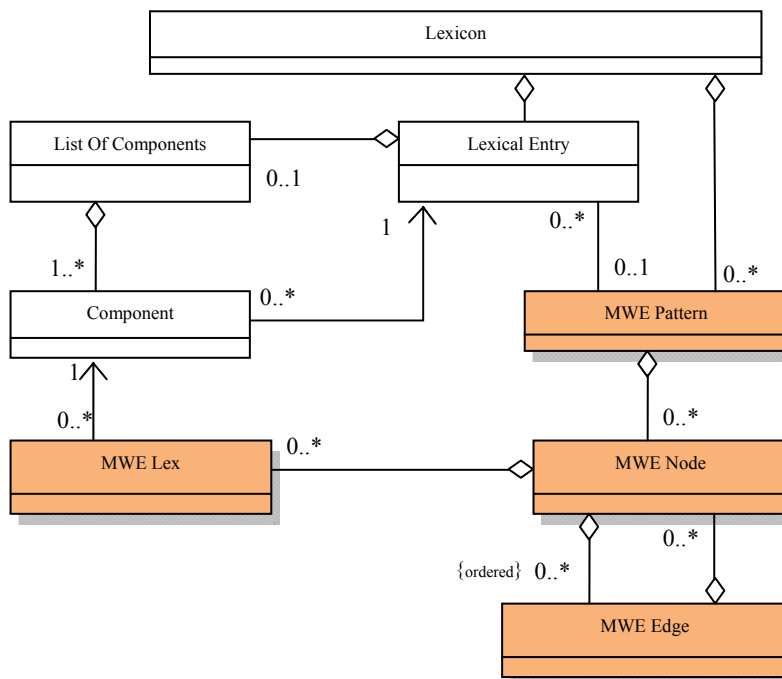
در همه زبان‌ها، اصطلاحات چند کلمه‌ای شامل دسته وسیعی از پدیده‌های مجزا و در عین حال مرتبط همچون مرتب‌سازی، افعال گروهی، ترکیب‌های اسم به اسم و غیره می‌باشد. در برخی سیستم‌ها و سنت‌های زبان‌شناسی، اصطلاحات کوتاه را نیز جز اصطلاحات چند کلمه‌ای می‌دانند. هر چند که برخی از اصطلاحات چند کلمه‌ای ثابت هستند و مانند *ad hoc* اجازه تداخل را نمی‌دهند، اما برخی دیگر کاملاً منعطف بوده و اجازه هرگونه تغییر و تعدیلی را می‌دهند.

ز-۲ بودن یا نبودن الگوهای اصطلاحات چند کلمه‌ای

شرح برخی اصطلاحات چند کلمه‌ای با استفاده از بسط الگوی ساختاردهنده ممکن است، اما برخی موارد، محدود به اصطلاحات چند کلمه‌ای ساده و بدون هر نوع تغییری می‌باشند. علی‌رغم این موضوع، این پیوست تحلیل تمامی اصطلاحات چند کلمه‌ای را بر اساس دستور زبان در بر دارد.

ز-۳ نمودار کلاس

الزاماً کلاس الگوی اصطلاحات چند کلمه‌ای، همان طور که در شکل زیر ملاحظه می‌شود، یک دستور ساخت عبارت می‌باشد.



شکل ز-۱- مدل مقوله الگوی اصطلاحات چند کلمه‌ای

ز-۴ شرح مدل الگوی اصطلاحات چند کلمه‌ای

ز-۴-۱ کلاس الگوی اصطلاحات چند کلمه‌ای

کلاس الگوی اصطلاحات چند کلمه‌ای، کلاسی است نمایانگر نوع معینی از پدیده‌های ترکیبی واژگان. یک الگو همیشه به لیست نمونه مولفه‌های مربوط از نمونه مدخل واژگانی ارجاع دارد. الگوی اصطلاحات چند کلمه‌ای نبایستی برای یک نمونه مدخل واژگانی که به نمونه فهرست مولفه‌ها متصل نیست، مورد استفاده قرار گیرد. یک نمونه الگوی اصطلاحات چند کلمه‌ای با استفاده از نمونه‌های گره اصطلاحات چند کلمه‌ای شرح داده می‌شود.

ز-۴-۲ کلاس گره اصطلاحات چند کلمه‌ای

کلاس گره اصطلاحات چند کلمه‌ای، کلاسی است نمایانگر جزئیات ساختار اصطلاحات چند کلمه‌ای. یک نمونه ترکیب‌دهنده می‌تواند به صفر یا چند نمونه محدوده اصطلاحات چند کلمه‌ای متصل باشد.

ز-۴-۳ کلاس محدوده اصطلاحات چند کلمه‌ای

کلاس محدوده اصطلاحات چند کلمه‌ای، کلاسی است نمایانگر جزء کوچکی از اطلاعات به عنوان کلاس گره اصطلاحات چند کلمه‌ای. نمونه محدوده اصطلاحات چند کلمه‌ای به طور بازگشتی به نمونه گره اصطلاحات چند کلمه‌ای متصل می‌باشد.

ز-۴-۴ کلاس واژگان اصطلاحات چند کلمه‌ای

کلاس واژگان اصطلاحات چند کلمه‌ای، کلاسی است نمایانگر ارجاع به یک جزء واژگانی. هر چند که هدف کل این مجموعه که ایجاد یک نمایش کلی از ترکیب‌های اصطلاحات چند کلمه‌ای در یک زبان مفروض می‌باشد، لیکن به مولفه‌ها به طور مستقیم اشاره نشده است، اما در عوض این مولفه‌ها توسط ترتیب خاصشان همان طوری که در نمونه فهرست مولفه‌ها مشخص شده است، مورد ارجاع قرار می‌گیرد.

پیوست ژ
(اطلاعاتی)

مثال‌های الگوهای اصطلاح چند کلمه‌ای NLP

ژ-۱ مثال آرایش کلاس

کلاس‌ها ممکن است با ویژگی‌های مندرج در جدول ژ-۱ آرایش شوند:

جدول ژ-۱ - آرایش کلاس برای الگوهای اصطلاحات چند کلمه‌ای NLP

نام کاس	مثالی برای ویژگی‌ها	توضیح
الگوی اصطلاحات چند کلمه‌ای	شناسه تفسیر	هدف نمودار الگوی اصطلاحات چند کلمه‌ای به اشتراک‌گذاری تمامی حالاتی را که دارای این ساختار می‌باشند، است. نمونه‌های الگوی اصطلاحات چند کلمه‌ای، منابع مشترک می‌باشند و بنابراین با یک شناسه همراه هستند، بدین صورت آن نمونه‌ها می‌توانند توسط ارتباطات بازگشتی نشانه‌گذاری شوند.
گره اصطلاحات چند کلمه‌ای	ساختار نحوی محدوده معنایی شمار دستوری	
محدوده اصطلاحات چند کلمه‌ای	عملکرد	
واژگان اصطلاحات چند کلمه‌ای	راس ساختار ردیف جداساز نمودار	

ژ-۲ مثال شرح کلمه

مثال "to throw somebody to the lions" می‌باشد، به شکل ژ-۱^۱ رجوع کنید:

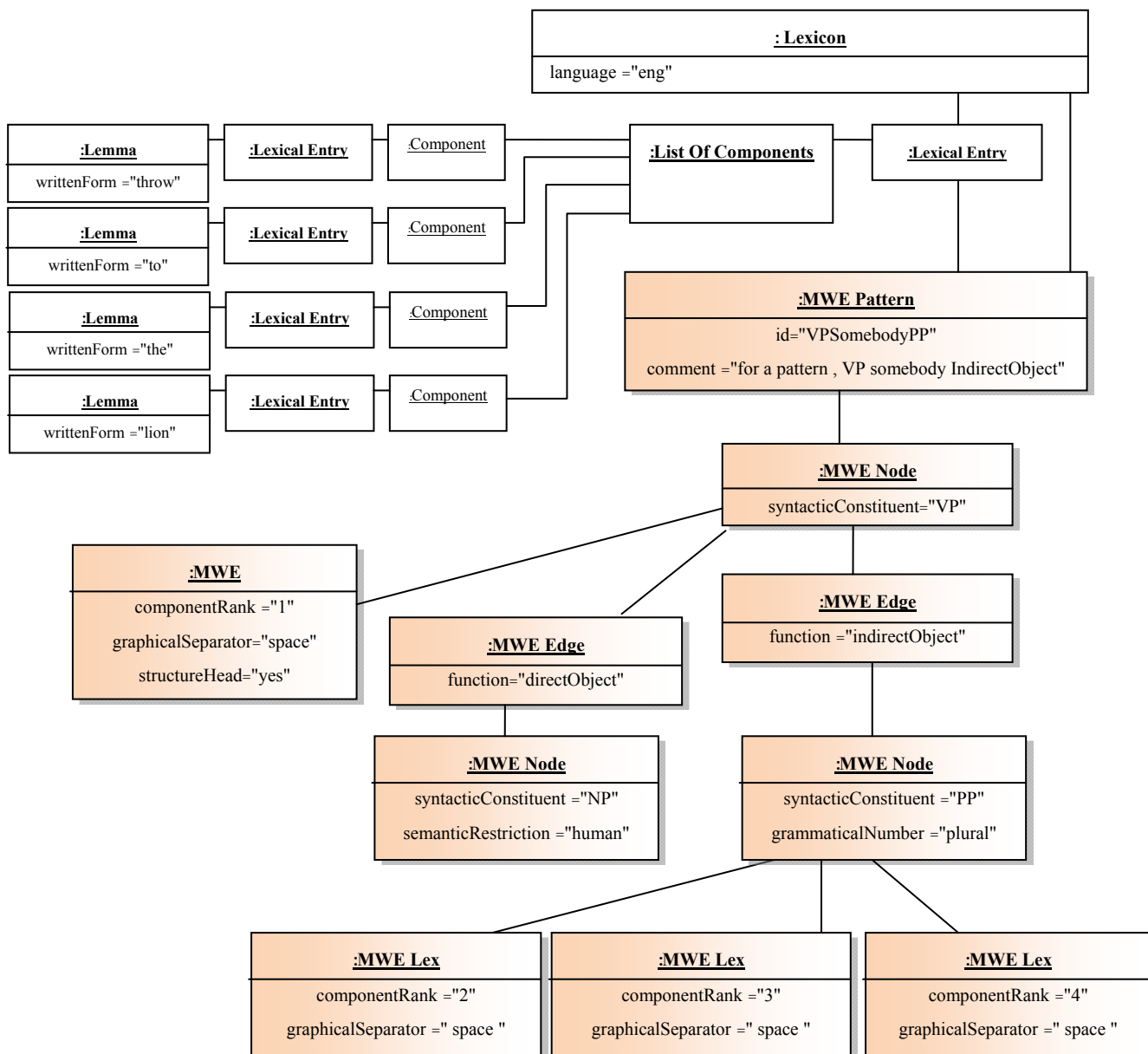
ساختار، شامل سه زیر ساختار می‌باشد:

- یک فعل کاملاً مشخص: throw (پرت کردن)، توسط ردیف یک در نمونه فهرست مولفه‌ها بدان اشاره می‌شود.

- یک گروه اسمی: somebody (کسی)، دارای مفهوم کاملاً مشخصی نمی‌باشد و تنها محدودیتی در هسته عبارت آن در نظر گرفته شده است که می‌تواند /human/ باشد.

- یک گروه اسمی کاملاً مشخص: to the lions (برای شیرها)، توسط ردیف ۲،۳ و ۴ در نمونه فهرست مولفه‌ها بدان اشاره می‌شود. این عبارت حرف اضافه‌ای به عنوان /plural/ برچسب می‌خورد.

۱- برای این که فهم این شکل آسان‌تر شود، محتوی کادرهای سایه‌دار برای نمونه‌های مربوط به مقوله‌هایی که در بسته کنونی تعریف شده‌اند، مورد استفاده قرار می‌گیرند. محتوی کادرهای مربوط به نمونه‌های مقوله‌هایی که در بسته دیگر تعریف شده، سایه ندارند.



شکل ژ-۱- مثال الگوی اصطلاحات چند کلمه‌ای

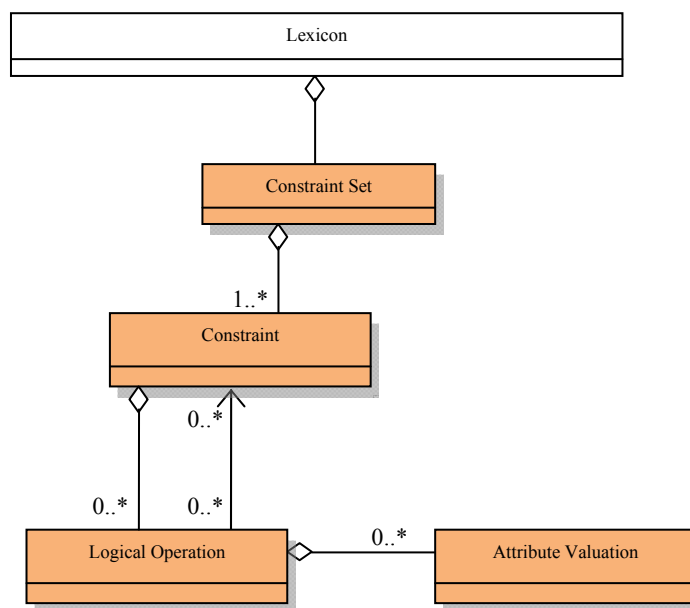
پیوست س
(الزامی)
بسط بیان محدودیت

س-۱ اهداف

هدف، شرح محدودیت‌های روی جفت‌های صفت - مقدار می‌باشد. محدوده این محدودیت‌ها در نمونه فرهنگ لغت وجود دارد.

س-۲ نمودار کلاس

بسط اصطلاح محدودیت مطابق با شکل زیر سازمان‌دهی می‌شود.



شکل س-۱- الگوی اصطلاح محدودیت

س-۳ شرح الگوی اصطلاح محدودیت

س-۳-۱ دسته محدودیت

دسته محدودیت، کلاسی است نمایانگر محدودیت‌ها. این کلاس به کلاس فرهنگ لغت با کاردینالیته صفر به چند متصل می‌شود.

مثال - دسته محدودیت، چگونگی ترکیب هر قسمت از مقدار گفتار مربوط به یک زبان مفروض را با مشخصه‌های دستوری آن زبان نشان می‌دهد.

نکته - می‌توان نمونه دسته محدودیت را برای نمایش‌های صرفی یک نمونه فرهنگ لغت و همچنین برای نمایش‌های نحوی همان فرهنگ لغت مفروض تعریف کرد.

س-۳-۲ محدودیت

محدودیت، کلاسی است نمایانگر یک یا چندین اصطلاح دودویی^۱ که می‌بایستی در یک نمونه فرهنگ لغت مفروض، رعایت شوند.

س-۳-۳ عملیات منطقی

عملیات منطقی، کلاسی است که یک اصطلاح دودویی را بین نمونه‌های ارزیابی ویژگی و نمونه‌های محدودیت نمایش می‌دهد.

مثال - یک نمونه عملیات منطقی می‌تواند یک حرف ربط (برای مثال یک "و" منطقی) و یا یک حرف ربط جداساز (برای مثال "یا" منطقی) را ارائه دهد.

نکته - می‌توان نمونه محدودیت را به طور بازگشتی به عنوان یک ترکیب منطقی از نمونه‌های محدودیت‌های دیگری تعریف کرد.

س-۳-۴ ارزیابی ویژگی

ارزیابی ویژگی، کلاسی است که یک جفت بین نام ویژگی یک کلاس LMF و مقدار این ویژگی بخصوص را ارائه می‌دهد.

مثال - نمونه ارزیابی ویژگی می‌تواند جفت اجزای کلام و صفت باشد.

1-Boolean

پیوست ش
(اطلاعاتی)
مثال بیان محدودیت

ش-۱ مثال آرایش کلاس

به جدول ش-۱ مراجعه شود.

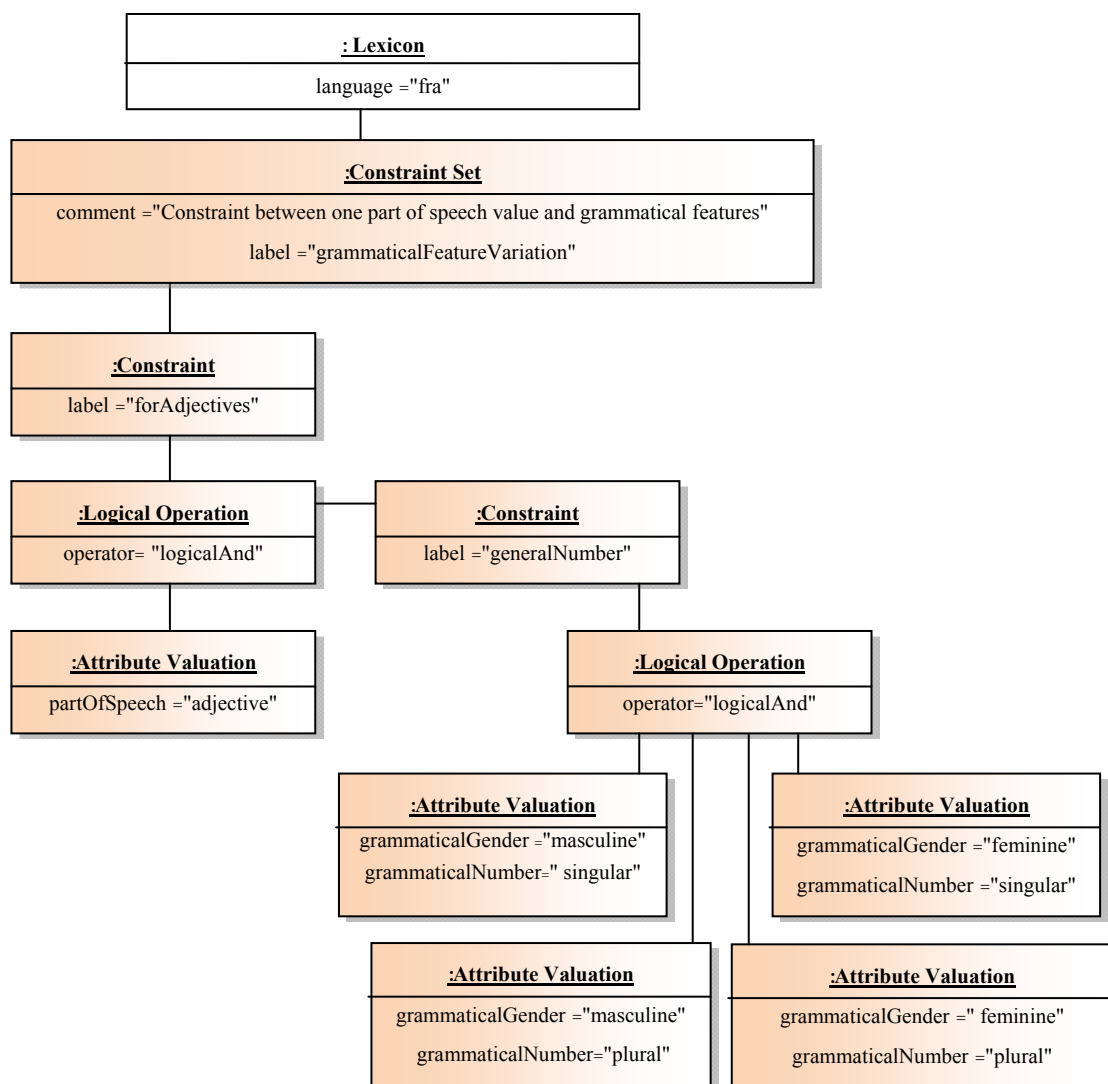
جدول ش-۱- آرایش کلاس برای اصطلاح محدودیت

نام کلاس	مثالی برای ویژگی‌ها	تفسیر
دسته محدودیت	برچسب تفسیر	
محدودیت	برچسب	
عملیات منطقی	عملگر	مقادیر برای ویژگی عملگر /logicalAnd/ ، /logicalOr/ ، /logicalNot/ می‌باشند.
ارزیابی ویژگی	اجزای کلام شمار دستوری جنسیت دستوری	مقدار معینی چون /any/ برای تعیین این که، تعیین ویژگی اجباری می‌باشد، به کار می‌رود و در حال ممکن است مقدار /any/ مورد استفاده قرار بگیرد. نام کلاس برای این ویژگی مشخص نشده است.

ش-۲ مثال اصطلاح محدودیت

در مثال زیر از زبان فرانسه، نمونه دسته محدودیت، محدودیتی را برای مشخص ساختن اینکه اجزای کلام صفت در شمار دستوری و جنسیت دستوری برای یک دسته مقادیر تعریف شده، تغییر می‌کند. مقادیر قابل قبول برای جنسیت دستوری مذکر و مونث می‌باشند که برای نمونه جنسیت خنثی مورد قبول نمی‌باشد. مقادیر قابل قبول برای شمار دستوری، مفرد و جمع می‌باشند، که برای نمونه دوگانگی مورد قبول نمی‌باشد. اگر با دقت ملاحظه کنید نمونه ارزیابی ویژگی برای اجزای کلام به مقدار بخصوصی تنظیم شده است که آن صفت می‌باشد. ویژگی‌ها و مقادیر معتبر برای تثبیت ترکیبات مجاز، توسط حرف ربط جداساز منطقی ارائه می‌شوند. به شکل ش-۱^۱ مراجعه شود.

۱- برای این که فهم این شکل آسان‌تر شود، محتوی کادرهای سایه‌دار برای نمونه‌های مربوط به رده‌هایی که در بسته کنونی تعریف شده‌اند، مورد استفاده قرار می‌گیرند. محتوی کادرهای مربوط به نمونه‌های رده‌هایی که در بسته دیگر تعریف شده، سایه ندارند.



شکل ش-۱- مثال اصطلاح محدودیت

پیوست ص

(اطلاعاتی)

ارتباط با چارچوب نشانه‌گذاری واژه‌شناسی^۱ (TMF) و سیستم‌های بازنمایی مبتنی بر مفهوم

ص-۱ مقدمه

هدف این پیوست ایجاد برخی دستورالعمل‌ها برای چگونگی ارتباط یک فرهنگ لغت سازگار با LMF با سیستم‌های خارجی، مانند منابع TMF و سیستم‌های بازنمایی مبتنی بر مفهوم دیگری مانند هستی‌شناسی می‌باشد.

LMF بر ویژگی‌های زبان‌شناسی کلمات تمرکز می‌کند. ایجاد یک سیستم سازمانی با دانش پیچیده، هدف یک فرهنگ لغت نمی‌باشد.

ص-۲ تنظیمات

دو نوع منبع واژگانی مختلف وجود دارند:

ص-۲-۱ پیکربندی ۱: نمونه منبع واژگانی، تک زبانه می‌باشد، به طوری که فقط یک نمونه فرهنگ لغت با نمونه منبع واژگانی پیوند داده شده است. در این موقعیت، یک ارتباط خارجی به نمونه مفهوم متصل می‌شود.

ص-۲-۲ پیکربندی ۲: نمونه منبع واژگانی، چند زبانه می‌باشد، به طوری که بیش از یک نمونه فرهنگ لغت با نمونه منبع واژگانی پیوند داده می‌شود و نمونه‌های محور مفهوم به منظور ارتباط با نمونه‌های مفهوم مرتبط با زبان‌های مختلف ثبت می‌شوند. در این موقعیت، یک ارتباط خارجی به نمونه محور مفهوم متصل می‌شود.

همان طوری که در هسته مرکزی شرح داده شد، ویژگی منحصر به فرد می‌باشد، اما نمونه مفهوم (نمونه محور مفهوم، به طور بازگشتی) ممکن است نیاز به ارتباط با بیش از یک گره سیستم خارجی داشته باشد. بنابراین ارتباط توسط جفت-مقدار ارائه نمی‌شود بلکه توسط یک کلاس معینی ارائه می‌شود.

در یک تنظیمات تک زبانه، ارتباط خارجی توسط نمونه منبع خارجی تک زبانه، همان طوری که در مجموعه معنایی پردازش زبان طبیعی مشخص شده است، نمایش داده می‌شود. ویژگی‌های `/externalSystem/` و `/externalReference/` برای رجوع به نام‌های سیستم خارجی به طور بازگشتی و همچنین جهت تعیین گره مرتبط در این سیستم خارجی مفروض ایجاد می‌شوند.

در یک پیکره‌بندی چند زبانه، ارتباط خارجی توسط نمونه منبع خارجی چند زبانه، همان طوری که در بسته نشانه‌گذاری چند زبانه NLP مشخص شده است، نمایش داده می‌شود. ویژگی‌های `/externalSystem/` و `/externalReference/` برای ارجاع به نام‌های سیستم خارجی و همچنین جهت تعیین گره مرتبط در این سیستم خارجی مفروض ایجاد می‌شوند.

1-Termonological markup framework

پیوست ض

(اطلاعاتی)

تعریف نوع سند^۱ (DTD) برای LMF

ض-۱ عمومی

بخش زیر فقط برای اطلاع‌رسانی ایجاد شده است و هسته مرکزی و پیوست‌های آن را پوشش می‌دهد. اجزای XML برای DTD، از نمودارهای کلاس UML با نام یکسانی انتقال کد یافته‌اند. آرایش کلاس به عنوان دسته اجزای مهارتی^۲ اجرا می‌شود.

یک کاربر می‌تواند در مورد اختیار DTD یا شما برای اجرای LMF تصمیم بگیرد. می‌توان از ساختارهای XML که در استاندارد نمایش ساختار ویژگی (ISO24610-1) تعریف شده است، نیز استفاده کرد.

ض-۲ تعریف نوع سند برای LMF

```
<?xml version='1.0' encoding="UTF-8"?>
  <!-- LMF DTD -->
  <!--##### Core package-->
  <!ELEMENT LexicalResource (feat*, GlobalInformation, Lexicon+, SenseAxis*, TransferAxis*,
ContextAxis*)>
  <!ATTLIST LexicalResource
    dtdVersion CDATA #FIXED "16">
  <!ELEMENT GlobalInformation (feat*)>
  <!ELEMENT Lexicon (feat*, LexicalEntry+, SubcategorizationFrame*, SubcategorizationFrameSet*,
    SemanticPredicate*, Synset*, SynSemCorrespondence*, MorphologicalPattern*,
    MWEPattern*, ConstraintSet*)>
  <!ELEMENT LexicalEntry (feat*, Lemma, WordForm*, Stem*, ListOfComponents?, RelatedForm*,
    TransformCategory*, Sense*, SyntacticBehaviour*)>
  <!ATTLIST LexicalEntry
    id ID #IMPLIED
    morphologicalPatterns IDREFS #IMPLIED
    mwePattern IDREF #IMPLIED>
  <!ELEMENT Sense (feat*, Sense*, Equivalent*, Context*, SubjectField*,
    PredicativeRepresentation*, SenseExample*,
    Definition*, SenseRelation*, MonolingualExternalRef*)>
  <!ATTLIST Sense
    id ID #IMPLIED
    synset IDREF #IMPLIED>
  <!ELEMENT Definition (feat*, Statement*, TextRepresentation*)>
  <!ELEMENT Statement (feat*, TextRepresentation*)>
  <!ELEMENT TextRepresentation (feat*)>
  <!--##### Package for Morphology -->
  <!ELEMENT Lemma (feat*, FormRepresentation*)>
  <!ELEMENT WordForm (feat*, FormRepresentation*)>
  <!ELEMENT Stem (feat*, FormRepresentation*, GrammaticalFeatures*)>
  <!ELEMENT FormRepresentation (feat*)>
  <!ELEMENT RelatedForm (feat*, FormRepresentation*)>
```

1-Document Type Definition

2-Feat elements

```

<!ATTLIST RelatedForm
  targets IDREFS #IMPLIED>
<!ELEMENT ListOfComponents (feat*, Component+)>
<!ELEMENT Component (feat*)>
<!ATTLIST Component
  entry IDREF #REQUIRED>
  <!--##### Package for MRD -->
<!ELEMENT Equivalent (feat*, TextRepresentation*)>
<!ELEMENT Context (feat*, TextRepresentation*)>
<!ELEMENT SubjectField (feat*, SubjectField*)>
  <!--##### Package for Syntax -->
<!ELEMENT SyntacticBehaviour (feat*)>
<!ATTLIST SyntacticBehaviour
  id ID #IMPLIED
  senses IDREFS #IMPLIED
  subcategorizationFrames IDREFS #IMPLIED
  subcategorizationFrameSets IDREFS #IMPLIED>
<!ELEMENT SubcategorizationFrame (feat*, LexemeProperty?, SyntacticArgument*)>
<!ATTLIST SubcategorizationFrame
  id ID #IMPLIED
  inherit IDREFS #IMPLIED>
<!ELEMENT LexemeProperty (feat*)>
<!ELEMENT SyntacticArgument (feat*)>
<!ATTLIST SyntacticArgument
  id ID #IMPLIED
  target IDREF #IMPLIED>
<!ELEMENT SubcategorizationFrameSet (feat*, SynArgMap*)>
<!ATTLIST SubcategorizationFrameSet
  id ID #IMPLIED
  subcategorizationFrames IDREFS #IMPLIED
  inherit IDREFS #IMPLIED>
<!ELEMENT SynArgMap (feat*)>
<!ATTLIST SynArgMap
  arg1 IDREF #REQUIRED
  arg2 IDREF #REQUIRED>
  <!--##### Package for Semantics -->
<!ELEMENT PredicativeRepresentation (feat*)>
<!ATTLIST PredicativeRepresentation
  predicate IDREF #REQUIRED
  correspondences IDREFS #REQUIRED>
<!ELEMENT SemanticPredicate (feat*, Definition*, SemanticArgument*, PredicateRelation*)>
<!ATTLIST SemanticPredicate
  id ID #REQUIRED
  semanticTypes IDREFS #IMPLIED>
<!ELEMENT SemanticArgument (feat*, ArgumentRelation*)>
<!ATTLIST SemanticArgument
  id ID #IMPLIED
  semanticTypes IDREFS #IMPLIED>
<!ELEMENT ArgumentRelation (feat*)>
<!ATTLIST ArgumentRelation
  targets IDREFS #IMPLIED>
<!ELEMENT SynSemCorrespondence (feat*, SynSemArgMap*)>
<!ATTLIST SynSemCorrespondence
  id ID #REQUIRED>

```

```

<!ELEMENT SynSemArgMap (feat*)>
<!ELEMENT PredicateRelation (feat*)>
<!ATTLIST PredicateRelation
  targets          IDREFS #IMPLIED>
<!ELEMENT SenseExample (feat*)>
<!ATTLIST SenseExample
  id              ID #IMPLIED>
<!ELEMENT Synset (feat*, Definition*, SynsetRelation*, MonolingualExternalRef*)>
<!ATTLIST Synset
  id              ID #IMPLIED>
<!ELEMENT SynsetRelation (feat*)>
<!ATTLIST SynsetRelation
  targets          IDREFS #IMPLIED>
<!ELEMENT MonolingualExternalRef (feat*)>
<!ELEMENT SenseRelation (feat*)>
<!ATTLIST SenseRelation
  targets          IDREFS #REQUIRED>
  <!--##### Package for Multilingual notations -->
<!ELEMENT SenseAxis (feat*, SenseAxisRelation*, InterlingualExternalRef*)>
<!ATTLIST SenseAxis
  id              ID #IMPLIED
  senses          IDREFS #IMPLIED
  synsets        IDREFS #IMPLIED>
<!ELEMENT InterlingualExternalRef (feat*)>
<!ELEMENT SenseAxisRelation (feat*)>
<!ATTLIST SenseAxisRelation
  Targets          IDREFS #REQUIRED>
<!ELEMENT TransferAxis (feat*, TransferAxisRelation*, SourceTest*, TargetTest*)>
<!ATTLIST TransferAxis
  id              ID #IMPLIED
  syntacticBehaviours IDREFS #IMPLIED>
<!ELEMENT TransferAxisRelation (feat*)>
<!ATTLIST TransferAxisRelation
  targets          IDREFS #REQUIRED>
<!ELEMENT SourceTest (feat*)>
<!ATTLIST SourceTest
  syntacticBehaviours IDREFS #REQUIRED>
<!ELEMENT TargetTest (feat*)>
<!ATTLIST TargetTest
  syntacticBehaviours IDREFS #REQUIRED>
<!ELEMENT ContextAxis (feat*, ContextAxisRelation*)>
<!ATTLIST ContextAxis
  id              ID #IMPLIED
  examples        IDREFS #IMPLIED>
<!ELEMENT ContextAxisRelation (feat*)>
<!ATTLIST ContextAxisRelation
  targets          IDREFS #REQUIRED>
  <!--##### Package for morphological patterns -->
<!ELEMENT MorphologicalPattern (feat*, TransformSet*, TransformCategory*, Affix*, AffixSlot*,
AffixTemplate*)>
<!ATTLIST MorphologicalPattern
  id              ID #IMPLIED>
<!ELEMENT TransformSet (feat*, Process*, GrammaticalFeatures*, TransformCategory*)>
<!ELEMENT GrammaticalFeatures (feat*)>

```



```

<!ELEMENT Process (feat*, Condition*)>
<!ELEMENT Condition (feat*, GrammaticalFeatures*, TransformCategory*)>
<!ELEMENT Affix (feat*, AffixAllomorph*, GrammaticalFeatures*)>
<!ELEMENT AffixAllomorph (feat*, FormRepresentation*, Condition*)>
<!ELEMENT AffixTemplate (feat*, TemplateSlot*)>
<!ELEMENT TemplateSlot (feat*, Affix*)>
<!ATTLIST TemplateSlot
  affixslots  IDREFS #IMPLIED>
<!ELEMENT AffixSlot (feat*)>
<!ATTLIST AffixSlot
  affixs      IDREFS #IMPLIED>
<!ELEMENT TransformCategory (feat*)>
  <!--##### Package for MWE patterns -->
<!ELEMENT MWEPattern (feat*, MWENode*)>
<!ATTLIST MWEPattern
  id          ID #REQUIRED>
<!ELEMENT MWENode (feat*, MWEEdge*, MWELEX)>
<!ELEMENT MWEEdge (feat*, MWENode*)>
<!ELEMENT MWELEX (feat*)>
  <!--##### Package for Constraint expression -->
<!ELEMENT ConstraintSet (feat*, Constraint*)>
<!ELEMENT Constraint (feat*, LogicalOperation*)>
<!ATTLIST Constraint
  id          ID #IMPLIED>
<!ELEMENT LogicalOperation (feat*, AttributeValuation*, Constraint*)>
<!ATTLIST LogicalOperation
  Constraints IDREFS #IMPLIED>
<!ELEMENT AttributeValuation (feat*)>
  <!--##### for data category adornment: feat stands for feature-->
<!ELEMENT feat EMPTY>
  <!-- att=constant to be taken from the DataCategoryRegistry -->
  <!-- val=free string or constant to be taken from the DCR-->
<!ATTLIST feat
  att  CDATA #REQUIRED
  val  CDATA #REQUIRED>

```

پیوست ط
(اطلاعاتی)
کتابنامه

- [1] IETF BCP 47 (ed. Phillips, A., Davis, M.), currently (June, 2006) represented by RFC 4646” Tags for Identifying Languages ,“and RFC 4646” Matching of Language Tags“
- [2] RUMBAUGH ,J., JACOBSON ,I. and BOOCH ,G .*The unified modeling language reference manual* ,2nd ed.,Addison Wesley, 2004
- [3] CLIPS. 2000, ff. Clips: Browsing Semantic and Syntactic Data
http://www.ilc.cnr.it/clips/SYN_SEM/browsing.htm
- [4] MEL'ČUK ,I., CLAS ,A., POLGUÈRE ,A .*Introduction à la lexicologie explicative et combinatoire* .Duculot. Bruxelles, 1995 (Dictionnaire Explicatif et Combinatoire
- [5] CALZOLARI ,N., MC NAUGHT ,J. and ZAMPOLLI ,A .*Eagles, editors introduction* , 1996
<http://www.ilc.cnr.it/EAGLES96/edintro/edintro.html>
- [6] *Standards and best Practice for Multilingual Computational Lexicons*) ed. Calzolari N., Bertagna F., Lenci A., Monachini M.). MILE (The Multilingual ISLE Lexical Entry). ISLE CLWG Deliverable D2.2 and 3.2 Pisa, 2003
- [7] Wordnet 2.1 <http://wordnet.princeton.edu>
- [8] FELLBAUM ,C. A semantic network of English: the mother of all WordNets. In : *EuroWordNet: a multilingual database with lexical semantic networks*) ed. Vossen), Kluwer academic publishers, 1998
- [9] ANTONI-LAY ,M.-H., FRANCOPOULO ,G. and ZAYSSER ,L. A generic model for reusable lexicons: the GENELEX project ,*Literary and linguistic computing* ,9)1 ,(1994 ,pp. 47-54
- [10] CARLSON ,D .*Modeling XML applications with UML* ,Addison Wesley, 2001
- [11] MEL'ČUCK ,I .*Cours de morphologie générale* ۵ ,volumes, Presses de l'Université de Montréal, 1993-2000
- [12] FRADIN ,B .*Nouvelles approches en morphologie* ,Presses universitaires de France, 2003
- [13] MATTHEWS ,P.H .*Morphology* ,2nd ed, Cambridge University press, 1991
- [14] MATTHEWS ,P.H. *Inflectional morphology: a theoretical study based on aspects of latin verb conjugation*, Cambridge University press, 1972
- [15] STUMP ,G .*Inflectional morphology: a theory of paradigm structure* ,Cambridge University press, 2001
- [16] COMRIE ,B .*Language universals and linguistic typology* ,2nd ed, University of Chicago press, 1989
- [17] ARONOFF ,M .*Morphology by itself: stems and inflectional classes* ,MIT press, 1994
- [18] WRIGHT ,S.E .*A global data category registry for interoperable language resources* , LREC Lisbon, 2004
- [19] IDE ,N. and ROMARY ,L .*A registry of standard data categories for linguistic annotation* ,LREC Lisbon, 2004
- [20] FRANCOPOULO ,G., DECLERCK ,T., MONACHINI ,M. and ROMARY ,L .*The relevance of standards for research infrastructure* ,LREC Genoa, 2006
- [21] MEL'ČUK ,I .*Dictionnaire explicatif et combinatoire du français contemporain* ,4 volumes. Presses de l'Université de Montréal, 1984-1999

- [22] BLACHÈRE ,R. and GAUDEFROY-DEMOMBYNES ,M .*Grammaire de l'arabe classique* .Maisonneuve and Larose, 2004
- [23] LOMBARD ,D .*Introduction à l'indonésien* .Archipel, Paris, 1991
- [24] FRANCOPOULO ,G., BEL ,N., GEORGE ,M., CALZOLARI ,N., MONACHINI ,M., PET ,M. and SORIA ,C .*Lexical Markup Framework: ISO standard for Semantic Information in NLP Lexicons* .GLDV (Gesellschaft für linguistische Datenverarbeitung), Lexical-Semantic and ontological resources workshop, Tübingen, 2007
- [25] HUDAK ,T. Thai. In :*The world's major languages*) ed. Comrie, B.), Oxford University press, Oxford, 1987
- [26] HIGBIE ,J .*Let's speak Thai* .Orchid press, Bangkok, 2004
- [27] BROHÉE ,J.M. and RENAUD ,J .*Vocabulaire hongrois* .Ophrys, Paris, 2001
- [28] KHEMAKHEM ,A., GARGOURI ,B., ABDELWAHED ,A. and FRANCOPOULO ,G .*Modélisation des paradigmes de flexion des verbes arabes selon la norme LMF-ISO 24613* , TALN, Toulouse, 2007
- [29] RUPPENHOFER ,J., ELLSWORTH ,M., PETRUCK ,M., JOHNSON ,C. and SCHEFFCZYK ,J .*FrameNet II: Extended Theory and Practice* ,2006
<http://framenet.icsi.berkeley.edu/book/book.html>
- [30] PUSTEJOVSKY ,J .*The generative lexicon* .MIT press, 1995
- [31] ISO 16642:2003 ,*Computer applications in terminology — Terminological markup framework*
- [32] ISO 24610-1 ,*Language resource management — Features structures — Part 1: Feature structure representation*
- [33] ISO/IEC 10646:2003 ,*Information technology — Universal Multiple-Octet Coded Character Set (UCS)*