

بہ نام

قافلہ سالار جہان قدم

مرسلہ پیوند گلوی قلم



وزارت علوم، تحقیقات و فناوری

پژوهشگاه علوم انسانی و مطالعات فرهنگی

پژوهشکده زبان‌شناسی

گروه زبان‌شناسی همگانی

پایان‌نامه کارشناسی ارشد رشته زبان‌شناسی همگانی
طرحی برای تهیه پیکره‌ی تاریخی متون نثر زبان فارسی
از قرن پنجم تا هفتم هجری

استاد راهنما:

دکتر سید مصطفی عاصی

استاد مشاور:

دکتر عبدالحسین فرزاد

پژوهشگر:

سعیده قندی

مهر ماه ۱۳۹۳

سپاس‌گزاری

دهلیزهای دانشکده‌ی ادبیات دانشگاه تهران، اگرچه ذوق مرا به نگهبانی دادند^۱، اما دریچه‌ای شدند برای ورود به دنیای زبان‌شناسی. زبان‌شناسی را همان روزها که هنوز دانشجوی رشته ادبیات فارسی بودم و به کلاس‌های حامد عبدی می‌رفتم، شناختم. با او شناختم...

اما در باورم هم نمی‌گنجید که پیچک‌های سبز پژوهشگاه علوم انسانی در انتظارم باشند... اولین روز ورود مردی را ملاقات کردم که لبخندها و صبوری‌اش روحم را مجذوب خود کرد... که هرکسی را بهر کاری ساختند... و انگار که دکتر عاصی را برای این دو صفت آسمانی ساخته باشند. از ایشان بسیار آموختم...

صبوری را که برای من اولین و مهم‌ترین درس بود... آموختم "هرچیز که خوار آید، یک روز به کار آید!"

تلاش و امید را یادگرفتم... خستگی ناپذیری و پشتکار را یاد گرفتم و بسیار درس‌هایی که این سطور، رسانای نارسایی هستند برای ابرازشان... از ایشان بسیار سپاسگزارم و خود را همواره مدیون ایشان می‌دانم.

نمی‌شود به اتاق دکتر عاصی رفت و سری به اتاق همسایه مهربان ایشان نزد! آن‌جا می‌توان بانوی گشاده‌رویی را دید که به من نگاه آموخت... نگاهی دوباره به هرآنچه می‌دیدم... از خانم دکتر افراشی به خاطر این درس بزرگ متشکرم.

در آخرین اتاق راهروی طبقه سوم، مردی پشت میز استادی نشسته است که همه او را "بزرگوار" می‌دانند و به حق که بزرگوارند... اگر امروز ذره‌ای و فقط ذره‌ای دقت را چاشنی کارهایم کرده‌ام، حاصل هم نشینی با ایشان است... لطف و محبتی که ایشان نسبت به من دارند، بارها بارها بیش‌تر از حق استاد و دانشجو است... از آقای دکتر مدرسی سپاسگزارم.

هر بار که به فکر موضوع پایان‌نامه افتادم، تمام تلاشم را کردم که اگر هم قرار است با ادبیات آشتی کنم، این آشتی نسبتی با پایان‌نامه‌ام نداشته باشد! اما نشد و چه نشدی! درست در قعر ادبیات فرو رفتم... حالا، اما دوست که شده‌ایم هیچ، رفیق گرمابه و گلستان هم شده‌ایم... از آقای دکتر فرزاد در این آشتی سهم بزرگی داشتند سپاسگزارم... تشویق‌ها و راهنمایی‌های ایشان همیشه همراهام بود.

^۱شاعری وارد دانشکده شد... دم در، ذوق خود را به "نگهبانی" داد... (قیصر امین‌پور)

بار دیگر بیهقی خواندم و فهمیدم که بیهقی فقط لغت معنی نیست... می شود که یک ادیب و زبان شناس کنار هم باشند و به کسی به اصطلاح برنخورد! می شود که با هم باشند و این بار ادبیات را جور دیگری ببینند... که قرار نیست چیزی از قداست متن کم شود و به عکس، همه ی آنچه سال ها همگان بدیهی می پنداشتند، روش تر شود... که اگر نگوئیم ناگفتنی های بسیاری رخ می نماید... در این راه، دوستم، دیرینه دوستم، نصیبه، شریک ام بود... که به حق، اگر نبود در همان آغاز راه رها شده بودم. دوستی را در حق ام تمام کرد... از او ممنونم... کاش می توانستم مانند او شعری بسرایم... شعری برای قدرانی اش

در این راه، نمی توانم از امیدهای همیشگی آقای دکتر عاصی چشم پوشی کنم... که اگر نبود صبوری و همراهی ایشان، آدمی سست عنصر و بی اراده بودم.

آقای دکتر صفار مقدم، که همیشه مرا مورد لطف خود قرار داده اند، زحمت خواندن این سیاه مشق ها را پذیرفتند... امیدوارم که بعد از خواندن این خط ها، همچنان، من در نظرشان همان دانشجوی تلاشگر سابق باشم!

نمی توانم از "گهواره جنبانم"^۱ یاد نکنم. که اگر نبود دست هایش، نبودم امروز، مادرم... از مردی یاد می کنم که جهان بینی به من آموخت، پدرم و دیدگان هردوشان را می بوسم...

می خواهم از یگانه دوستی یاد کنم که در غبار جاده ی مرگ ناپدید شد... گلنسا، که اگر امروز بود پناهی بود برای دست های خسته ام...

از دوستان مهربانی که کمک های بی دریغشان دلگرمی بینهایتی بود؛ خانم آيسان رشید و آقای علی معصومی و هر که از ذهن کوچک من فراموش شد...

و از کسی که "رفتن، همیشه رفتن، رفتن و رسیدن..." را با او تجربه کردم.

از همسرم ممنونم...

بی هیچ پیش و پسی...

مرا گهواره جنبان بود مادر (ابوالحسن نوروزی)

^۱ به عهد خردسالی نازپرور

پیشکش

به همهی کسانی که از آنان آموختم و می آموزم...

چکیده

دیگر کمتر کسی هست که از اهمیت رایانه و نقش آن در زندگی امروزه غافل باشد. در علم زبان‌شناسی نیز، نقش آن بیش از پیش روشن شده‌است. تکیه بر داده‌های واقعی و عینی زبان، به صورت گسترده‌ای رواج یافته و شرط بسیاری از پژوهش‌های نظری و کاربردی شده‌است. این مهم با ساخت پیکره‌های زبانی میسر می‌شود: مجموعه‌ای از متن‌های نوشتاری یا گفتاری مکتوب که می‌توان در توصیف و تحلیل زبان از آن بهره گرفت. این پژوهش تلاشی در جهت ساخت اولین پیکره‌ی تاریخی از متون نثر قرن پنجم تا هفتم است. بدین منظور، ابتدا گزیده‌ای از متون این دوره بر اساس معیارهای مشخصی تعیین و با روشی یکسان تایپ شد. در گام بعد، همه‌ی متن‌ها وارد پایگاه داده‌های زبان فارسی شده و نمایه شدند. هدف از این کار، دستیابی بسیار سریع کاربران به اطلاعات مورد نظر و امکان ایجاد فهرست‌های بسامدی، واژه‌نما و آماری است. پیکره‌ی حاضر، تنها پیکره‌ای خام نیست، بلکه متن‌ها دارای نشانه‌گذاری‌هایی در سطح شناسنامه‌ی متن، برچسب‌های دستوری و آوایی است.

کلیدواژه‌ها: زبان‌شناسی پیکره‌ای، پیکره‌ی زبانی، برچسب‌دهی دستوری، پیکره‌ی تاریخی زبان

فارسی، متون نثر قرن پنجم تا هفتم

فهرست مطالب

.....	فصل ۱: کلیات
۲.....	۱-۱ مقدمه
۳.....	۲-۱ بیان مسئله
۵.....	۳-۱ پرسش‌های پژوهش
۵.....	۴-۱ فرضیه‌های پژوهش
۶.....	۵-۱ اهمیت پژوهش
۷.....	۶-۱ روش پژوهش
۸.....	۷-۱ ساختار پایان‌نامه
.....	فصل ۲: مروری بر آثار پیشین
۱۱.....	۱-۲ مقدمه
۱۱.....	۲-۲ مطالعات غیر ایرانی انجام‌شده در زمینه‌ی پیکره‌های زبانی
۱۵.....	۲-۲-۱ پیکره‌های نسل اول
۲۵.....	۲-۲-۲ پیکره‌های نسل دوم، پیکره‌های غول‌آسا
۳۰.....	۳-۲ مطالعات ایرانی انجام‌شده در زمینه‌ی پیکره‌های زبانی
۳۱.....	۳-۲-۱ پایگاه‌های داده
۳۱.....	۳-۲-۲ پیکره‌های متنی/نوشتاری
۴۴.....	۳-۲-۳ پیکره‌های دوزبانه
۳۶.....	۳-۲-۴ پیکره‌های گفتاری
۳۸.....	۳-۲-۵ پیکره‌های زبان‌آموز
۳۸.....	۴-۲ پیکره‌های تاریخی
۹۹.....	۴-۲-۱ پیکره‌های تاریخی برای زبان انگلیسی
۹۹.....	۴-۲-۱-۱ پیکره‌های در زمانی تاریخی
۳۳.....	۴-۲-۱-۲ پیکره‌های هم‌زمانی تاریخی
۳۳.....	۴-۲-۱-۳ پیکره‌های تجزیه‌شده
۴۴.....	۴-۲-۲ پیکره‌های تاریخی برای زبان‌های غیرانگلیسی
۴۵.....	۴-۲-۳ پیکره‌های تاریخی زبان فارسی
۴۷.....	۵-۲ جمع‌بندی
.....	فصل ۳: مبانی نظری

۹۹ ۱-۳ مقدمه
۵۰ ۲-۳ زبان‌شناسی پیکره‌ای
۵۳ ۱-۲-۳ دو رویکرد در زبان‌شناسی پیکره‌ای
۵۳ ۲-۲-۳ دو روش تحلیل داده‌ها در زبان‌شناسی پیکره‌ای
۵۴ ۳-۲-۳ پرسش‌های زبان‌شناسی پیکره‌ای
۵۵ ۴-۲-۳ انتقادات به زبان‌شناسی پیکره‌ای
۵۷ ۵-۲-۳ مهم‌ترین کاربردهای زبان‌شناسی پیکره‌ای
۵۷ ۱-۵-۲-۳ آموزش و یادگیری زبان
۵۸ ۲-۵-۲-۳ تحلیل گفتمان
۵۹ ۳-۵-۲-۳ مطالعات ادبی و سبک‌شناسی
۵۹ ۴-۵-۲-۳ مطالعات ترجمه
۶۰ ۵-۵-۲-۳ زبان‌شناسی قانونی
۶۰ ۶-۵-۲-۳ کاربردشناسی
۶۱ ۷-۵-۲-۳ زبان‌شناسی اجتماعی
۶۲ ۸-۵-۲-۳ فرهنگ‌نگاری
۶۳ ۹-۵-۲-۳ زبان‌شناسی تاریخی
۶۳ ۱۰-۵-۲-۳ دستور
۶۴ ۱۱-۵-۲-۳ تغییر زبان
۶۵ ۱۲-۵-۲-۳ معنی‌شناسی
۶۶ ۳-۳ پیکره‌ی زبانی
۶۸ ۱-۳-۳ ویژگی‌های پیکره
۷۱ ۲-۳-۳ اندازه‌ی پیکره‌ها
۷۲ ۳-۳-۳ مراحل ساخت پیکره
۷۳ ۴-۳ انواع پیکره‌های زبانی
۷۵ ۱-۴-۳ انواع پیکره‌های تخصصی
۷۵ ۱-۱-۴-۳ پیکره‌های منطقه‌ای یا جغرافیایی
۷۵ ۲-۱-۴-۳ پیکره‌های گویشی
۷۵ ۳-۱-۴-۳ پیکره‌های غیراستاندارد
۷۵ ۴-۱-۴-۳ پیکره‌های زبان‌آموز

۷۵۳-۴-۱-۵ پیکره‌های تاریخی
۷۶۳-۴-۱-۶ پیکره‌های تک‌زبانه
۷۶۳-۴-۱-۷ پیکره‌های دو زبانه
۷۶۳-۴-۱-۸ پیکره‌های چند زبانه
۷۶۳-۴-۱-۹ پیکره‌های موازی
۷۶۳-۴-۱-۱۰ پیکره‌های ایستا
۷۶۳-۴-۱-۱۱ پیکره‌های پویا یا مانیتور یا پایشگر
۷۷۳-۴-۱-۱۲ پیکره‌های نوشتاری
۷۷۳-۴-۱-۱۳ پیکره‌های گفتاری
۷۷۳-۴-۱-۱۴ پیکره‌های خام
۷۷۳-۴-۱-۱۵ پیکره‌های نشانه‌گذاری شده
۷۸۳-۴-۱-۱۶ پیکره‌های وب
۷۸۳-۴-۱-۱۷ پیکره‌های تطبیقی
۷۸۳-۴-۱-۱۷ پیکره‌های آموزشی
۷۹۳-۵-۱-۵ تحلیل پیکره‌ها
۷۹۳-۵-۱-۱ نشانه‌گذاری پیکره‌ها
۸۱۳-۵-۱-۱ سه رویکرد به نشانه‌گذاری زبانی
۸۲۳-۵-۱-۲ فواید پیکره‌ی نشانه‌گذاری شده
۸۳۳-۵-۱-۲ پردازش و آماده‌سازی متن
۸۳۳-۵-۲-۱ ابن واژه‌سازی
۸۵۳-۵-۲-۱ نوع و نمونه
۸۵۳-۵-۲-۲ برچسب‌دهی
۸۸۳-۵-۲-۱ انواع برچسب‌ها
۹۱۳-۵-۳-۳ تجزیه
۴۴۳-۶-۱ خروجی‌های حاصل از تحلیل پیکره
۴۴۳-۶-۱ فهرست‌های بسامدی
۹۷۳-۶-۲ فهرست واژگانی
۹۷۳-۶-۳ کلیدواژه‌ها
۹۹۳-۶-۴ کلیدواژه‌های مهم

۹۹ ۵-۶-۳ چندنگاشت
۹۹ ۶-۶-۳ باهم آیی
۱۰۲ ۷-۶-۳ هم ساختی
۱۳۳ ۸-۶-۳ فهرست واژه‌نما
۱۰۶ ۷-۳ پیکره‌های تاریخی
۱۰۶ ۱-۷-۳ چارچوب زمان و مسائل مرتبط با آن
۱۰۷ ۱-۷-۳-۱ آغاز و پایان پیکره‌ی تاریخی
۱۰۸ ۷-۳-۱-۲ تاریخ نمونه برداری متون پیکره
۱۰۸ ۷-۳-۱-۳ اندازه‌ی پیکره
۱۱۰ ۷-۳-۲ متون پیکره‌ی تاریخی
۱۱۱ ۷-۳-۳ محدودیت‌ها و ویژگی‌های پیکره‌های تاریخی
۱۱۴ ۷-۳-۴ پردازش پیکره‌های تاریخی
۱۱۷ ۷-۳-۵ نشانه‌گذاری پیکره‌ها
۱۲۰ ۸-۳ جمع‌بندی
	فصل ۴: توصیف و تحلیل پیکره
۱۲۳ ۱-۴ مقدمه
۱۲۳ ۲-۴ زبان فارسی
۱۲۶ ۳-۴ گستره‌ی زمانی پیکره
۱۲۶ ۴-۴ گستره‌ی مکانی پیکره
۱۲۷ ۵-۴ علت انتخاب دوره‌ی مورد بحث
۱۲۹ ۵-۴-۱ نثر فارسی از اوسط قرن پنجم تا اوایل قرن هفتم هجری
۱۰۰ ۶-۴ متون انتخابی برای پیکره
۱۳۲ ۶-۴-۱ اهمیت متون انتخابی
۱۳۷ ۷-۴ معرفی پیکره
۱۳۷ ۷-۴-۱ اندازه‌ی پیکره
۱۳۸ ۷-۴-۲ ساختار زبانی پیکره
۱۳۸ ۷-۴-۳ انتخاب و تایپ متن‌ها و چالش‌های پیش رو
۱۴۱ ۷-۴-۳-۱ بازیابی و ویرایش
۱۴۱ ۷-۴-۴ درونداد

۱۴۲ ۵-۷-۴ نشانه‌گذاری
۱۳۳ ۱-۵-۷-۴ نشانه‌گذاری شناسنامه‌ی متن
۱۳۳ ۲-۵-۷-۴ برچسب دستوری
۱۴۴ ۱-۲-۵-۷-۴ چالش‌های برچسب‌دهی
۱۴۶ ۳-۵-۷-۴ برچسب‌دهی آوایی
۱۴۶ ۶-۷-۴ کاربردان پیکره
۱۴۷ ۱-۶-۷-۴ سازمان‌ها
۱۴۷ ۲-۶-۷-۴ اشخاص
۱۴۷ ۷-۷-۴ سودمندی‌های پیکره
۱۹۹ ۸-۷-۴ امکانات پیکره
۱۵۰ ۹-۷-۴ ساختار رایانه‌ای پیکره
۱۵۱ ۱۰-۷-۴ کاربری‌های پیکره
۱۵۱ ۸-۴ جمع‌بندی
	فصل ۵: نتیجه‌گیری
۱۵۴ ۱-۵ مقدمه
۱۵۴ ۲-۵ خلاصه‌ی پژوهش
۱۵۶ ۳-۵ ارزیابی فرضیه‌ها
۱۵۷ ۴-۵ پیشنهادهایی برای پژوهش‌های آتی
۱۵۷ کتابنامه
۱۵۹ منابع فارسی
۱۶۱ منابع درون‌داده‌ی پیکره
۱۶۲ منابع انگلیسی
	پیوست‌ها
۱۶۸ ۱. دوره‌های پوشش داده‌شده‌ی پیکره‌ی هلسینکی
۱۷۰ ۲. صد واژه‌ی پرسامد پیکره ملی بریتانیا
۱۷۴ ۳. فهرستی از تمام متون نثر قرن پنجم تا هفتم هجری
۱۹۵ ۴. شیوه‌نامه‌ی تایپ متون
۱۹۷ ۵. جدول اصلاحات متون
۱۹۹ ۶. فهرست سرتیترهای متون

- پیوست ۷. نمونه‌ای از شناسنامه‌ی یک متن - کتاب ترجمه‌ی رساله‌ی قشیریه ۲۰۱
- پیوست ۸
- ۱-۸: نمونه‌ای از جمله‌های برچسب‌دهی شده‌ی دستوری دستی کتاب تاج‌التراجم ۲۳۳
- ۲-۸: نمونه‌ای از جمله‌های برچسب‌دهی شده‌ی دستوری دستی کتاب تاریخ بیهقی ۲۰۶
- ۳-۸: نمونه‌ای از جمله‌های برچسب‌دهی شده‌ی دستوری دستی کتاب تذکره‌الاولیاء ۲۱۰
- ۴-۸: نمونه‌ای از جمله‌های برچسب‌دهی شده‌ی دستوری دستی کتاب ترجمه‌ی رساله‌ی قشیریه ۲۱۳
- ۵-۸: نمونه‌ای از جمله‌های برچسب‌دهی شده‌ی دستوری دستی کتاب تمهیدات ۲۱۷
- ۶-۸: نمونه‌ای از جمله‌های برچسب‌دهی شده‌ی دستوری دستی کتاب ذخیره‌ی خوارزمشاهی ۲۲۰
- ۷-۸: نمونه‌ای از جمله‌های برچسب‌دهی شده‌ی دستوری دستی کتاب سفرنامه‌ی ناصر خسرو ۲۲۲
- ۸-۸: نمونه‌ای از جمله‌های برچسب‌خورده‌ی دستوری دستی کتاب قابوس نامه ۲۲۶
- ۸-۹: نمونه‌ای از جمله‌های برچسب‌دهی شده‌ی دستوری دستی کتاب کلیله و دمنه ۲۲۹
- ۸-۱۰: نمونه‌ای از جمله‌های برچسب‌دهی شده‌ی دستوری دستی کتاب لغت فرس ۲۳۳
- پیوست ۹: برچسب‌های دستوری پایگاه داده‌های زبان فارسی برای پیکره‌ی معاصر و تاریخی ۲۳۶
- پیوست ۱۰: نمونه‌ای از جمله‌های برچسب‌دهی شده‌ی آوایی خودکار در کتاب تاریخ بیهقی ۲۰۰
- پیوست ۱۲
- ۱-۱۲: نمونه‌ای از فهرست واژه‌نمای کلمه‌ی «لشکر» در کتاب تاریخ بیهقی ۲۴۷
- ۲-۱۲: نمونه‌ای از فهرست واژه‌نمای کلمه‌ی «طریقت» در کتاب تذکره‌الاولیا ۲۵۲
- ۳-۱۲: نمونه‌ای از فهرست واژه‌نمای کلمه‌ی «فایده» در کتاب ذخیره‌ی خوارزمشاهی ۲۵۷
- ۴-۱۲: نمونه‌ای از فهرست واژه‌نمای کلمه‌ی «شهر» در کتاب سفرنامه‌ی ناصر خسرو ۲۶۰
- ۵-۱۲: نمونه‌ای از فهرست واژه‌نمای کلمه‌ی «سیاست» در کتاب کلیله و دمنه ۲۶۵
- پیوست ۱۳
- واژه‌نامه‌ی فارسی به انگلیسی ۲۶۹
- پیوست ۱۴
- واژه‌نامه‌ی انگلیسی به فارسی ۲۸۰

فهرست جدول‌ها

- جدول ۱. نمونه‌ی یک جمله‌ی بن‌واژه‌سازی‌شده ۸۴
- جدول ۲. نمونه‌ی یک جمله‌ی برچسب‌دهی‌شده‌ی دستوری از پیکره‌ی ملی بریتانیا ۹۱
- جدول ۳. مقایسه‌ی کلیدواژه‌هایی در دو پیکره‌ی *il* اُ بی و *if* اِل اُ بی ۹۶
- جدول ۴. بیست باهم‌آیند از واژه‌های *girl* و *boy* در پیکره‌ی ملی بریتانیا ۱۰۲
- جدول ۵. نمونه‌هایی از فهرست واژه‌نمای کلمه‌ی *window* ۱۰۲
- جدول ۶. نمونه‌ای تصادفی و مرتب‌نشده از فهرست‌های واژه‌نمای و کلمه‌ی *girl* در پیکره‌ی اِل اُ بی ۱۰۴

فصل ۱: کلیات

۱-۱ مقدمه

انسان‌ها به طور طبیعی موجوداتی کنجکاو در کسب دانش هستند. رایانه‌ها نیز برای پاسخ به این کنجکاوای بشر کمک فراوانی کرده‌اند، تا جایی که امروز، دیگر، نمی‌توان نقش آن‌ها را در این راستا نادیده گرفت. نقش واهمیت روزافزون این وسیله در تمام علوم و به‌خصوص در علم زبان‌شناسی برای همگان بیش از پیش روشن است.

رایانه‌ها در زبان‌شناسی تحولات بسیاری ایجاد کرده و افق‌های تازه‌ای پیش روی زبان‌شناسان گسترده‌اند. "زبان‌شناسی رایانشی"^۱ که پیوند رایانه و زبان‌شناسی است، از جوان‌ترین حوزه‌های زبان‌شناسی محسوب می‌شود و به عنوان «مطالعه‌ی زبان با کمک رایانه» تعریف شده است.

"زبان‌شناسی پیکره‌ای"^۲ شاخه‌ی دیگری است که حاصل پیوند رایانه و علم زبان‌شناسی است. شاخه‌ای که بنیادی روش‌شناختی برای پژوهش‌های زبانی به شمار می‌آید.

استفاده از "پیکره‌های زبانی"^۳ و "روش‌های پیکره‌بنیاد"^۴ در پژوهش‌های زبانی سنتی دیرینه است. زبان‌شناسان ساخت‌گرای آمریکایی از بلوم فیلد^۵ (۱۹۹۳) تا هریس^۶ (۱۹۵۱) در توصیف زبانی از رویکردهای تجربی استفاده می‌کردند. در زبان‌شناسی نظری امروز نیز، تکیه بر داده‌های واقعی و عینی به صورت گسترده‌ای رواج یافته و به کار گرفته می‌شود. مدافعان و حامیان متأخر شامل، بیرنر^۷ (۱۹۹۴)، مکفارلند^۸ (۱۹۹۷) و مایکلیس^۹ (۱۹۹۶) نیز، در بسیاری از پژوهش‌های خود، به‌جای تعمیم‌های ازقبل پیشنهادشده درباره‌ی ساخت‌های زبانی، به

^۱.computational Linguistics

^۲. corpus Linguistics

^۳.corpora

^۴.corpus-based

^۵. L. Bloomfield

^۶.Z. Harris

^۷.Birner

^۸. Macfarland

^۹. Michaelis

پیکره‌ها روی آوردند و آن را اساس کار خود قرار دادند(بیر^۱ و همکاران، ۱۹۹۸: ۲۳۳).

با توجه به اهمیت موضوع، استفاده از رایانه‌ها و پژوهش‌های پیکره‌ای برای زبان فارسی، از دید پژوهش‌گران ایرانی دور نمانده و سال‌های زیادی است که در کشور آغاز شده و هر ساله ارتقا یافته است.

در همین راستا، پژوهش حاضر به یکی از جنبه‌های موضوع "زبان فارسی و رایانه" می‌پردازد.

هدف از این پژوهش، ارائه‌ی نمونه‌ای برای ساخت نخستین پیکره‌ی تاریخی از متون نثر کلاسیک زبان فارسی از قرن چهارم تا چهاردهم هجری است که به بررسی ویژه‌ی گزیده‌ی متون میانه‌ی قرن پنجم تا آغاز قرن هفتم هجری می‌پردازد. این پیکره همراه با برچسب دستوری دستی، برچسب خودکار آوایی (برای واژه‌هایی که قابل شناسایی است)، نشانه‌گذاری‌های فهرست بسامدی و فهرست واژه‌نما به صورت برخط و رایگان، به عنوان فاز تاریخی در پایگاه داده‌های زبان فارسی قرار می‌گیرد.

در این فصل، پس از بیان مسئله و پرسش‌های پژوهش، به توضیح فرضیه‌های پژوهش، اهمیت موضوع و روش‌های پژوهش پرداخته شده است. در پایان نیز، ساختار کلی پایان‌نامه ترسیم شده است.

۱-۲ بیان مسئله

در قلب زبان‌شناسی پیکره‌ای مفهوم "پیکره" قرار گرفته است. کِنِدی^۲ (۱۹۹۸: ۱) می‌نویسد: «پیکره‌ی زبانی مجموعه‌ای از متن‌های نوشتاری یا گفتاری آوانویسی شده است که می‌توان در توصیف و تحلیل زبان از آن بهره گرفت». پیکره‌های زبانی را می‌توان از جنبه‌های مختلفی طبقه‌بندی کرد. "پیکره‌های تاریخی"^۳، یکی از انواع

^۱. S. Bayer

^۲. G. Kennedy

^۳. historical Corpora

پیکره‌ها هستند و شامل متونی از یک یا چندین دوره‌ی تاریخی‌اند. این پیکره‌ها برای بررسی زبانِ دوره‌های گذشته و مطالعه‌ی چگونگی تغییرات زبان در طول زمان، کاربرد دارند (بیکر^۱ و همکاران، ۲۰۰۶: ۸۵).

با توجه به ارتباط دو رشته‌ی ادبیات و علم زبان‌شناسی و رشد روزافزون کارایی و سودمندی فناوری اطلاعات و رشته‌های مرتبط با آن، از جمله زبان‌شناسی پیکره‌ای، لازم است که این علم در به‌روز شدن و تسهیل دسترسی به داده‌های ادبی یاری‌رسان باشد. چه بسیاری از فعالیت‌های علمی در حوزه‌ی زبان، ادبیات و زبان‌شناسی به داده‌های مشابهی نیاز دارند که هر یک برای خود تکه‌ای یا گوشه‌ای از گستره‌ی زبان را دربر می‌گیرند.

با توجه به موارد بیان‌شده، مسئله‌ی این پایان‌نامه این است که چگونه می‌توان برای زبان فارسی، با توجه به گستردگی و تنوع آن، پیکره‌ای تاریخی ساخت؟ پرسش‌های پژوهش، معطوف به همین مسئله هستند.

طرح حاضر به عنوان اولین پیکره‌ی تاریخی از متون نثر کلاسیک زبان فارسی، سعی بر آن دارد تا تصویری از ارتباط ادبیات و زبان‌شناسی به دست دهد.

برای تهیه‌ی این پیکره، متون نثر را براساس کتاب تاریخ ادبیات در ایران دکتر ذبیح الله صفا به دوره‌های مختلف تقسیم‌بندی می‌کنیم.

در مرحله‌ی نخست، شناسنامه‌ی متون را در بخشی مجزا در پایگاه داده‌های زبان فارسی قرار می‌دهیم که کمک بزرگی به کاربران این پایگاه می‌کند. سپس متون را به روشی یکسان تایپ می‌کنیم.

در دومین مرحله، همه‌ی متن‌ها وارد پایگاه داده شده، نمایه می‌شوند. هدف از این کار، دستیابی بسیار سریع کاربران، در کسری از ثانیه، به اطلاعات مورد نظر و امکان ایجاد فهرست‌های بسامدی^۲، واژه‌نما^۳ و آماری^۴ است. مزیت این روش دقت فراوان آن و تحلیل متون از زوایای مختلف است که هر کدام از آن‌ها به روش‌های دستی

1. P. Baker et al

2. frequency lists

3. concordance

4. statistics

متضمن صرف وقت بسیار زیاد و از آن مهم‌تر، دقت پایین‌تر است. در مرحله‌ی سوم، این متن‌ها را با به صورت دستی برچسب‌دهی دستوری می‌کنیم. این پیکره با توجه به فازهای مختلف آن به‌روز خواهد شد و از مزایای ارزشمند آن، دسترسی رایگان به همه‌ی اطلاعات و امکان استفاده‌ی تمام کاربران از آن، با توجه به نیازهای مختلف آنان است. همچنین امکان تهیه‌ی پیکره‌های خاص و موردی با توجه به موضوع و زمان تالیف متون فراهم است که خود در بررسی‌های مقابله‌ای در رشته‌های مختلف و مخصوصاً زبان و ادبیات فارسی و زبان‌شناسی حائز اهمیت فراوانی است.

۳-۱ پرسش‌های پژوهش

- ۱- پیکره‌ی تاریخی مطلوب باید شامل چه متونی باشد؟
- ۲- نشانه‌گذاری این متون چگونه و در چه سطحی انجام می‌شود؟
- ۳- روش درون‌داد و ویژگی‌های ویرایشی و تایپی این پیکره چیست؟

۴-۱ فرضیه‌های پژوهش

- ۱- پیکره‌ی تاریخی مطلوب برای متون نثر زبان فارسی (پیش از مشروطه) باید نماینده‌ی گویایی از دوره‌های گوناگون زبانی باشد و نمونه‌های شاخص هر دوره را دربرگیرد.
- ۲- بهره‌گیری از معتبرترین نسخه‌های موجود در هر دوره، درون‌داد رایانه‌ای آن‌ها با دقت بالا و افزودن اطلاعات بیشتر برای غنی‌سازی متون
- ۳- وارد کردن پیکره به پایگاه داده‌های زبان فارسی و بهره‌گیری از توان‌مندی‌های پردازشی آن و ارائه‌ی خدمات از راه اینترنت

۵-۱ اهمیت پژوهش

زبان‌شناسان همواره بر بنیاد نهادن بررسی‌های زبانی و استدلال‌ها و استنتاجی مبتنی بر داده‌های واقعی زبان تاکید داشته و دارند و همواره بهترین راه برای استفاده از داده‌های مستند را ایجاد و استفاده از پیکره‌های زبانی می‌دانسته‌اند. از طرفی، گسترش روزافزون رایانه و علوم مختلف میان‌رشته‌ای، این امکان را برای زبان‌شناس فراهم می‌آورد که بتواند پیکره‌های بسیار عظیم زبانی را به‌وجود آورد و آن‌ها را به گونه‌های مختلف طبقه‌بندی و نشانه‌گذاری کند و به طرق بسیار گوناگون از اطلاعات به‌دست آمده بهره‌مند شود.

رشد زبان‌شناسی پیکره‌ای و پیکره‌های زبانی در دهه‌ی ۱۹۸۰ و سرعت پردازش رایانه‌ها، این امکان را به پژوهشگران داد تا کاری که ممکن بود ماه‌ها به طول بینجامد و مستلزم مخارج بسیاری باشد، ظرف مدت چند دقیقه و حتی کسری از ثانیه و با دقت بالا به انجام برسانند (حاجی عبدالحسینی، ۱۳۷۵:۱).

همچنان که اشاره شد، برای زبان فارسی، نیز، پژوهش‌های پیکره‌ای متعددی از سال‌های پیش آغاز شده و همچنان ادامه دارد و ما در فصل دوم از آن‌ها یاد خواهیم کرد. تا آن‌جا که نگارنده بررسی کرده است، پیکره‌ای که تمام متون آن تاریخی بوده و به بررسی تحول زبان فارسی در ادوار مختلف بپردازد، وجود ندارد و بیشتر پیکره‌ها از متون فارسی معاصر یا گفتار روزمره بوده است. تنها پیکره‌ای که شامل متون تاریخی است، پیکره‌ی فرهنگستان زبان و ادب فارسی است که در واقع حاصل ماشینی کردن فیش‌های دستی بوده و شامل پاره‌هایی از متون نظم و نثر از ابتدای قرن چهارم تا سال ۱۳۸۲ است. اما این پیکره، نیز، در حقیقت گزیده‌ی متون است و تمام یک متن را دربر ندارد و برای تهیه‌ی فرهنگ جامع زبان فارسی طراحی شده و در اختیار همگان نیست.

از آن‌جا که نگارنده، خود، دانشجوی رشته‌ی زبان و ادبیات فارسی بوده است با خلأ پیکره‌ی تاریخی روبرو بود. چه بسیاری از پایان‌نامه‌های این رشته به بررسی‌های سبک‌شناسی اختصاص دارد و در این راه، مقایسه‌ی ویژگی‌های زبانی دوران مختلف، کاربرد واژه‌ها در بافت‌های گوناگون و بسامد آن‌ها نقشی اساسی دارند. علاوه بر صرف زمان و هزینه طولانی در این کار، امکان اشتباه و عدم دقت به دلیل دستی بودن محاسبات وجود دارد. سالیان سال

این روند ادامه داشته و تا کنون نیز ادامه دارد. لذا لزوم تهیه‌ی چنین پیکره‌ای بیش از گذشته احساس می‌شود. علاوه بر اهمیت در سبک‌شناسی، این پیکره می‌تواند به "تحلیل محتوا"، نیز، کمک کند. اما شاید بتوان مهم‌ترین فایده‌ی آن را در بررسی‌های زبان‌شناختی و تحلیل‌های زبانی دانست. کاربردهای دیگر چنین پیکره‌ای در حوزه‌ی فرهنگ‌نگاری، دستور نویسی، آموزش زبان و ادب فارسی، و مانند این‌هاست. امید است که پژوهش حاضر گامی در جهت ارتقای زبان و ادب فارسی باشد.

۱-۶ روش پژوهش

در این پژوهش از منابع کتابخانه‌ای استفاده می‌شود. زیرا داده‌های این طرح از متون قدیمی و تنها براساس نسخه‌های تصحیح شده به دست می‌آید. در حقیقت، منابع براساس کتاب‌های تصحیح شده و در دسترس، تایپ می‌شوند و از آنجاکه قرار است از سکوی پایگاه داده‌های زبان فارسی برای ارائه‌ی آن استفاده شود، از امکانات پردازشی این پایگاه برای نشانه‌گذاری و نمایه کردن متون بهره‌گیری می‌شود.

در میانه‌ی قرن پنجم تا آغاز هفتم هجری، یازده ژانر^۱ (گونه‌ی سبکی^۲) عرفانی، تاریخی، سفرنامه‌نویسی، فرهنگ لغت، تفسیر قرآن، پزشکی اسلامی و داروشناسی، فابل(داستان‌هایی از زبان حیوانات)، اندرزنامه، کشورداری و کلام است وجود دارد. ما شاخص‌ترین متن هر ژانر را انتخاب کرده‌ایم، سپس متون را با شیوه‌ای یکسان تایپ کرده و پس از اعمال اصلاحاتی بر روی آن، وارد پایگاه داده‌های زبان فارسی می‌کنیم. در مرحله‌ی آخر نیز آن‌ها را به صورت دستی برچسب‌دهی می‌کنیم.

1. genre

^۲ از این پس، برای سهولت نوشتار، از اصطلاح عام‌تر «ژانر» استفاده می‌کنیم.