

نشریه پژوهش‌های زبان‌شناسی

سال هفتم، شماره دوم، شماره ترتیبی ۱۳، پاییز و زمستان ۱۳۹۴

تاریخ وصول: ۱۳۹۴/۶/۲۲

تاریخ اصلاحات: ۱۳۹۴/۱۲/۵

تاریخ پذیرش: ۱۳۹۵/۱/۳۰

صص ۱۱۴-۱۰۱

مشکلات انطباق دسته‌های هم‌معنای فارسی با دسته‌های هم‌معنای وردنت پرینستون

اکبر حسابی*

چکیده

در این پژوهش به بررسی مشکلات موجود در انتخاب معادل برای دسته‌های هم‌معنای (دسمعناهای) فارسی پرداخته شد. اهمیت این پژوهش به این لحاظ است که نتایج آن می‌تواند در طرح‌های آینده مربوط به تکمیل فارسی‌نت و ساخت شبکه‌های واژگانی دیگر یا تخصصی مورد استفاده قرار گیرد. با توجه به وجود سه دسته از مشکلات جهت انطباق وردنتها بر همدیگر شامل الف. مشکلات ناشی از سختی در تمایز معنایی در وردنت مبدأ ب. مشکلات ناشی از اصول حاکم بر وردنت مبدأ و منابع زبان مقصد و ج. مشکلات ناشی از تفاوت‌های ذاتی بین زبان مبدأ و مقصد، به تحلیل موارد یافت شده در فارسی‌نت پرداخته شده است. این پژوهش سعی بر پاسخ به سوالات زیر داشته است: مشکلات موجود بر سر راه انطباق دسمعناهای فارسی‌نت با وردنت چه می‌باشد؟ چه مشکلاتی بسامد بیشتری دارند؟ آیا تفاوتی بین مشکلات موجود در انطباق فارسی‌نت با وردنت و انطباق در سایر وردنت‌های موجود وجود دارد؟ حجم نمونه مورد بررسی که به شکل تصادفی انتخاب شده‌اند ۱۵۵۲ دسته می‌باشد که شامل دسمعناهایی است که اولین عضو آنها با حروف <ع، غ، ف و ق> شروع شده‌اند. مشکلات یافته شده به هشت دسته تقسیم گردیدند و بسامد و درصد هر مشکل ارائه گردید و برای حل برخی از این مشکلات راه‌هایی پیشنهاد شد. این راه‌ها می‌تواند در فرایند تکمیل فارسی‌نت و یا پیش از ساخت شبکه‌های واژگانی دیگر یا شبکه‌های واژگانی تخصصی مد نظر قرار گیرد.

واژه‌های کلیدی: شبکه‌واژگانی، وردنت پرینستون، فارسی‌نت، انطباق اولیه، انطباق ثانویه، زبان‌شناسی رایانشی.

۱- مقدمه

با توجه به حجم عظیم اطلاعات موجود در جهان امروز و محدودیت انسان در بررسی آن‌ها، پردازش زبان طبیعی از نیازهای غیر قابل انکار جوامع امروز می‌باشد؛ تا بدین طریق بتوانند سریعتر و به شکل بهینه به اطلاعات مورد نیاز خود دست یابند. اما پردازش زبان با این اهداف همیشه به راحتی ممکن نبوده و خود نیازمند ابزارهایی است تا نتایج حاصله دقیقتر و در واقع همان نتایجی باشد که کاربر مد نظر دارد. اگر به جستجوهای واژه‌ای در موتورهای جستجوی معمول توجه گردد متوجه حجم عظیمی از نتایج نامربوط در جستجو می‌شویم که ناشی از پردازش به لحاظ صورت یا فرم در موتورهای جستجو است.

جهت حل این مشکل و مسائل دیگر در پردازش زبان طبیعی خصوصاً مشکل ابهام در واژه‌های مورد جستجو، دادگانی که توجه افراد بسیاری را در حوزه پردازش زبان طبیعی به خود جلب نموده است شبکه‌واژگانی^۱ است. این دادگان متشکل از واحدهایی به نام دسته‌هم‌معنا^۲ (دسمعنا) می‌باشد که عبارت از واژه‌های قابل جایگزینی در بافتهای خاص به جای یکدیگر است (فلباوم، ۱۹۸۸). دسته‌های هم‌معنا خود با روابط واژگانی از قبیل شمولیت، جزء‌واژگی، بخش‌واژگی، عضو‌واژگی و غیره پس از بررسی توسط آزمونهایی که به این منظور طراحی شده‌اند، به یکدیگر مرتبط شده و شبکه‌ای از واژه‌ها ایجاد می‌شود که تصور بر این است که با توجه به یافته‌های روان‌شناسی و عصب‌شناسی زبان شبیه آن چیزی است که انسانها در مغز خود دارند (حسابی و همکاران، ۱۳۸۸).

روابط موجود در شبکه‌های واژگانی به دو دسته روابط درون‌زبانی و روابط میان‌زبانی قابل تقسیم است. روابط درون‌زبانی میان دسته‌های هم‌معنا باعث ایجاد شبکه‌واژگانی می‌گردد و روابط میان‌زبانی شبکه‌ای را به شبکه‌دیگر انطباق می‌دهد. روابط درون‌زبانی بین دسته‌های هم‌معنا قابل تقسیم به روابط میان‌اسامی، روابط میان‌افعال، میان‌صفات و میان‌قیود می‌باشد. این روابط در شبکه‌واژگانی پرینستون که تمایز مشخصی بین مقوله‌های کلام لحاظ می‌گردید بیشتر برجسته بود ولی در شبکه‌های واژگانی بعدی مانند شبکه‌واژگانی اروپا و بالکان روابط بین مقوله‌ای نیز لحاظ گردیدند.

روابط درون‌زبانی در پرینستون ۱/۵ عبارت‌اند از مترادف^۳ (بین اعضای دسته‌های هم‌معنا)، تضاد^۴، شمول^۵، جزء‌واژگی^۶، استلزام^۷، نوعی‌از^۸، باعث شدن^۹، همچنین ببینید^{۱۰}، مشتق از^{۱۱}، مشخصه^{۱۲}، صفت مرتبط^{۱۳}، شبیه به^{۱۴} و صورت مصدری^{۱۵} (وسن، ۲۰۰۲: ۱۵). این روابط بجز رابطه‌ی مترادف که بین اعضای یک دسته‌هم‌معنا وجود دارد در جدول ۱ قابل مشاهده می‌باشند. در ادامه ضمن توضیح این روابط به ارائه‌ی مثال برای هر رابطه می‌پردازیم.

¹ wordnet

² Synset or synonym set

³ synonymy

⁴ antonymy

⁵ hyponymy

⁶ meronymy

⁷ entailment

⁸ troponym

⁹ cause

¹⁰ Also see

¹¹ Derived from

¹² attribute

¹³ Relational adjective

¹⁴ Similar to

¹⁵ participle

جدول (۱) روابط معنایی در شبکه‌واژگانی پرینستون ۱/۵ (برگرفته از وسن، ۲۰۰۲: ۱۵)

Table 1: WordNet1.5 Relations

Relation	PoS linked	Example	EWN
ANTONYMY	noun/noun; verb/verb; adjective/adjective	man/woman; enter/exit; beautiful/ugly	yes
HYPONIMY	noun/noun	slicer/knife	yes
MERONYMY	noun/noun	head/nose	yes
ENTAILMENT	verb/verb	buy/pay	SUBEVENT or CAUSE
TROPONYM	verb/verb	walk/move	HYPONIMY
CAUSE	verb/verb	kill/die	yes
ALSO SEE	verb/adjective		no
DERIVED FROM	adjective/adverb	beautiful/beautifully	yes
ANTONYM	noun/noun; verb/verb	heavy/light	yes
ATTRIBUTE	noun/adjective	size/small	XPOS HYPONYM
RELATIONAL ADJ	adjective/noun	atomic/ atomic bomb	PERTAINS TO
SIMILAR TO	adjective/adjective	ponderous/heavy	no
PARTICIPLE	adjective/verb	elapsed/ elapse	no

همان‌طور که از جدول (۱) برمی‌آید رابطه تضاد بین دسته‌های هم‌معنای متعلق به سه مقوله اسامی، افعال و صفات وجود دارد که مثال‌های مرد/زن، وارد شدن/خارج شدن و زیبا/زشت دارای این رابطه می‌باشند (ستون چهارم اشاره به روابطی دارد که در یورونت حفظ گردیده یا تغییر نام پیدا کرده‌اند). رابطه دیگر شمولیت است که بین اسامی برقرار است و نمونه آن در انگلیسی رابطه slicer/knife است و یا رابطه بین زنبق و گل در فارسی که زنبق نوعی از گل است. رابطه سوم رابطه جزء‌واژگی است که بین اسامی برقرار است و یک اسم جزئی از اسم دیگر است مانند رابطه بین "بینی" و "سر" که بینی جزئی از سر است. این رابطه به شکل کلی در نظر گرفته شده و به اقسام مختلف تقسیم نگردیده است. رابطه بعدی رابطه استلزام می‌باشد که بین افعال برقرار است مانند رابطه خرید/فروش که یکی مستلزم دیگری است. رابطه دیگر رابطه باعث شدن است که بین افعال برقرار است، مانند رابطه بین کشتن و مردن. روابط دیگر نیز همان‌طور که از جدول (۱) برمی‌آید بین مقوله‌های مختلف کلام برقرار است.

موفقیت این دادگان طی دهه‌های گذشته چنان قابل توجه بوده است که شاهد این موضوع می‌باشیم که برای اغلب زبان‌های دنیا طرح‌های ایجاد آن به نتیجه رسیده است یا در حال انجام است. حتی برای زبان‌هایی با گویشوران محدود نیز این دادگان در حال ساخته شدن است (اوراو و همکاران ۲۰۱۴؛ فلبام و وسن ۲۰۱۲). با مراجعه به تارنمای انجمن جهانی شبکه‌واژگانی^۱ می‌توان زبان‌هایی که برای آن‌ها شبکه‌واژگانی طراحی و ایجاد شده است مشاهده نمود. برای زبان فارسی نیز مانند سایر زبان‌ها طی طرحی شبکه‌واژگانی ساخته شد که در ادامه به آن پرداخته شده است.

۲- ساخت شبکه‌واژگانی بان فارسی (فارس‌نت)

ساخت شبکه‌واژگانی برای زبان فارسی به سال ۲۰۰۷ و طرح کیوان و همکاران (۲۰۰۷) در دانشگاه پرینستون برمی‌گردد که به نتیجه عملی منجر نگردید. بعد از این طرح طی پژوهشی فامیان و آقاجانی (۲۰۰۷) طرحی را برای ساخت شبکه‌واژگانی صفات ارائه دادند و سپس روحی‌زاده و همکاران (۲۰۰۸)، منصور (۱۳۸۷) و منصور و

همکاران (۲۰۱۱) اقدام به ارائه طرح برای شبکه‌واژگانی افعال نموده و در انتها حسابی (۲۰۰۹) طرح ایجاد شبکه‌واژگانی اسامی را ارائه نمود. اما آنچه جای خالی خود را نشان می‌داد ایجاد شبکه‌واژگانی زبان فارسی در واقعیت بود. این امر سرانجام طی طرحی که گزارش آن (شمس‌فرد و همکاران، ۲۰۱۰) در کنفرانس بین‌المللی شبکه‌واژگانی ارائه شد محقق گردید و با استفاده از نتایج طرح‌های گذشته و همکاری گروهی متخصصان رایانه و زبان‌شناسان فارسی‌ت‌پا به عرصه وجود نهاد. طی این طرح ۱۰۰۰۰ دسته‌هم‌معنا (دسمعنا) در دادگان فارسی قرار داده شد و علاوه بر روابط درون‌زبانی با استفاده از روابط میان‌زبانی با دسمعناهای وردنت پرینستون (از این به بعد آن را وردنت می‌خوانیم) انطباق داده شد تا به این وسیله با تمامی وردنت‌های جهان متصل گردند. مروگر برخط فارسی‌نت^۱ از طریق وب قابل مشاهده است. اخیراً اطلاعات مربوط به شبکه‌واژگانی دیگری نیز برای زبان فارسی بر روی تارنمای جهانی شبکه‌واژگانی قابل مشاهده است. این پژوهش به بررسی مشکلات موجود در انطباق دسمعناهای فارسی‌نت با وردنت و دسته‌بندی آن‌ها پرداخته است و به دنبال پاسخگویی به سوالات زیر بوده است:

۱- چه مشکلاتی در انطباق دسمعناها وجود دارد؟

۲- بسامد کدام یک از مشکلات بیشتر است؟

۳- آیا شباهتی بین مشکلات موجود در انطباق فارسی‌نت با وردنت و انطباق دیگر شبکه‌های واژگانی و وردنت وجود دارد؟

در ادامه پس از مرور فرایند ساخت شبکه‌واژگانی فارسی به روش پژوهش و بررسی داده‌ها و سپس تحلیل آن‌ها پرداخته شده است.

۳- چهارچوب نظری

انطباق در شبکه‌های واژگانی عبارت از یافتن معادلی برای دسمعناهای شبکه‌واژگانی مبدأ در زبان مقصد می‌باشد که برای آن زبان شبکه‌واژگانی در حال ساخته شدن است و یا برعکس یافتن معادلی در شبکه‌واژگانی مبدأ برای دسمعنا در شبکه‌واژگانی مقصد است. این امر با توجه به روش‌های ساخت شبکه‌های واژگانی به شکل‌های متفاوتی مدیریت می‌شود. سه روش برای ساخت شبکه‌های واژگانی وجود دارد که عبارتند از تلفیق (ادغام)، گسترش (توسیع) و ساخت بالا به پایین. (حسابی، ۲۰۰۹).

در روش تلفیق، ساخت شبکه‌واژگانی به صورت مستقل بدون استفاده از شبکه‌های واژگانی دیگر انجام می‌پذیرد مانند آنچه در شبکه‌واژگانی وردنت پرینستون به عنوان اولین شبکه‌واژگانی رخ داد. فامیان روش ساخت شبکه‌های واژگانی هلندی، سوئدی و آلمانی را از این دست می‌داند (فامیان، ۱۳۸۶).

در این روش بعد از ساخت، انطباق دسمعناها با سایر شبکه‌های واژگانی در صورت نیاز انجام می‌پذیرد. این نوع انطباق را انطباق ثانویه می‌نامیم.

در روش گسترش، دسمعناهای وردنت پرینستون ترجمه شده (معادلایی یا انطباق اولیه) و روابط معنایی بین آن‌ها نیز از این وردنت گرفته می‌شود و سپس شبکه‌واژگانی مقصد بر این اساس شکل می‌گیرد. نتیجه شبکه‌واژگانی ای است که

^۱ <http://nlp.sbu.ac.ir/farsnet>

سوگیری زیادی به وردنت دارد. وسن (۲۰۰۲) شبکه‌های واژگانی فرانسوی و اسپانیایی و بتی ولی و همکاران (۲۰۰۰) ایتالیایی را از این نوع شبکه‌های واژگانی می‌دانند.

به گفته حسابی (۲۰۰۹) در روش ساخت بالا به پایین مفاهیم پایه (دسمعناهای) مستقل از زبان که با بررسی شبکه‌های زبانی موجود به دست آمده و در همه آن‌ها مشترک بوده، در زبان مقصد معادل‌سازی شده و سپس واژگان پربسامد زبان با استفاده از منابع موجود (مانند پیکره‌های زبانی) شناسایی و واژگانی که دارای جایگاه بالایی در سلسله مراتب معنایی بوده و روابط معنایی زیادی با مفاهیم دیگر دارند شناسایی می‌گردند و با استفاده از آن‌ها دسمعناهای ویژه زبان ساخته می‌شوند. با بررسی و مقایسه دسمعناهای ویژه زبان با مفاهیم پایه مستقل از زبان و حذف موارد یکسان، هسته شبکه واژگانی شکل می‌گیرد. سپس با استفاده از منابع زبانی اقدام به گسترش این هسته می‌گردد. در این روش معادلیابی اولیه و ثانویه انجام می‌پذیرد که آن را متفاوت از دو روش دیگر می‌نماید.

با توجه به اینکه در ساخت فارسی‌نت از روش ساخت بالا به پایین استفاده شده است (حسابی و همکاران، ۱۳۸۸) انطباق اولیه و ثانویه هر دو وجود دارد. نکته دیگر در انطباق شبکه‌های واژگانی بر یکدیگر مشکلات انطباق است. کریستیا و همکاران (۲۰۰۴) به مشکلات موجود بر سر راه انطباق دسمعناهای وردنت با شبکه‌های واژگانی رومانیایی پرداخته‌اند و به وجود سه دسته از مشکلات جهت انطباق وردنت‌ها بر همدیگر اشاره نموده‌اند:

۱. مشکلات ناشی از سختی تمایز معنایی در وردنت مبدأ (وردنت پرنستون) شامل:
 - ا. دسمعناهای یکسان یا دارای اشتراک
 - ب. دارای تعاریف مشابه یا غیر متمایز
 - ج. دارای تعاریف مکرر
 - د. مثالهای غیر ممیز
 - ه. فقدان شامل‌های ممیز^۱ یا داشتن شامل‌های یکسان
 - و. فقدان زیرشمول‌های ممیز
 - ز. اشتراک واژگانی بین دسمعناها و شامل‌هایشان
 - ح. تضاد بین اطلاعات مربوط به شامل‌ها و مثال‌ها
۲. مشکلات ناشی از اصول حاکم بر وردنت مبدأ و منابع زبان مقصد
 - ا. همه معانی در وردنت مبدأ یکسانند در حالی که منابع زبان مقصد حاوی معانی اصلی و مشتقی می‌باشند که ویژه‌تر هستند.

ب. منابع زبان مقصد برای اعضای دسمعناها دسته‌بندی‌های کلی‌تری (و نه با دقت) نسبت به وردنت مبدأ دارند

^۱ شامل‌های غیر ممیز عبارتند از دسمعناهایی یکسان که با توجه به آن‌ها نمی‌توان بین دسمعناهای زیرشمول آن‌ها تمایز قائل شد. برای مثال کریستیا و همکاران (۲۰۰۴): (۱۳۳) به معانی (۷) و (۱۰) snap اشاره می‌نمایند که دارای دسمعناهای یکسان move به عنوان شامل می‌باشند و بنابراین نمی‌توان با استفاده از این رابطه شمولیت بین معانی آن‌ها تمایز قایل شد.

snap(7) -- (move with a snapping sound; "bullets snapped past us")

snap(10), click, flick -- (cause to make a snapping sound; "snap your fingers")

move -- (move so as to change position, perform a nontranslational motion; "He moved his hand slightly to the right").

- ج. همپوشی^۱ معنایی نظام‌مند در منابع زبانی مقصد
- د. معانی مجازی جدید که در زبان مقصد وجود ندارد
- ه. عدم وجود اصطلاحات در منابع زبان مقصد
- و. کلمات یا عبارات بسیار زشت که در منابع زبان مقصد وجود ندارند
- ز. عبارتهای ساختگی در وردنت مبدأ
۳. مشکلات ناشی از تفاوت‌های ذاتی بین زبان مبدأ و مقصد
- أ. معانی‌ای در زبان مبدأ که در زبان مقصد وجود ندارند
- ب. معانی‌ای در زبان مقصد که در زبان مبدأ وجود ندارند
- ج. معانی‌ای که در زبان مبدأ با یک کلمه بیان می‌شوند ولی در زبان مقصد با معادل‌های مرکب آورده می‌شوند.

در زمان طراحی و ایجاد شبکه‌واژگانی اروپا (وسن، ۲۰۰۲) جهت انطباق دسمعناهای مبدأ با دسمعناهای وردنت پرینستون چندین رابطه معرفی گردید که عبارت بودند از: معادل-مترادف (EQ_SYNONYM) که در موارد تطابق یک به یک مورد استفاده قرار می‌گرفت و مهم‌ترین رابطه بود و روابط پیچیده‌ای مانند معادل-نزدیک به مترادف (EQ_NEAR_SYNONYM) که مواقعی که یک دسمعنا مبدأ با چندین دسمعنا وردنت پرینستون تطابق داشته یا یک دسمعنا پرینستون با چندین دسمعنا مبدأ تطابق داشتند مورد استفاده قرار می‌گرفت؛ معادل-دارای-شامل (EQ_HAS_HYPERONYM) وقتی دسمعنا خاصتر از دسمعنا پرینستون بود یا خلاء واژگانی وجود داشت؛ معادل-دارای-زیرشمول (EQ_HAS_HYPONYM) که اشاره به دسمعنا مبدأئی داشت که تنها می‌توانست به دسمعنا پرینستون وصل گردد که به لحاظ معنایی خاص‌تر بود و تعدادی دیگر از روابط پیچیده مانند EQ_HAS_HOLONYM, EQ_HAS_MERONYM, EQ_INVOLVED, EQ_ROLE and EQ_CO_ROLE این روابط برای ایجاد معادل برای مواردی بود که معانی دقیقاً منطبق بر همدیگر نبودند.

۴- روش

در این پژوهش سعی گردید تا موارد خاص یافت شده در انطباق دسمعناهای فارسی با وردنت مورد بررسی قرار گیرند و برای آن‌ها دسته‌بندی‌ای ارائه گردد. منظور از موارد خاص دسمعناهایی هستند که ارائه معادل برای آن‌ها با توجه به منابع موجود مشکل بوده است؛ به عبارت دیگر، معادل آن‌ها جهت انطباق با دسمعناهای وردنت در منابع موجود نداشته است یا سازندگان با توجه به دانش زبانی و منابع نتوانسته‌اند برایشان معادلی پیدا نمایند. در این پژوهش معادل‌های ارائه شده برای دسمعناهایی که با کلماتی که با حروف <ع>، <غ>، <ف> و <ق> شروع شده‌اند، به کمک بررسی اعضای دسمعنا، تعریف آن و مثال‌های مربوط به کاربرد کلمات مورد بررسی قرار گرفت. انتخاب دسمعناهای فوق به صورت تصادفی بوده است. برای جلوگیری از همپوشانی، هر دسمعنا با توجه به کلمه‌اول موجود در آن مورد بررسی قرار گرفت.

^۱ clashing

با توجه به اینکه جهت حرکت برای بررسی دسمعناها از سمت دسمعنای فارسی با توجه به تعریف و مثال یا مثال‌هایش به سمت دسمعنای پیشنهادی انگلیسی بود، مشکلات به چند دسته کلی تقسیم گردید.

۵- نتایج

در زیر به تقسیم‌بندی و ذکر نمونه‌هایی از این موارد پرداخته شد و بسامد رخداد هر کدام مشخص گردید. کلیه نمونه‌ها از تارنمای فارسی^۱ گرفته شده است و مشکلات یافت شده به هشت دسته تقسیم گردید.

۱- در مواردی برای دسمعنای فارسی هیچ معادل انگلیسی موجود نیست یعنی این موارد خلاء واژگانی^۲ می‌باشند (مشکل ۳-ب در قسمت چهارچوب نظری). بعضی از این نوع خلاءها مانند آنچه در شبکه‌واژگانی اروپا انجام گرفت با استفاده از رابطه معادل-دارای-شامل به دسمعناهای شبکه‌مبدأ متصل گردیدند. این خلاءها از نوع خلاءهای کاربردشناختی^۳ بودند. هفتاد و پنج مورد از این نوع خلاء واژگانی یافت شد. اطلاعات اعلامی نظیر اسم افراد، شهرها و روستاها، رودها، مکان‌ها، اسامی غذاها و خوراکی‌ها (عدس پلو، قورمه سبزی، فسنجان، قابلی، قطاب، قیسی، قره قروت و...)، بعضی مشاغل (حوزوی، قاری، قالی باف)، بعضی اعمال (قرتی بازی، فتوا دادن، غسل کردن، قمه زنی) بعضی اسامی (فطریه، عده: مدت زمانی که زن بعد از طلاق نباید ازدواج کند، قنوت، قانون: نام ساز، قبضه: ریش) بعضی اشیا (علامت: در عزاداری جلو دسته حرکت می‌دهند) همه و همه از این دست هستند. به علاوه، مدخل‌هایی که مربوط به مراسم، سنن و مفاهیم ملی و مذهبی هستند هم جز این دسته محسوب می‌شوند. بررسی کاملتری از خلاءهای واژگانی در فارسی را در حسابی (۱۳۹۴) می‌توان یافت. نکته مسئله آفرین این است که برای بعضی از آن‌ها معادل‌هایی آورده شده است که خالی از اشکال نمی‌باشد. به مثال زیر توجه نمایید که از تارنمای فارسی گرفته شده است و برای آن معادل نامناسب ارایه گردیده است:

فاتحه چی (Adjective)

ویژگی آن که در مجالس ترحیم، هنگام ورود اشخاص، فاتحه بر زبان می‌آورد و به دست حضار گلاب می‌ریزد یا جزوات قرآن برای قرائت به حضار می‌دهد

"پیرمرد فاتحه چی روی دستمان گلاب ریخت و برایمان قرآن آورد"

<p>WordNet , prayer , supplication the act of communicating with a deity (especially as a petition or in adoration or contrition or thanksgiving); "the priest sank to his knees in prayer"</p>

در مورد فوق واژه فارسی واژه‌ای فرهنگی است که ریشه در مذهب دارد. این واژه به لحاظ معنایی با دسمعنای انگلیسی معادل نیست و به نظر می‌رسد نمونه‌ای از خلاء واژگانی فرهنگی در انگلیسی باشد که قابل انطباق با دسمعنایی

^۱ <http://nlp.sbu.ac.ir/farsnet>

^۲ Lexical gap

^۳ در شبکه اروپا خلاءها به دو نوع فرهنگی (مفهومی که در فرهنگ انگلیسی یا پرنستون موجود نیست) و کاربردشناختی (ناشی از تفاوت‌های واژگانی شدگی بین دو زبان است؛ به عبارت دیگر، مفهوم وجود دارد، اما با یک صورت واژگانی واحد بیان نمی‌گردد) تقسیم شدند.

در انگلیسی نمی‌باشد مگر اینکه قالب یا طرحواره مربوطه به انگلیسی وارد شده و سپس دسمعنا قرض گرفته شود. همچنین دسمعناهای غسل - غسل ارتماسی - عرصه - قبله - قبله یاب نیز برای انطباق دچار همین مشکل هستند.

۲- مشکل دیگر مواردی است که چند مدخل مجزا برای یک مفهوم واحد وجود دارد به عبارت دیگر گستره معنایی معادل انگلیسی وسیعتر از مفهوم دسمعنا فارسی است. این مشکلی است که تحت شماره ۲-آ بیان گردیده است و در شبکه اروپا با رابطه معادل-دارای-شامل انطباق یافته است. برای روشن شدن این قضیه به مثال زیر توجه فرمائید:

عمل آمدن، فرورده شدن، پرورده شدن، پرورش یافتن (Verb)

درست و آماده شدن ماده‌ای غذایی مانند ترشی، مربا و امثال آن

"می‌دانی مربای توت فرنگی عمل می‌آید؟"

WordNet
, cure
prepare by drying, salting, or chemical processing in order to preserve; "cure meats"; "cure pickles"; "cure hay"

دسمعنا انگلیسی فعل که در این مدخل مشاهده می‌کنیم به معنای "انداختن موادی مثل ترشی یا گوشت را دودی کردن و..." می‌باشد و مثال هم دال بر همین موضوع است. مدخل زیر نمونه دیگری از این موارد است:

عمل آمدن (Verb)

ساخته یا آماده شدن و به شکل دلخواه در آمدن

"چند ماه لازم است بگذرد تا ترشی به خوبی عمل بیاید"

WordNet

ظاهراً این مورد هم چیزی جدای از مدخل بالا نیست، حال آنکه دو مدخل متفاوت به آن‌ها اختصاص داده شده است. در مورد مدخل دوم، همان‌طور که مشاهده می‌شود، معادل انگلیسی هم پیشنهاد نشده است که شاید با رفع مشکل شرح داده شده (ادغام دو مدخل) این مسئله هم حل شود: اگر دو مدخل جدا هستند پس معانی آن‌ها باید تفاوت‌هایی داشته باشد و این موضوع در مثال هم منعکس شود و اگر این‌طور نیست بهتر است دو مدخل یکی شوند.

این مشکل ناشی از تفاوت در گستره معنایی واژه‌های زبان‌های مبدأ و مقصد می‌باشد. در فارسی هر مفهوم با استفاده از واژه‌ای خاص بیان می‌گردد که ناشی از استفاده تخصصی‌تر از واژه‌ها می‌باشد. موارد دیگر که این تفاوت گستره در آن‌ها مشاهده شده است عبارتند از: دسمعناهای *عائله*، *نانخور* - *عاریتی* - *عبور ممنوع* - *عصاکش* - *قاضی القضاة* - *قبا* که در مجموع ۸ مورد را به خود اختصاص می‌داد.

۳- مشکل جدیدی که قابل افزودن به مشکلات یافت شده در بررسی انطباق شبکه‌واژگانی رومانایی با وردنت است عبارت از بعضی از موارد است که دسمعنا فارسی معادل انگلیسی دارد (یعنی خلاء واژگانی نیست) ولی این معادل در وردنت انگلیسی ثبت نشده است. به نمونه زیر برگرفته از وبگاه فارس‌نت توجه کنید:

غوره (Noun)

میوه نارس انگور، دارای رنگ سبز و طعم ترش

"غوره انگورهای بی دانه را برای آبگیری استفاده می‌کنند"

WordNet , sour grapes disparagement of something that is unattainable

معادل انگلیسی این واژه sour grapes می‌باشد ولی در وردنت انگلیسی sour grapes در معنایی اصطلاحی به کار رفته است و با وجود موجود بودن این معادل در انگلیسی در وردنت ثبت نشده است. این مسئله باعث شده یا معادلی برای این موارد در فارس‌نت ذکر نگردد یا معادل ذکر شده صحیح نباشد. دسمعناهای فازمتر، فازسج - فوتبال دستی از دیگر موارد این نوع خلاء در پرنستون است. در مجموع ۱۹ مورد از این مشکل در داده‌های بررسی شده یافت شد.

۴- مشکل مشاهده شده دیگر منطبق با مشکل شماره ۲-ج در چهارچوب نظری است. در برخی موارد، دسمعنا انگلیسی پیشنهادی در فارس‌نت فاقد اشکال است اما دسمعنا(های) دیگری هم وجود دارند که قابل انطباق با آن هستند. این موارد در یورونت با عنوان معادل-نزدیک به مترادف با دسمعناهای پرنستون انطباق یافتند. در چنین مواردی جهت تکمیل نقص موجود، ضمن حفظ دسمعناهای انگلیسی ارائه شده در فارس‌نت، دسمعنا(های) جدید هم باید اضافه شوند. به نمونه‌های که برگرفته از وبگاه فارس‌نت است توجه نمایید:

عاشق بودن (Verb)

علاقه زیاد به چیزی داشتن

"او عاشق بستنی است."

WordNet , love , enjoy get pleasure from; "I love cooking"
--

همان‌طور که پیداست دسمعنا انگلیسی موجود در فارس‌نت درست است ولی مورد زیر را هم می‌توان به آن افزود:

love – (have a great affection or liking for; "I love French food"; "She loves her boss and works hard for him")

یک نمونه دیگر از حرف "ق":

قدر دانستن، قدرشناسی کردن (Verb)

ارزش چیزی را دانستن

"قدر مادرت را بدان."

WordNet
, acknowledge , recognize , recognise
express obligation, thanks, or gratitude for; "We must acknowledge the kindness she showed towards us"

دسمعنای پیشنهادی فارس‌نت درست است ولی دسمعنای appreciate هم قابل افزودن است:

appreciate – (recognize with gratitude; be grateful for)

مدخل‌های "فاش کردن، هویدا کردن، افشا کردن، آشکار کردن، نمایان ساختن"، "فتنه انگیز، فتنه گر"، "فتوت، جوانمردی"، "فداکاری، ایثار"، "قابل استناد، مستدل..."، "قابل دسترس"، "قابل چاپ"، "عادلانہ"، "عاشقانه، عاشق وار..." هم از این دست هستند. در مجموع ۱۱ مورد از این اشکال یافت گردید.

۵- مشکل دیگر مطابق با مشکل ۳-ج مطرح شده در چهارچوب نظری است. دلیل بروز چنین اشکالاتی تفاوت در گستره معنایی واژگان دو زبان است: واژه‌ای در فارسی طیف وسیعی از معانی مرتبط را در بر می‌گیرد ولی برای اشاره به چنین طیفی در انگلیسی بسته به کلمات همایند^۱ آن باید از واژگان متفاوتی استفاده کرد. این وضعیت عدم تطابق میان کلمات دو زبان را *divergent polysemy* می‌گویند. در یورونت از رابطه معادل-نزدیک به مترادف استفاده گردیده است. به مثال زیر برگرفته از وبگاه فارس‌نت توجه کنید:

غلاف (Noun)

پوشش

"علی شمشیر را در غلاف فرو کرد."

WordNet
, shell , case , casing
the housing or outer covering of something; "the clock has a walnut case"

در فارسی برای چاقو، شمشیر، گیاه... از کلمه "غلاف" استفاده می‌شود، حال آنکه در انگلیسی کلمات متفاوتی به کار می‌رود:

sheath – (a protective covering for a knife or sword)

scabbard – (a sheath for a sword or dagger or bayonet)

pod, cod, seedcase – (the vessel that contains the seeds of a plant (not the seeds themselves))

¹ collocations

۶- مشکل دیگر ناشی از تقسیم بندی ویژه زبان فارسی است که منجر به این مسئله می‌شود که دسمعناهای معادل آن‌ها در وردنت پرینستون یافت نشود مانند تقسیم علوم به نقلی و عقلی و ایجاد دو دسمعنا *علوم عقلی* و *علوم نقلی*. این مشکل را می‌توان با مشکل ۳-آ در چهارچوب نظری یکسان دانست.

۷- مشکل جدید در انطباق مربوط به مواردی است که دسمعنایی در وردنت مبدأ وجود ندارد و برای انطباق آن از واژه‌ی متضاد بهره‌گیری می‌شود مانند دسمعنا *غیرعشائری* که با استفاده از دسمعنا *urban* انطباق داده شده است.

۸- مشکل دیگر تفاوت در بیان یک مفهوم با مقوله‌های کلامی متفاوت است. بسیاری از صفات در انگلیسی با اسامی بیان می‌شود. مانند کلمات *steel* و *iron* یا *silver* که در فارسی هم با اسم و هم صفت بیان می‌گردند و تمایز دقیق بین مقوله‌ها منجر به مشکل انطباق می‌گردد. هرچند در وردنت برای *silver* هم مقوله اسم و هم صفت لحاظ شده ولی برای *steel* و *iron* اینچنین نیست و این عدم یکدستی منجر به مشکل انطباق مقوله صفت گردیده است. در یورونت با استفاده از رابطه معادل-نزدیک به مترادف این موارد انطباق داده شده‌اند.

جدول ۲ حاوی بسامد و درصد دسمعناهایی است که برای انطباق آن‌ها مشکلات فوق‌الذکر مشاهده گردید.

جدول ۲- بسامد و درصد مشکل‌های یافت شده در نمونه

مشکل	اول	دوم	سوم	چهارم	پنجم	ششم	هفتم	هشتم	کل
بسامد	۷۵	۸	۱۹	۱	۲	۲	۱	۱	۱۰۹
درصد مشکل در مجموع مشکل‌ها	۶۹	۷,۳	۱۷,۴	۰,۹	۱,۸	۱,۸	۰,۹	۰,۹	٪۱۰۰
درصد مشکل نسبت به کل دسمعناهای نمونه (۱۵۵۲ دسمعنا)	۴,۴۵	۰,۴۷	۱,۱۲	۰,۰۶	۰,۱۲	۰,۱۲	۰,۰۶	۰,۰۶	٪۶,۴۶

در قسمت بعد به پاسخگویی سوالات مطرح شده در این پژوهش پرداخته شده است.

۶- بحث و نتیجه‌گیری

در این قسمت ضمن پاسخگویی به سوالات مطرح شده سعی بر ارائه راه‌حلهایی برای موارد ذکر شده است. هرچند در نمونه مورد بررسی همه مشکلات مطرح شده در چهارچوب نظری قابل بررسی نبود (به علت عدم وجود برخی موارد ذکر شده) اما متداول‌ترین مشکل در نمونه مورد بررسی، خلاءهای واژگانی در زبان انگلیسی و خلاءهای واژگانی وردنت (دسمعناهای موجود در انگلیسی که هنوز در وردنت وارد نشده‌اند) بود. مشکلات موجود در انطباق فارسی‌نت با وردنت در قسمتی شبیه به مشکلات مطرح شده در انطباق شبکه‌واژگانی رومانیایی (کریستیا و همکاران، ۲۰۰۴) با وردنت بوده است که این موارد در قسمت نتایج و در زیربخش‌های ۱، ۲، ۴، ۵ و ۶ آورده شده و به قسمت‌هایی که در چهارچوب نظری با آن مشابه هستند اشاره گردیده است.

علاوه بر اینکه مشکلات جدیدی (در زیر بخش‌های ۳، ۷ و ۸) نیز مطرح گردیدند که قابل بررسی در وردنت رومانیایی یا زبان‌های دیگر می‌باشند. در ادامه راه‌حلهای زیر برای برطرف نمودن بخشی از مشکلات پیشنهاد می‌گردد:

۱. اگرچه در فارسی‌نت از رابطه نزدیک به معادل (*near equivalent*) و در یورونت از معادل-نزدیک به مترادف برای رابطه بین دسمعناهای شبکه مبدأ و وردنت استفاده گردیده است اما به نظر می‌رسد برای حل مشکل ۱ می‌توان از

انواع معادل‌ها بهره گرفت. مانند معادل کامل (در مواردی که خلاء ناشی از عدم درج برخی واژه‌ها در وردنت بوده است مانند مورد معادل غوره)، معادل نسبی (در مواردی که تناظر یک به یک بین دسمعنای فارسی و انگلیسی وجود ندارد ولی می‌توان آن‌ها را از جهاتی معادل نسبی دانست مانند غسل کردن و ablution که هر دو از جنبه مذهبی نوعی شستشو می‌باشند)، معادل فرهنگی (برای مثال برای آتش معادل فرهنگی soup لحاظ گردیده است) و معادل متقابل (برای این معادل به مواردی مانند غیرشهری که با rural انطباق یافته می‌توان اشاره نمود) بهره گرفت. نکته دیگر این است که ظاهراً به نظر می‌رسد برای بعضی از خلاءهای واژگانی فرهنگی نمی‌توان هیچ معادل نزدیک یا دور واژگانی شده‌ای یافت که به نظر می‌رسد ناشی از خلاء قالب^۱ یا طرحواره^۲ مربوطه در زبان انگلیسی است که برای بیان آن گاهی نیاز است از قالبها یا طرحواره‌های مشابه استفاده نمود و سپس براساس آن‌ها مفهوم جدید را معرفی کرد مانند طرحواره‌های مربوط به عزاداری.

۲. مشکل ۲: به نظر می‌رسد راه‌حل این مشکل برقراری روابط یک به چند بین وردنت مبدأ و مقصد است به عبارت دیگر چند دسمعنای فارسی با یک دسمعنای انگلیسی انطباق یابند. در یورونت از رابطه معادل-دارای-شامل استفاده گردیده است در مواردی که دسمعنای فارسی خاصتر است مانند دسمعناهای *عائله*، *نانخور*، *عاریتی* - عبور ممنوع - *عصاکش* - *قاضی القضاة* - *قبا* می‌توان از معادل عام بهره گرفت به عبارت دیگر دسمعنای وردنت را معادل عامتر نامید که شبیه به رابطه مورد استفاده در یورونت است.

۳. راه‌حل پیشنهادی برای مشکل ۳ یافتن معادل مناسب برای آن در زبان انگلیسی و ثبت آن در فارس‌نت با برچسبی مانند خلاء وردنت است و سپس این موارد را می‌توان به سازندگان وردنت گزارش نمود مانند معادل‌های *فازمتر* که عبارتند از {test light, test lamp, voltage tester, mains tester} و *فوتبال دستی* {table football, foosball, futsal}.

۴. راه‌حل پیشنهادی برای مشکل ۴ نیز افزودن امکان رابطه یک به چند بین دسمعنای فارسی و دسمعناهای انگلیسی است که در یورونت با رابطه معادل-نزدیک به مترادف انطباق یافته‌اند. به نظر می‌رسد این مسئله ناشی از دسمعناهای با تفاوت‌های جزئی در وردنت است که این تفاوت‌ها در زبان فارسی ممیز نمی‌باشند. این رابطه را می‌توان معادل خاصتر نامید. برای مثال برای انطباق دسمعنای {سخاوت، بخشندگی، گشاده دستی، ...} معادل {generosity, generousness} با تعریف the trait of being willing to give your money or time آورده شده است. اما آن را با دسمعناهای دیگر وردنت که عبارتند از {generosity, unselfishness} با تعریف acting generously و munificence, largess, liberality in bestowing gifts; extremely liberal and largesse, magnanimity, openhandedness با تعریف generous of spirit می‌توان انطباق داد.

۵. مشکل ۵: راه‌حل پیشنهادی برای این مشکل نیز ارتباط یک به چند بین دسمعناهاست که یک معادل برای دسمعناهای متفاوت به کار می‌رود و ناشی از کلمات همایند با آرگومان متفاوت در دسمعناهاست. در یورونت از رابطه معادل-نزدیک به مترادف استفاده گردید که در فارس‌نت آن را می‌توان معادل همایندی نامید.

¹ frame

² script

۶. مشکل ۶: به نظر می‌رسد برای این موارد نمی‌توان معادلی یافت. آن‌ها را می‌توان خلاء طبقه‌بندی نامید. مانند تقسیم اسلحه به سرد و گرم در فارسی‌نت که در انگلیسی معادلی ندارد.

۷. مشکل ۷: با توجه به اینکه در ترجمه یکی از فونمی که مترجم می‌تواند از آن بهره‌گیری استفاده از کلمات متضاد می‌باشد، در برخی موارد از این فن می‌توان در انطباق بهره‌گرفت هرچند یافتن معادل‌های واژگانی شده از پیش نیازهای ساخت شبکه‌واژگانی است. مسئله این است که همه متضادها ضرورتاً متقابل نیستند، به نظر می‌رسد این نوع انطباق را می‌توان تنها برای متقابل‌های غیر مدرج (non-gradable antonyms) استفاده نمود. این رابطه را می‌توان معادل-متضاد منفی نامید. برای مثال دسمعنای غیرعشائیری را می‌توان با urban و rural منطبق نمود.

۸. مشکل ۸: راه‌حل این مشکل انطباق بین مقوله‌ای است که افزودن این امکان به فارسی‌نت باعث پویایی بیشتر می‌گردد. در یورونت این رابطه معادل-نزدیک به مترادف نامیده شده است که در فارسی‌نت آن را می‌توان معادل بین مقوله‌ای نامید.

با مرور روابط مطرح شده در یورونت جهت انطباق دسمعناهای دارای مشکل، قابل مشاهده است که در چهار مورد از رابطه معادل-نزدیک به مترادف استفاده شده است. به نظر می‌رسد این راه‌حل‌ها بتوانند بخشی از مشکلات انطباق را حل نمایند، هرچند برای بررسی همه مشکلات مطرح شده در پژوهش بر روی شبکه‌واژگانی رومانیایی نیازمند بررسی کلیه دسمعناهای فارسی‌نت و وردنت است که در پژوهشی محدود نمی‌گنجد و نیازمند طرحی ملی است. اما می‌توان در انطباق بین شبکه‌های واژگانی از روابط دقیقتری مانند معادل عامتر، معادل خاصتر، معادل هم‌ایندی، معادل بین مقوله‌ای استفاده نمود. نتایج این پژوهش در تکمیل اطلاعات فارسی‌نت می‌تواند بکار گرفته شود و همچنین در طراحی و ایجاد شبکه‌های واژگانی تخصصی مورد استفاده قرار گیرد.

منابع

- حسابی، اکبر و عاصی، سید مصطفی و شمس فرد مهرنوش و عرب یارمحمدی، مهسا (۱۳۸۸). "طراحی و ایجاد شبکه‌واژگانی اسامی زبان فارسی". *مجله بین‌المللی ارتباطات و فناوری اطلاعات*. جلد یکم. ش ۳. صص ۱-۱۳.
- حسابی، اکبر (۱۳۹۴). "واژه‌های ویژه زبان فارسی؟: خلأهای موجود در انطباق واژگانی فارسی‌نت با شبکه‌واژگانی پرنستون (وردنت)". *جستارهای زبانی*. دوره ۶. ش ۶. صص ۷۵-۱۰۳.
- فامیان، علی رضا قلی (۱۳۸۶). *بررسی و تحلیل روابط معنایی صفت برای طراحی شبکه واژگانی صفات زبان فارسی*. رساله دکتری زبان‌شناسی همگانی. تهران: دانشگاه تربیت مدرس.
- منصوری، نیلوفر (۱۳۸۷). *بازنمایی معنایی افعال مرکب در وردنت فارسی*. رساله دکتری زبان‌شناسی همگانی. تهران: دانشگاه تهران.

- Bentivogli, L., Pianta, E. and Pianesi, F. (2000). "Coping with lexical gaps when building aligned multilingual WordNets". In: *Proceedings of LREC 2000*. Athens. Greece. pp: 12-20.
- Cristea, D., Mihaila, C., Forascu, C., Trandabar, D., Husarcus, M., Haja, G. and Postolache, O. (2004) "Mapping Princeton WordNet Synsets onto Romanian Wordnet Synsets," *Romanian Journal of Information Science and Technology*, 7(1-2), 125-145.
- Famian, A. and Aghajani, D. (2007) "Towards Building a WordNet for Persian Adjectives," *Proceedings of the 3rd Global WordNet conference*. Korea. pp: 307-308
- Fellbaum, C. (ed.) (1998) *WordNet: An Electronic Database*. Cambridge, MIT press.

- Fellbaum, C. and Vossen, P. (ed.) (2012) Proceedings: GWC2012 6th International Global WordNet Conference. Matsue, Japan.
- Hesabi, A. (2009) Designing and Developing the Persian Noun WordNet. Unpublished PhD dissertation. Tehran. Allameh Tabatabai University.
- Keyvan, F., Borjan, H., Kasheff, M. and Fellbaum, C. (2007) "Developing PersiaNet: the Persian Word-Net," Proceedings of the 3rd Global WordNet conference. Korea. pp: 315-318.
- Mansoori, N., Shamsfard, M. and Rouhizadeh, M. (2011) "Compound verbs in Persian WordNet." *International Journal of lexicography*. 1-18.
- Orav, H., Fellbaum, C., and Vossen, P. (ed.) (2014) Proceedings of the Seventh Global WordNet Conference. Tartu, Estonia.
- Rouhizadeh, M., Shamsfard, M. and Yarmohammadi, M. (2008) "Building a WordNet for Persian Verbs" A. Tanács, D. Csentes, V. Vincze, Ch. Fellbaum, P. Vossen (eds.) In: Proceedings of the Fourth Global WordNet Conference, Szeged, Hungary. pp: 406-412.
- Shamsfard, M., Hesabi, A., Fadaei, H., Mansoori, N., Famian, A., Bagherbeigi, S., Fekri, E., Monshizadeh, M. and Assi, M. (2010) "Semi Automatic Development of FarsNet; The Persian WordNet", *The fifth Global WordNet conference*. India. pp: 413-418
- Vossen, P. (2002) *EuroWordNet: General Document*. Netherland, University of Amsterdam. Version 3, Final, July 1, 2002. <http://hdl.handle.net/1871/11116>.