

واژگان زبانی زبان کُردی سورانی: KSLexicon
(KSLexicon: Kurdish-Sorani Generative Lexicon)

هاوری حسینی^۱

هادی ویسی^۲

محمد محمدامینی^۳

چکیده

در این مقاله واژگان زبانی زبان کُردی سورانی را به نام KSLexicon معرفی می‌کنیم. برای این منظور، ساختار تصریفی کلمه در زبان کُردی سورانی فرمول‌بندی و ارائه می‌شود که براساس آن یک واحدساز صرفی برای استخراج واژگان زبانی از پیکره کُردی طراحی شده است. اگرچه زبان کُردی گویشوران فراوانی دارد و سورانی یکی از دو گویش عمده این زبان است، از وجود منابع زبانی کافی برای پردازش زبان طبیعی بهره‌مند نیست. واژگان زبانی کُردی سورانی تلاشی برای فراهم آوردن واژگان زبانی این زبان است. واژگان استخراج شده در این تحقیق شامل ۳۵۰۰۰ واژه زبانی کُردی می‌باشد که دارای برچسب مقوله واژگانی می‌باشند.

^۱ دانشجوی کارشناسی ارشد زبان‌شناسی رایانشی، دانشگاه صنعتی شریف.

Hawre.hosseini1988@gmail.com

^۲ استادیار دانشکده علوم و فنون نوین، دانشگاه تهران.

H.veisi@ut.ac.ir

^۳ کارشناس مهندسی نرم‌افزار، دانشگاه باهنر کرمان.

Mohammadamini.mohammad@gmail.com

کلیدواژه‌ها: واژگان زایا^۴، ساختار تصریفی^۵، پردازش زبان طبیعی^۶، زبان گُردی سورانی^۷

۱. مقدمه

با پیشرفت علوم کامپیوتر و هوش مصنوعی در انتهای نیمه اول قرن بیستم، حوزه پردازش زبان طبیعی به عنوان یک حوزه بین‌رشته‌ای بسیار مورد توجه قرار گرفت. اگرچه این حوزه به دستاوردهای شایانی دست یافته است؛ با این حال، هنوز زبان‌هایی وجود دارند که در این زمینه مورد تحقیق و پژوهش قرار نگرفته‌اند. یکی از دلایل این امر عدم وجود منابع زبانی اعم از پیکره‌های مختلف زبانی است. زبان گُردی به عنوان یکی از زبان‌های هندوایرانی یکی از آن زبان‌هایی است که علی‌رغم داشتن شمار زیادی از گویشوران، از نبود منابع محاسباتی زبانی رنج می‌برد.

واژگان^۸ به مجموعه واژه‌های داخل ذهن گویشوران اطلاق می‌گردد (اسلامی و همکاران، ۱۳۸۳) که یکی از منابع مهم در تولید و ساخت تکنولوژی‌های پردازش زبان طبیعی مانند بازشناسی خودکار گفتار^۹، تبدیل رایانه‌ای متن به گفتار^{۱۰}، ترجمه ماشینی^{۱۱} و غیره می‌باشد. گویشوران زبان

⁴ generative lexicon

⁵ inflectional system

⁶ natural language processing

⁷ Kurdish-Sorani language

⁸ lexicon

⁹ automatic speech recognition

¹⁰ computer text to speech system

¹¹ machine translation

اطلاعات مختلف واژه‌ها را در ذهن خود در قالب توانش زبانی دارند (Bloomfield^{۱۲}, 1933: 247). یکی از این اطلاعات مربوط به اطلاعات ساخت‌واژی^{۱۳} واژه‌های مختلف می‌باشد. بدین شکل که گویشوران در ذهن خود اطلاعاتی در مورد مقولهٔ واژگانی هر واژه را ذخیره می‌کنند. اما، اینکه گویشور ساخت‌های متفاوت تصریفی کلمات را برای هر واژهٔ خاص در ذهن خود ذخیره کند، اصل اقتصاد در زبان را نقض می‌کند (اسلامی و همکاران، ۱۳۸۳). بنابراین، گویشوران ساختار تصریفی زبان خود را برای ساخت صورت کلمات متفاوت بصورت قواعدی در ذهن خود دارند و در صورت نیاز برای استفاده از واژه‌ها درون جمله‌ها، قواعد را بر روی واژه‌ها اعمال می‌کنند و صورت‌های مختلف کلمات را می‌سازند. برای مثال واژهٔ "خانه" صورت‌های تصریفی پرشماری دارد که در ذهن اهل زبان ذخیره نمی‌گردند. بلکه، تنها خود واژهٔ "خانه" و قواعد تصریفی نظیر اضافه کردن تک‌واژ جمع در ذهن ذخیره می‌گردد و صورت کلمهٔ "خانه‌ها" ساخته می‌شود.

بر اساس این ویژگی ذهن گویشوران زبان، واژگان با برچسب مقولهٔ واژگانی^{۱۴} می‌تواند زمینه را برای انجام مطالعات زبان‌شناختی در حوزهٔ ساخت‌واژه^{۱۵} فراهم کند. علاوه بر آن، واژگان به‌منظور تولید برچسب‌زن اجزای کلام^{۱۶} و تحلیل‌گر ساخت‌واژی^{۱۷} و ابزارهای لازم دیگر برای استفاده در حوزه‌های مختلف پردازش زبان طبیعی ضروری است.

¹² Bloomfield L.

¹³ morphological

¹⁴ part-of-speech tag

¹⁵ morphology

¹⁶ part-of-speech tagger

¹⁷ morphological parser

در این پژوهش، تلاش شد تا ابتدا برچسب مقوله واژگانی واژه‌های فارسی درون فرهنگ لغت گُردی-فارسی استخراج و به واژه‌های گُردی نسبت داده شوند. سپس، با استفاده از واژه‌های موجود و ساختار اشتقاقی زبان گُردی، واژه‌های داخل پیکره متنی گُردی را ارزیابی کنیم و آنها را با مقوله واژگانی شناخته شده براساس ساختار اشتقاق و تصریف در واژگان خود قرار دهیم.

در ادامه این مقاله، ابتدا در بخش ۲ مروری بر کارهای پیشین خواهیم داشت. سپس در بخش ۳ شیوه ساخت واژگان زایای گُردی سورانی (KSLexicon) و ویژگی‌های آن را ارائه می‌دهیم. نهایتاً، در بخش ۴ ساختار تصریف کلمه در زبان گُردی گویش سورانی فرمول‌بندی و ارائه می‌گردد. بخش ۵ به نتیجه‌گیری اختصاص دارد.

۲. کارهای پیشین

زبان گُردی یکی از زبان‌ها با منابع اندک پردازش زبانی می‌باشد. به همین دلیل، کارهای پردازشی انجام‌گرفته بر روی این زبان اندک می‌باشند. برای داشتن یک واژگان زایا نیاز به داشتن پیکره‌ای متنی از متون معاصر زبان داریم، زیرا واژگان بایستی شامل واژه‌های زایای زبان باشد. روش‌های مختلفی برای استخراج واژه‌ها و تولید واژگان پیشنهاد شده‌است. اما در زبان گُردی، تنها یک کار در این زمینه صورت گرفته است. والتر^{۱۸} و ساگوت^{۱۹} در (Walther, G., & Sagot, B, 2010) روشی متشکل از سه گام را برای ایجاد واژگان زبان گُردی سورانی پیشنهاد دادند. آنها از متون

¹⁸ Walther G.

¹⁹ Sagot B.

جمع‌آوری شده از وبلاگ برنامه تلویزیونی Ruwange به‌عنوان پیکره استفاده کرده‌اند (Ruwange Blog, 2015). پیکره آنها شامل ۵۹۰۵۶۸ واحد^{۲۰} و ۶۲۹۹۳ واحد منحصربه‌فرد بوده و واژگان آنها به نام SoraLex شامل ۱۷۶۰۰ مدخل تصریف‌شده برابر با ۱۳۳۱۵ نوع تصریفی جداگانه و ۴۶۸ مدخل ریشه‌ای می‌باشد. این واژگان، همان‌طور که از آمار پیدا است، واژگان کوچکی است.

۳. واژگان KSLexicon

واژگانی که ما در این تحقیق معرفی می‌کنیم KSLexicon نام دارد که شامل ۳۵۰۰۰ مدخل با ۲۸ برچسب می‌باشد. این واژگان بصورت نیمه‌خودکار جمع‌آوری گشته‌است که در ادامه، جزئیات گردآوری و تنظیم آن ارائه می‌گردند.

رویه گردآوری واژگان ما رویه‌ای دومرحله‌ای شامل استخراج واژگان اولیه از فرهنگ لغت گُردی-فارسی و برچسب‌زنی واژگان استخراجی با کمک واژگان فارسی، و درنهایت توسعه واژگان با کمک پیکره متنی بود. هرکدام از مراحل و چالش‌های پیش‌رو برای هر مرحله در ادامه تشریح می‌شود.

۱.۳. استخراج واژگان از فرهنگ لغت گُردی-فارسی

²⁰ token

در این تحقیق از فرهنگ لغت کُردی-فارسی هنبانه بورینه (هزار^{۲۱}، 1989) استفاده شده است که تنها منبع الکترونیک از واژگان کُردی سورانی قابل دسترسی عمومی است. این واژگان شامل ۵۵۰۰۰ مدخل از واژگان کُردی و برابره‌های فارسی آنها است. رویه به این ترتیب بوده است که برچسب برابره‌های فارسی واژگان کُردی موجود در فرهنگ لغت کُردی-فارسی را برای واژگان کُردی انتخاب کرده‌ایم. برای این کار، ما آن دسته از واژگان کُردی را که برابره‌های فارسی آنها در واژگان اف-لکسیکان (اسلامی و همکاران، ۱۳۸۳) قابل استخراج بود، استخراج کردیم.

یکی از چالش‌های پیش‌رو در این بخش برابره‌های فارسی موجود در واژگان کُردی بود. برابره‌های فارسی برای واژگان کُردی هنبانه بورینه غالباً به شکل عبارت و چندکلمه‌ای هستند، در صورتی که واژگان موجود در اف-لکسیکان به شکل یک کلمه‌ای هستند. برخی از برابره‌های فارسی نیز شامل عبارات چندکلمه‌ای و یک کلمه‌ای بودند. ما برابره‌های تک کلمه‌ای را برای واژه کُردی استخراج نمودیم.

چالش دیگر مشکلات خط فارسی در پردازش رایانه‌ای و همچنین شیوه‌های متفاوت نوشتن مانند فاصله‌گذاری‌های متفاوت برای یک کلمه‌ی واحد بود. یکی از پردازش‌های ما بر روی برابره‌های داخل فرهنگ لغت کُردی و همچنین واژگان داخل اف-لکسیکان برای یکدست‌سازی آنها بود. در ابتدا و بدون انجام یکدست‌سازی و مراحل پردازش، تعداد ۱۲۰۰۰ مدخل فرهنگ لغت را در اف-لکسیکان یافتیم، که این تعداد پس از انجام مراحل پردازشی به ۲۶۰۰۰ مدخل افزایش یافت.

²¹ Hajar

۲،۳. توسعه واژگان با کمک پیکره متنی گردی سورانی

برای گردآوری واژگان زایا نیاز به پیکره‌ای از متون زبان گردی سورانی داشتیم. نظر به اینکه فرهنگ لغت گردی-فارسی هنبانه بورینه واژه‌های لهجه‌های دیگر گردی غیر از سورانی را نیز در خود دارد، استخراج واژگان زایا از پیکره متنی گردی سورانی ضروری می‌نمود. رویه کلی کار به این صورت انجام گرفت که از واژگان استخراج شده به‌عنوان واژه‌های پایه استفاده کردیم. سپس، با توجه به ابزارهای اشتقاقی ساخت واژه‌های جدید صورت کلمات موجود در پیکره متنی گردی سورانی را پردازش و همچنین با توجه به ساختار تصریفی خاص هر مقوله واژگانی، واژه‌های ریشه را استخراج نمودیم.

نخستین پیکره بزرگ زبان گردی سورانی، پیکره متنی آسوسافت^{۲۲} است که شامل ۱۵۲ میلیون واحد می‌باشد (Veisi et al, 2015) که در این مطالعه از آن استفاده کردیم. پیکره‌های محدود و کوچک دیگری برای زبان گردی سورانی وجود دارند که یکی از آنها به نام پیوان شامل ۱۸ میلیون واحد است (SheykhEsmaili et al, 2013).

برای تحلیل صورت‌کلمه‌های داخل پیکره متنی از یک طرف نیاز به قواعد و ابزارهای ساخت کلمات جدید از واژه‌های ساده زبان داشتیم، و از طرفی دیگر نیاز به یک تحلیل‌گر ساخت‌واژی بود تا صورت‌کلمات را به‌عنوان ورودی دریافت کند و پس از جداسازی تک‌واژه‌های تصریفی، واژه

²² AsoSoft Text Corpus

غیرتصریفی را به‌عنوان خروجی بدهد. به منظور برآوردن نیاز اول، از منابع ساخت‌واژی گردی سورانی استفاده نمودیم که در آن قواعد زایش واژه‌های جدید در زبان گردی سورانی شرح داده شده‌اند (Marf A., 2014; Rokhzadi, A., 1393). نمونه قواعد به این شکل می‌باشند:

"به" + صفت (Adj) + "ی" --- < قید (Adv)

"به" + باش (Adj) + "ی" --- < به‌باشی (Adv)

برای تحلیل صورت‌کلمات از نظر زایش واژه جدید، نیاز داشتیم تا تک‌واژه‌های تصریفی کلمات شناسایی شوند. شناسایی تک‌واژه‌های تصریفی از دو جهت حائز اهمیت بودند. اولاً، شناسایی تک‌واژه‌های تصریفی خاص هر مقوله کمک می‌کرد تا نسبت به مقوله واژگانی واژه پس از جداسازی تک‌واژه‌های تصریفی اطمینان حاصل کنیم. دوماً، نیاز داشتیم که واژه را بدون هیچ تک‌واژ تصریفی‌ای به واژگان نهایی خود اضافه کنیم. بنابراین، ساختار تصریفی زبان گردی سورانی استخراج و فرمول‌بندی گشتند و براساس آن تحلیل‌گر ساخت‌واژی طراحی و پیاده‌سازی شد. سپس، متن پیکره متنی با قوانین زایش گردی سورانی و تحلیل‌گر ساخت‌واژی پردازش گشت. پس از پردازش‌های متفاوت برای اطمینان از مقوله واژگانی ساخت‌های استخراج شده، به واژگان استخراج شده برچسب مقوله واژگانی آنها زده شد. پس از این مراحل، تعداد ۹۰۰۰ واژه به واژگان اضافه گشت.

البته، برخی از مقوله‌های واژگانی نظیر انواع ضمائر و صفات اشاره به صورت دستی استخراج و برچسب زده شدند. انواع مقوله‌های واژگانی درون واژگان در جدول ۱ آمده‌اند.

جدول ۱. مقوله‌های واژگانی در واژگان KSLexicon

برچسب‌ها	مقوله‌های واژگانی	شماره
N1	اسم	۱
N2	اسم خاص مکان	۲
N3	اسم خاص اشخاص	۳
Ad1	صفت	۵
Ad2	صفت اشاره	۶
Ad3	صفت مبهم	۷
Ad4	صفت پرسشی	۸
V1	بن ماضی	۹
V2	بن مضارع	۱۰
V3	فعل اسنادی	۱۱
V4	فعل وجه‌نما	۱۲
Vpr	پیشوند فعل	۱۳
V5	فعل کمکی	۱۴
PrS	ضمیر فاعلی	۱۵
PrO	ضمیر مفعولی	۱۶
PrD	ضمیر اشاره	۱۷
PrR	ضمیر تاکیدی / ضمیر انعکاسی	۱۸

PrM	ضمیر مشترک / ضمیر متقابل	۱۹
Adv	قید	۲۰
Con	حرف ربط	۲۱
Pre	حرف اضافه	۲۲
PrePo	حرف اضافه پسین	۲۳
Spec	شاخص	۲۴
Intj	صوت	۲۵
Sign	علامت	۲۶
Alp	حروف الفبا	۲۷
Abb	اختصارات	۲۸

۴. ساختار تصریف در زبان گُردی سورانی

همان‌گونه که پیش‌تر عنوان شد، هر زبانی بسته به نظام ساخت‌واژی خود ابزارهایی برای تصریف کلمات در اختیار دارد و از آنها برای ایجاد صورت‌کلمات خاصی استفاده می‌کند. اهل زبان دانش تصریف کلمات را به عنوان توانش زبانی در اختیار دارند. اگرچه استثناهایی در ساخت‌های تصریفی وجود دارند، اما به نسبت قواعد موجود ناچیز می‌باشند و برای اهل زبان داشتن یک واژگان در ذهن و همچنین توانایی ساختن صورت‌های تصریفی با استفاده از قواعد کاملاً اقتصادی می‌باشد. در ادامه، ساخت‌های تصریفی کلمه در زبان گُردی سورانی و قواعد آنها به شیوه‌ی فرمول‌بندی‌شده می‌آید. بدیهی است که این فرمول‌بندی‌ها در کارهای محاسباتی و پردازش زبان طبیعی بسار حائز اهمیت می‌باشند.

۱.۴. فعل

فعل در زبان گُردی سورانی پیچیده‌ترین ساخت تصریفی را دارا می‌باشد. سعی شده‌است تا جایی که ممکن است فرمول‌های قواعد تصریفی افعال خلاصه و دربرگیرنده‌ی قواعد مربوط به بیشترین ساخت‌ها باشند.

۱.۱.۴. افعال زمان ماضی

در ساخت‌های افعال ماضی قواعد همه‌ی حالت‌های اخباری، اخباری استمراری ناتمام، نقلی، بعید و التزامی به شکل زیر در دو فرمول جمع گذشته‌اند.

• تمام حالت‌های ماضی بجز زمانی که فعل ساده متعدی باشد:

[پیشوند فعلی) / (عنصر غیر فعلی)] + (تک‌واژ "یش/ش") + (ضمیر متصل فاعلی) * + [(تک‌واژ نفی) / (تک‌واژ وجه التزامی-پیشوند)] + (تک‌واژ استمراری) + ستاک گذشته + [(تک‌واژ ماضی نقلی ساز) / (تک‌واژ ماضی بعید ساز-ستاک گذشته فعل بودن)] + (تک‌واژ وجه التزامی-پسوند) + (ضمیر متصل فاعلی) ** + (ضمیر متصل مفعولی) + (تک‌واژ پسوندی تکرار "هوه")

* همراه فعل متعدی ** همراه فعل لازم

مثال‌ها: دهر + یش + م* + نه + ده + خست + - + - + - + **
+ ن + هوه

دهر + یش + - + * + نه + ده + كهوت + - + - + م**
+ - + هوه

• تمام حالت‌های ماضی زمانی که فعل ساده متعدی باشد:

[(تک‌واژ نفی) / (تک‌واژ وجه التزامی-پیشوند)] + (تک‌واژ "یش/ش") * +
(ضمیر متصل فاعلی) * + (تک‌واژ استمراری) + ستاک گذشته + [(تک‌واژ
ماضی نقلی ساز) / (تک‌واژ ماضی بعید ساز-ستاک گذشته فعل بودن)] +
(تک‌واژ وجه التزامی-پسوند) + (تک‌واژ "یش/ش") ** + (ضمیر متصل
فاعلی) ** + (ضمیر متصل مفعولی) + (تک‌واژ پسوندی تکرار "هوه")
* زمانی که پیشوند تک‌واژ نفی یا وجه التزام داشته باشیم.

** زمانی که پیشوند تک‌واژ وجه نفی یا وجه التزام نداشته باشیم.

مثال‌ها: نه + ش* + م* + - + گرت + بوو + - + - + **
+ ن + هوه

- + - + * + - + گرت + بوو + - + ش** + م**
+ ن + هوه

این دسته از افعال در حالت‌های استمراری، اخباری و التزامی در قالب افعال ساده و غیرساده فرمول‌بندی می‌گردند.

• افعال ساده

[تک‌واژ نفی / تک‌واژ وجه التزامی / تک‌واژ استمراری] + (تک‌واژ "ش/یش") + (ضمیر متصل مفعولی) + ستاک حال + (ضمیر متصل فاعلی) + (تک‌واژ پسوندی تکرار "وه")

مثال: ده + ش + یان + خه + م + وه

ب + ش + یان + خه + م + وه

• افعال غیرساده

[پیشوند فعلی / عنصر غیرفعلی] + (تک‌واژ "ش/یش") + (ضمیر متصل مفعولی) + [تک‌واژ نفی / تک‌واژ وجه التزامی / تک‌واژ استمراری] + ستاک حال + (ضمیر متصل فاعلی) + (تک‌واژ پسوندی تکرار "وه")

مثال‌ها: دهر + بیش + یان + ده + خه + م + وه

دهر + بیش + یان + ب + خه + م + وه

مقوله اسم به نسبت فعل از ساخت تصریفی ساده‌تری برخوردار است. وابسته‌های تصریفی اسم در گُردی همه پسوندی هستند. فرمول قواعد تصریفی اسم در زبان گُردی سورانی به شکل زیر می‌باشد:

اسم + (تک‌واژ تصغیر) + (تک‌واژ معرفه‌ساز) + (تک‌واژ جمع) + (تک‌واژ
 نکره‌ساز) / (تک‌واژ بند موصولی) + [تک‌واژ ("ش/یش") + ("ی")]
 اضافه) / (واژه‌بست‌های شخصی)* + (تک‌واژ "ش/یش") +
 (واژه‌بست‌های ربطی)** / (ضمیر متصل فاعلی)*** / (بخش دوم حرف
 اضافه)

مثال‌ها: کتیب + وک + هک + ان + - + یش + یان* + - +
 ن** (.)

کتیب + وک + هک + ان + - + - + یان* + یش + م***
 (وه‌رگرت.)

۳،۴. مصدر

مصدر + (تک‌واژ پسوندی تکرار "وه") + (تک‌واژ معرفه‌ساز) / (تک‌واژ
 نکره‌ساز) / (تک‌واژ بند موصولی) + [("ی" اضافه) + (تک‌واژ
 "ش/یش") + (واژه‌بست شخصی) + (تک‌واژ "ش/یش") + (ضمیر متصل
 فاعلی)* / (واژه‌بست‌های ربطی)** * / (بخش دوم حرف اضافه)]

مثال: خویندن + هوه + که + - + مان + یش + یان*
(دی.)

(بینه‌ری) خویندن + هوه + یه‌ک + یش + مان + - + ن**
(.)

۴،۴. صفت

ساخت تصریفی صفت تا حدودی مشابه ساخت تصریفی اسم می‌باشد.
فرمول قاعده تصریفی صفت در گردی به شکل زیر می‌باشد.

صفت + [تک‌واژ صفت تفضیلی] / [تک‌واژ صفت عالی] + [تک‌واژ
معرفه‌ساز] + [تک‌واژ جمع] + [تک‌واژ نکره‌ساز] / [تک‌واژ بند
موصولی] + [تک‌واژ "ش/یش"] + ["ی" اضافه] / [واژه بست
شخصی]** + [تک‌واژ "ش/یش"] + [واژه بست ربطی]** / [ضمیر
متصل فاعلی]** / [بخش دوم حرف اضافه]

مثال‌ها: ژیر + ترین + هک + ان + - + یش + یان* + - +
ن** (.)

ژیر + ترین + هک + ان + - + - + یان* + یش +
م*** (هه‌لبژارد.)

۵,۴. قید

قید + (تک واژ معرفه ساز) + (تک واژ جمع) + (تک واژ نکره ساز) / (تک واژ بند موصولی) + (تک واژ "ش / یش") + (واژه بست شخصی) / ("ی" اضافه) + (ضمیر متصل فاعلی) / (واژه بست ربطی) / (بخش دوم حروف اضافه)

مثالها: ئیوار + - + ان + یک + یش + یان + مان
(دی.)

(له) ئیوار + هک + ان + - + یش + یان + دا

۶,۴. ضمیر

ضمیر + (تک واژ معرفه ساز) + (تک واژ جمع) + (تک واژ نکره ساز) / (تک واژ بند موصولی) + ("ی" اضافه) / (تک واژ "ش / یش") + (واژه بست شخصی) + (ضمیر متصل فاعلی) / (واژه بست ربطی) / (بخش دوم حروف اضافه)

مثال: هه رکهس + - + - + یک + یش + یان + مان

ئهو + هک + ان + - + ی + - + -

۷,۴. حرف اضافه

حرف اضافه + (تک‌واژ "ش/یش") + (واژه‌بست شخصی) + [ضمیر متصل
فاعلی] / (واژه‌بست ربطی) / (بخش دوم حرف اضافه) / (ضمیر متصل
مفعولی)

مثال‌ها: یّ + ش + مان + ی

له‌سه‌ر + یش + ت + ه

۸,۴. عدد

عدد اصلی + (پسوند عدد ترتیبی) + (تک‌واژ معرفه‌ساز) + (تک‌واژ جمع)
+ [(تک‌واژ نکره‌ساز) / (تک‌واژ بند موصولی)] + (تک‌واژ "ش/یش") +
(واژه‌بست شخصی) + (تک‌واژ "ش/یش") + [ضمیر متصل فاعلی] /
(واژه‌بست ربطی) / (بخش دوم حروف اضافه)

مثال‌ها: دوو + ههمین + هک + ان + - + - + یان + یش

+ مان (وه‌رگرت).

دوو + ههم + - + - + یّک + یش + یان + -
+ م (دی.)

۵. نتیجه‌گیری

ما در این مقاله واژگان زایای زبان گُردی سورانی را به نام KSLexicon معرفی کردیم. این دادگان به صورت نیمه‌خودکار استخراج شده است و شامل ۳۵۰۰۰ مدخل می‌باشد. همچنین، در این مقاله ساختار تصریف

گردی سورانی به شکل فرمول‌بندی شده ارائه شد. یکی از ابزارهای ساخته شده برای تهیه این واژگان تحلیل‌گر ساختواژی بود که براساس ساختار تصریف معرفی شده در این مقاله ساخته شد.

فهرست منابع

Bloomfield, L. (1933). *Language*. New York : Holt, Rinehart, Winston.

Hajar (1989). *Hanbana Borina Kurdish-Persian Dictionary*. Soroush Publication.

اسلامی، محرم، مسعود شریفی آتشگاه، صدیقه علیزاده لمجیری و طاهره زندی. ۱۳۸۳. واژگان زبانی زبان فارسی، مجموعه مقالات اولین کارگاه پژوهشی زبان فارسی و رایانه، دانشگاه تهران، ایران.

Walther G. and Sagot B. (2010). *Developing a Large-scale Lexicon for a Less-Resourced Language*. In *SaLTMiL's Workshop on Lessresourced Languages (LREC)*.

Rokhzadi A. (1393), *Kurdish Phonetics and Grammar*, Sanandaj, Iran, Kurdistan Publication.

Sheykh Esmaili K., Salavati Sh., Yosefi S., Eliassi D., Aliabadi P., Hakimi Sh., and Mohammadi A. (2013). *Building a Test Collection for Sorani Kurdish*. In *Proceedings of the 10th IEEE/ACS International Conference on Computer Systems and Applications (AICCSA'13)*.

- Marf A. Rahman. (2014), Kurdish Morphology (Noun), Vol. 1 Part .1, Erbil Iraq Rojhalat Publication.
- Marf A. Rahman. (2014), Kurdish Morphology (Pronoun), Vol. 1 Part . 2, Erbil Iraq Rojhalat Publication.
- Marf A. Rahman. (2014), Kurdish Morphology (Adjective), Vol. 1 Part .3, Erbil Iraq Rojhalat Publication.
- Marf A. Rahman. (2014), Kurdish Morphology (Adverb and Number), Vol. 1 Part .4, Erbil Iraq Rojhalat Publication.
- Marf A. Rahman. (2014), Kurdish Morphology (Verb), Vol. 1 Part .5, Erbil Iraq Rojhalat Publication.
- Ruwange Blog. (2015). <http://ruwange.blogspot.com>
- Veisi H., MohammadAmini M., Hosseini H. (2015). AsoSoft Kurdish-Sorani Text Corpus. Springer Language Resources and Evaluation. (Submitted)

شماره تماس نویسندگان مقاله:

هاوری حسینی:

تلفن همراه: ۰۹۳۵۷۶۲۴۸۲۹

تلفن همراه یدک: ۰۹۳۷۵۶۵۰۹۳۱

محمد محمدامینی:

تلفن همراه: ۰۹۱۸۹۷۸۳۶۰۱

تلفن ثابت: ۰۹۳۶۸۷۴۹۴۹۲