

نشریه پژوهش‌های زبان‌شناسی

سال ششم، شماره دوم، شماره ترتیبی 11، پاییز و زمستان 1393

تاریخ وصول: 1393/05/04

تاریخ اصلاحات: 1393/11/13

تاریخ پذیرش: 1393/11/20

صص 71 - 88

## اعتبارسنجی گویش سنجی کرانه جنوب‌خاوری دریای مازندران به روش بسامد آوایی هونبروورس

\*سیف‌اله ملایی پاشایی

\*\*آرزو نجفیان

### چکیده:

در مقام بررسی‌های نوپای گویش‌شناسی رایانشی، داده‌های زبان مازندرانی - مستخرج از طرح ملی اطلس زبانی ایران - با اعمال الگوریتم لونیشتین کاوش، و منطقه‌های گویشی در دامنه شمالی رشته کوه‌های البرز مکان‌یابی شده است. در این مقاله کوشیده‌ایم با واکاوی داده‌ها به روش تحلیل بسامد واکه، ضمن تشخیص متغیرهای آوایی مسبب تفاوت‌های گویشی منطقه‌ای، پایایی و اعتبار یافته‌های گویش‌سنجی را بیازماییم. پژوهش حاضر، مبتنی بر روش توصیفی هم‌زمانی و به لحاظ دادگان، میدانی است.

مواد زبانی متشکل از 113 مصاحبه است. میانگین سنی گویشوران 42 سال، متوسط سواد آنها در حد آموزش ابتدایی و 34٪ آنها زن بودند. نقشه‌های بازنمودی و تفسیری حاصل، ضمن تأیید اعتبار یافته‌های پژوهش‌های پیشین با آلفای کرنباخ 0/84، تظاهر منطقه‌های گویشی را پشتیبانی می‌کند و بیانگر پایایی روش تحقیق و روایی پژوهش است.

**واژگان کلیدی:** زبان‌شناسی رایانشی، گویش‌شناسی، روش بسامد آوایی، زبان مازندرانی.

**1. مقدمه**

در حوزه جغرافیای گویشی، گونه‌گونی زبان‌ها یا توزیع جغرافیایی گویش‌وران یک زبان، بر روی یک نقشه موضوعی زبان‌شناختی نمایش داده می‌شود. به دیگر سخن، جغرافیای گویشی به بررسی پراکنش جغرافیایی ویژگی‌های زبانی مشخص، در پهنه جغرافیایی معین، می‌پردازد و آن را به صورت نقشه زبان‌شناختی بازمی‌نمایاند. بنابراین، بعد اولیه و مهمتر در این دست مطالعات گویش‌شناختی، گوناگونی‌های زبانی و بعد ثانویه، فاصله جغرافیایی است. در این حوزه «به تبع تغییرات نظری و روش‌شناختی در علوم نظیر زبان‌شناسی، جامعه‌شناسی، ریاضی و رایانه، روش‌های متعدد و متفاوتی بسط یافته است» (پاتریارکا<sup>1</sup> و دیگران، 2009:174). روش‌های مختلف آماری و ریاضی برای محاسبه مقداری و کمی میزان تمایزات و ویژگی‌های زبانی در نمونه‌های گسترده، به تدریج شکل گرفت؛ بعدها به کارگیری رایانه نیز برای تحلیل داده‌های زبانی حجیم میسر شد و رهیافت گویش‌سنجی پا گرفت. «سپس این شیوه مقبول افتاد و بر گویش‌های هلندی، ساردینی، نروژی، آلمانی، بلغاری، و سوئدی پیاده شد» (اُسِنُوا<sup>2</sup> و دیگران، 2010:1).

پیاده‌سازی روش‌های مختلف گویش‌شناسی، بویژه روش‌های نوین رایانشی، و ساخت و پرداخت روش‌شناسی نظام‌مند در این حوزه، برای گسترش و یک‌پارچه ساختن مطالعات گویشی ضروری می‌نماید. ملایی‌پاشایی (1393) با تحلیل انبوهه داده‌های گویشی زبان مازندرانی مستخرج از طرح ملی اطلس زبانی ایران به گویش‌سنجی دامنه شمالی البرز با اعمال الگوریتم لِنِشتین پرداخته است. در پی آن نجفیان و همکاران (1393) به روشی مشابه گونه‌گونی و پراکنش زبانگونه خاوری واقع در دامنه شمالی البرز خاوری در استان مازندران را به رشته تدقیق کشیدند. در مقاله پیش‌رو کوشیده‌ایم پایایی و اعتبار یافته‌های پژوهش‌ها را به محک روش تحلیل بسامد واکه ملهم از هوبنبروورس و هوبنبروورس<sup>3</sup> (2001) بیازماییم؛ چرا که هدف از نخستین کارهایی که به روش نو انجام می‌شود، در درجه اول سنجیدن اعتبار خود آن روش‌ها است. ابتدا به پیشینه مختصر پژوهش و سپس به مبانی نظری و روش‌شناسی گویش‌سنجی دامنه شمالی البرز پرداخته می‌شود. شیوه جمع‌آوری داده‌ها، قلمرو زمانی و مکانی پژوهش نیز در این بخش قرار دارد. سرانجام بررسی مقابله‌ای یافته‌ها، اعتبارسنجی آنها و نتیجه‌گیری بخش چهارم مقاله را شکل می‌دهد.

**2. پیشینه پژوهش**

مطالعات گویش‌سنجی فراوانی صورت گرفته است. از جمله بسیار، نربن و کِرتسچمار<sup>4</sup> (2003) به معرفی پاره‌ای فناوری‌های گویش‌سنجی پرداخته‌اند. پالاندر<sup>5</sup> و دیگران (2003) با تحلیل انبوهه داده‌های واجی و ساختوازی مستخرج از 198 مصاحبه، و خوشه‌نگاری به وسیله بسته نرم‌افزاری R، مناطق گذار<sup>6</sup> در فنلاند شرقی را تعیین کردند. پُرکیک<sup>7</sup> (2007) با تحلیل انبوهه مواد زبانی بلغاری، ضمن ارائه ساختار زبان‌شناختی تفاوت‌های گویشی، آن جایگزینی‌های آوایی که مسبب تفاوت‌های گویشی بلغاری هستند را بازشناخت. اُسِنُوا و دیگران (2010) نیز بر پایه تلفظ 36 واژه در

<sup>1</sup> M. Patriarca

<sup>2</sup> P. Osenova

<sup>3</sup> C. Hoppenbrouwers & G. Hoppenbrouwers

<sup>4</sup> J. Nerbonne & W. Kretzschmar

<sup>5</sup> M. Palander

<sup>6</sup> relic area

<sup>7</sup> J. Prokic

490 سایت<sup>1</sup> در بلغارستان، اطلسی گویشی بر پایه مقیاس‌گذاری چندبعدی ارائه نمودند. هیرینگا<sup>2</sup> و دیگران (2009) شاخص فاصله زبانی گویش‌های نروژی را بر پایه ویژگی‌های آکوستیک<sup>3</sup> - به هدف حذف اشتباه و سوگیری ناخودگاه فردی در آوانگاری - گویش‌سنجی کرده‌اند. برادران هونینوروس (1988؛ 2001)، و نربن و هیرینگا<sup>4</sup> (1988؛ 2004) بر اساس بسامد آوا به دسته‌بندی گویشی پرداختند. لی‌نونن<sup>5</sup> (2010) به تحلیل انبوهه مناطق گویشی سوئدی بر پایه ساخت واکه‌ای آن می‌پردازد. با اعمال این روش‌ها، ویلینگ<sup>6</sup> (2012) به بررسی کمی دگرگونی‌های زبان جامعه‌شناختی کاتلان می‌پردازد.

در ایران - علاوه بر ملایی‌پاشایی (1391؛ 1393) و نجفیان و همکاران (1393) - اسدپور (1390) در پایان‌نامه کارشناسی ارشد خود درجه تراکم تفاوت‌های واژگانی، آوایی، ساختواژی، و نحوی در میان گونه‌های ترکی آذری را گویش‌سنجی کرده است. وی نتیجه می‌گیرد که گونه‌های مختلف ترکی آذری گویش نیستند، بلکه زیرلهجه‌های آن محسوب می‌گردند. نیز رستم‌بیگ نفرشی (1393) برای تدوین اطلس گویشی و اندازه‌گیری فاصله‌های گویشی در نیمه جنوبی استان همدان از پرسشنامه‌ای مبتنی بر فهرست واژه‌های سوادش برای گردآوری داده‌ها از 15 سایت استفاده کرده و دریافته است که تنوع واژگانی در مقوله «طبیعت» بیشترین حد و در «اسامی حیوانات» کمترین حد را دارد.

### 3. پایگاه نظری و روش پژوهش

در این پژوهش دو روش از گویش‌سنجی، همانا روش تحلیل انبوهه و روش بسامد واکه، ارائه و مقایسه می‌شود تا از این رهگذر اعتبار چنین روش‌های نویی سنجیده شود. برای این کار به بررسی مقابله‌ای مطالعه‌ای میدانی در هر یک از این روش‌ها می‌پردازیم؛ این دو مطالعه در ادامه این بخش متعاقباً در (3-1) و (3-2) ارائه شده است. برای سنجش روایی و نیز مقایسه اعتبار یافته‌های دو پژوهش با یکدیگر از روش آلفای کرنباخ<sup>7</sup> بهره می‌جویم که آستانه قابل قبول برای آن 0/70 است (برای جزئیات بیشتر ر.ک. هیرینگا، 2004:170). وجه تشابه هر دو مطالعه، پایگاه دادگان و قلمرو مکانی همسان است، تا نتایج‌شان قابل مقایسه و اعتبارسنجی باشد. به این اعتبار، قلمرو مکانی پژوهش، حد فاصل کرانه جنوب‌خاوری دریای مازندران و دامنه شمالی البرز خاوری است؛ «البرز خاوری از دره فیروزکوه تا گرگان‌رود و مرز خراسان کشیده شده است» (جعفری، 1368:37). بنابراین، محدوده قلمرو پژوهش - یعنی حد فاصل خلیج گرگان تا نزدیک مرکز مازندران - طبق تقسیمات کشوری از خاور به باختر به شهرستان‌های بهشهر (شامل شهرهای بهشهر، گلوگاه و رستمکلا)، نکا، ساری (شامل شهرهای ساری، کیاسر، فریم، و سورک)، قائمشهر و جویبار بخش می‌شود. موقعیت جغرافیایی و محدوده سایت‌ها در تصویر (1) نمایانده شده است. در پایگاه دادگان پژوهش برای هر یک از این سایت‌ها

<sup>1</sup> برای اشاره به موقعیت جغرافیایی مورد نظر از این پس از کلیدواژه سایت (Site) استفاده می‌کنیم که بسته به گونه زبانی مورد بررسی برابر است با نقطه جغرافیایی، نقطه جمعیتی، مکان گردآوری، یا حوزه تقسیمات جغرافیایی که آن گونه‌زبانی بدانجا تعلق دارد یا گویش‌ور اهل آنجاست. در این مقاله به دلیل رایج تر بودن و گویاتر بودن این اصطلاح انگلیسی به ناچار از معادل‌سازی فارسی آن اجتناب شده است.

<sup>2</sup> W. Heeringa

<sup>3</sup> acoustic features

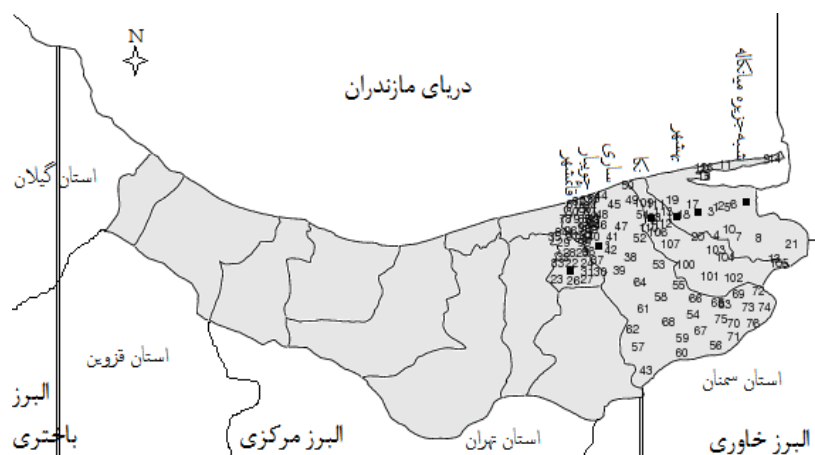
<sup>4</sup> J. Nerbonne & W. Heeringa

<sup>5</sup> T. Leinonen

<sup>6</sup> M. Wieling

<sup>7</sup> Cronbach's  $\alpha$  method

یک مصاحبه با یک گویشور منتخب وجود دارد. چون درج نام سایت‌ها در نقشه خواندن آنرا غیر ممکن می‌ساخت، نام آنها در جدول پیوست (1) فهرست شده است. خطوط شکسته پررنگ سیاه نشان‌دهنده مرز شهرستان بر اساس تقسیمات کشوری هستند.



تصویر 1) نقشه راهنمای قلمرو مکانی پژوهش، زمین ساخت رشته کوه البرز، تقسیمات کشوری استان مازندران، پراکنش و شماره 113 سایت مورد پژوهش در دامنه شمالی البرز خاوری را نشان می‌دهد. نام سایت‌ها در پیوست 1 درج شده است.

مواد زبانی پژوهش از طرح *ملی اطلس زبانی ایران* وام گرفته شد. گویش ور مورد مصاحبه می‌بایست بومی، بی‌سواد و یک‌زبان، ساکن دائم محل و دست کم میان‌سال، و فاقد مشکلات گفتاری می‌بود و شیوا و بی‌پروا سخن می‌گفت<sup>1</sup>. داده‌های زبانی زبان مازندرانی مورد تحلیل در مقاله پیش‌رو، از پاییز 1383 الی بهار 1384 به صورت هدایت شده و پرسشنامه‌ای تهیه گردید و تکمیل شد. از جنبه محتوایی کوشش شد واژه‌ها و جملات مندرج در این پرسشنامه ضمن در بر داشتن واژگان پایه، در حد امکان، با تنوعات جغرافیایی و فرهنگی پهنه ایران سازگاری داشته باشند.

در گام بعد، برای انجام تحلیل‌های گویش‌سنجی، معادل بومی 62 مدخل<sup>2</sup> واژگانی نخست پرسشنامه (جمعاً 7006 معادل) برای هر یک از 113 سایت قلمرو پژوهش از این داده‌ها استخراج شد. این واژگان در جدول پیوست (2) فهرست شده‌اند و شامل روابط خویشاوندی، اعضا بدن، اسامی حیوانات، عناصر طبیعت، ابزار و وسایل، و مفاهیم انتزاعی است. میانگین سنی گویشوران 42 سال، متوسط سواد آنها در حد آموزش ابتدایی و 34٪ آنها زن بودند. برای پردازش داده‌ها این واژگان باید آوانگاری می‌شد. بنابراین در گام سوم این کار با استفاده از 38 نویسه آی.پی.ای (جمعاً 29394 نویسه) آوانگاری شد. این پژوهش در آوانگاری رفتار توزیعی یا رابطه همنشینی آواها را مد نظر ندارد؛ بلکه هر آوا به طور جداگانه مورد بررسی قرار گرفته و ویژگی‌های آوایی آن فارغ از نقش تقابلی و معیار ممیز معنا بودن تعریف می‌شود.

<sup>1</sup> برای کسب اطلاعات کامل اندر مستندسازی، پایگاه داده‌ها، و نقشه این طرح ر.ک. پرمون (1385).

<sup>2</sup> مدخل (linguistic item or parameter) دال بر متغیر زبانی - اعم از آوا، واج، تکواژ، یا واحد نحوی - است که پیکره دادگان را می‌سازد و مبنای بررسی گونه مورد مطالعه و یا محاسبه شاخص فاصله زبانی قرار می‌گیرد.

## 3-1. تحلیل انبوهه

روش‌های کمی در گویش‌شناسی جغرافیایی اصطلاحاً روش‌های تحلیل انبوهه نامیده شده‌اند، به این دلیل که به جای توزیع ویژگی‌های زبانی منفرد، تفاوت بین چندین متغیر را محاسبه می‌کند و سپس برآیند نسبی این تفاوت‌ها در کل دادگان معین می‌شود. انگاره اصلی آن است که حجم بالای متغیرهای متعدد تصویر درست‌تری از ارتباط بین زبانگونه‌ها بدست می‌دهد. بسیاری از روش‌های تحلیل تک‌پارامتر، غیر رایانه‌ای هستند و تعداد کمی از ویژگی‌های زبانی را مد نظر قرار می‌دهند؛ بنابراین، شاید نتوان آنها را مطالعات جامعی در نظر گرفت. وانگهی، در رهیافت تحلیل انبوهه ضمن تأکید بر تعدد ویژگی‌های زبانی منتخب، همه آنها به شیوه‌ای یکسان مورد تحلیل، و برآیند یافته‌های همه پارامترها مبنا قرار می‌گیرد.

ملایی‌پاشایی (1393) با تحلیل انبوهه داده‌های گویشی زبان مازندرانی مستخرج از طرح ملی اطلس زبانی ایران به محاسبه شاخص فاصله زبانی 425 سایت در دامنه شمالی البرز با اعمال الگوریتم لونیشتین<sup>1</sup> پرداخته است. وی بر اساس شاخص فاصله زبانی<sup>2</sup>، خوشه‌نگار<sup>3</sup>، گروه‌های گویشی اصلی، و منطقه‌های گویشی را در قالب نقشه‌های پرتوی<sup>4</sup> بدست داده است. در آن پژوهش حد فاصل دامنه شمالی البرز و کرانه جنوبی دریای مازندران به سه گروه گویشی اصلی مازندرانی گالشی، مازندرانی خاوری، و مازندرانی باختری بخش شده است. زبانگونه خاوری خود به دو زیرگروه کوچکتر با اختلاف شاخص فاصله زبانی 0/01 بخش می‌شود: نخست شهرستان‌های گلوگاه، بهشهر، نکا، ساری، قائمشهر، و دو دیگر شهرستان‌های سوادکوه، بابلسر، بابل، فریدون‌کنار، محمودآباد، آمل، نور، و نوشهر. سپس، نجفیان و همکاران (1393) به همان روش به بررسی ریزبینه‌تر زبانگونه خاوری واقع در دامنه شمالی البرز خاوری در استان مازندران - که از لحاظ تقسیمات کشوری شهرستان‌های جویبار، قائمشهر، ساری، نکا، و بهشهر را در بر می‌گیرد - مبادرت ورزیده‌اند.

الگوریتم لونیشتین شیوه‌ای است برای مقایسه دو زنجیره (مثلاً زنجیره علائم آی.پی.ای) و محاسبه میزان تفاوت یا فاصله آنها. به دیگر سخن، شاخص فاصله زبانی، شاخصی کمی است که میزان تفاوت گونه‌های زبانی را به صورت فازی در بازه مجموعه اعداد طبیعی نشان می‌دهد. برای مقایسه دو گویش، تلفظ‌های واژگان مشابه در دو گویش جفت جفت با هم مقایسه می‌شوند، و سپس میانگین همه فاصله‌های زبانی لونیشتین بین واژگان به عنوان فاصله زبانی لونیشتین دو گویش محاسبه می‌شود. در این الگوریتم تفاوت بین دو رشته به صورت کمینه هزینه عملیات‌های نگاشت یک رشته به رشته‌ای دیگر - یعنی درج یک عنصر درون رشته، حذف یک عنصر از رشته، و جایگزینی یک عنصر از رشته با یک عنصری دیگر - تعریف می‌شود. برای مثال در جدول (1) در تلفظ معادل «اجاق» هزینه نگاشت رشته /cblə/ به رشته /calə/ به خاطر یک عملیات جایگزینی عنصر دوم یعنی واکه /B/ با واکه /a/ برابر با 1 است. در چنین نسخه‌های ساده این الگوریتم هزینه همه عملیات‌ها برابر 1 در نظر گرفته می‌شود. می‌توان برای عملیات‌های مختلف، همچنین عناصر مختلف مثلاً واکه‌ها یا علائم زیرزنجیری هزینه‌های مختلف تعریف کرد.<sup>5</sup>

<sup>1</sup> Levenshtein algorithm

<sup>2</sup> linguistic distance index

<sup>3</sup> dendrogram

<sup>4</sup> beam map

<sup>5</sup> برای آگاهی بیشتر اندر الگوریتم‌های تحلیل انبوهه و خوشه‌بندی داده‌های خطی ر.ک. جین و دوبر (1988)

جدول 1) هزینه نگاشت رشته /cɛlə/ به رشته /calə/ معادل «اجاق» به شیوه محاسبه ساده فاصله لونیشتین برابر با 1 است.

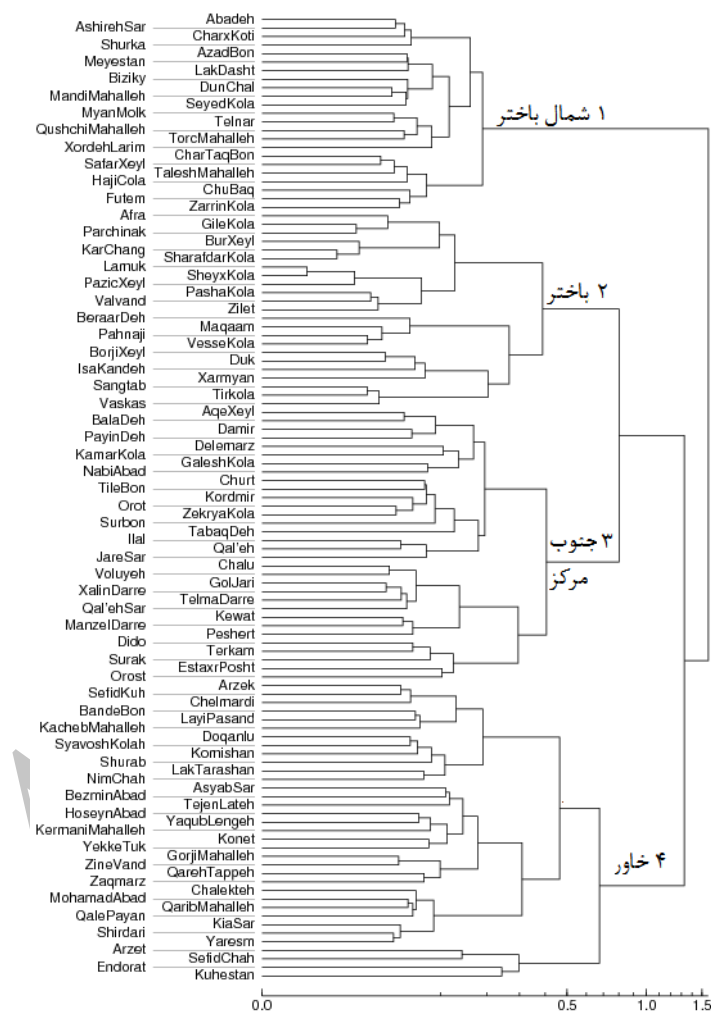
c	ɛ	l	ə	
c	a	l	ə	
0	1	0	0	1

در نجفیان و همکاران (1393) محاسبه فاصله لونیشتین آوانگاشت تلفظ 62 واژه در 133 سایت - یعنی  $62 \times 113 = 006,7$  رشته - در پایگاه دادگان پژوهش به شیوه بالا بین هر دو سایت تکرار شد. بنابراین، در نهایت  $(113 \times (113 - 1)) = 328,6$  شاخص فاصله لونیشتین بدست آمد. نخستین شیوه تحلیل بر اساس این شاخص‌ها، رسم خوشه‌نگار است. هدف از خوشه‌نگاری نیز تمیز گروه‌های گویشی اصلی است. این روش زبان‌گونه‌های شبیه‌تر به هم را به صورت سلسله‌مراتبی در گروه‌های گویشی خوشه‌بندی می‌کند. برای این کار نخست هر سایت خود یک گروه در نظر گرفته می‌شود. سپس دو گروهی که کمترین شاخص فاصله لونیشتین (بیشترین شباهت) را دارند در هم ادغام و تشکیل یک گروه جدید می‌دهند. فاصله این گروه جدید با گروه‌های قبلی محاسبه می‌شود. آنگاه این گروه جدید با گروه دیگری ادغام می‌شود و این فرآیند تا زبان‌گونه آخر ادامه می‌یابد تا دست آخر به خوشه‌ای از گروه‌های گویشی برسیم. در نمودار (1) سایت‌ها بر اساس شاخص فاصله لونیشتین بدست آمده به روش بالا، با الگوریتم وارد گروه‌بندی شده‌اند. به علت محدودیت نرم‌افزاری مجبور به استفاده از عدد و قلم لاتین شده‌ایم. در نمودار (1) پنج گروه گویشی اصلی قابل تفکیک است. این تعداد با نواحی کانونی نقشه شبکه‌ای تصویر (3) برابر است. گروه‌های گویشی 1، 2، و 3 در مقابل گروه‌های گویشی 4 و 5 تمایز باختری-خاوری را رقم می‌زنند. در گروه‌های باختری نخست سایت‌های شمال باختری حول شهرستان جویبار با فاصله درون‌گروهی حدود 0/25 با بیشترین فاصله از بقیه سایت‌ها (1/7) یک گروه گویشی تشکیل می‌دهند. سپس گروه گویشی 3 حول شمال شهرستان ساری و شهرستان قائمشهر خوشه‌ای دیگر را شکل می‌دهند. وانگهی، اگر اندک پراکنش گروه گویشی دوم در مناطق مرکزی نیمه شمالی را در نادیده انگاریم، می‌توان گفت این گروه گویشی تمایز شمال و جنوب را بازمی‌نماید، هر چند به لحاظ کمی این تمایز (0/43) اندک است. خوشه پنجم تنها با چهار عضو (سایت‌های آرزت، سفیدچاه، اندرات، و کوهستان) گروهی متفاوت را تشکیل می‌دهد. با بررسی مجدد مشخصات گویشوران خوشه‌های چهارم و پنجم در می‌یابیم این افراد گونه‌زبانی خود را مازندرانی گالشی معرفی کرده‌اند؛ در واقع گونه گالشی در هیچ گروه دیگری جای نمی‌گیرد. درج شاخص کمی فاصله بین گروه‌ها برای تشخیص گروه‌های گویشی اصلی ضروری است؛ چرا که تنوع و دسته‌بندی زبانگونه‌ها می‌تواند به موارد کوچکتر و کوچکتر ادامه یابد تا سرانجام به گفتار فردی برسد.

نقشه شبکه‌ای<sup>1</sup> تصویر (2) نیز مستقیماً بر اساس شاخص فاصله لونیشتین رسم می‌شود؛ مناطقی که به لحاظ زبانی همگن‌ترند و یک منطقه گویشی را شکل می‌دهند به صورت مجموعه‌ای از خطوط تیره ظاهر می‌شوند، و مرز بین آنها به صورت نوارهای سفید یا با خطوط روشن مشخص می‌شود. وانگهی، در نقشه بازنمودی شبکه‌ای تنها سایت‌های مجاور به هم متصل می‌شوند. در این نوع نقشه چون تنها تفاوت محلی سایت‌های مجاور نمایانده می‌شود، نسبت به نقشه پرتوی جامع نمی‌نماید و الگوهای کلی قلمرو پژوهش ممکن است پنهان بماند؛ منتها تصویری روشن‌تر از تفاوت‌های منطقه‌ای

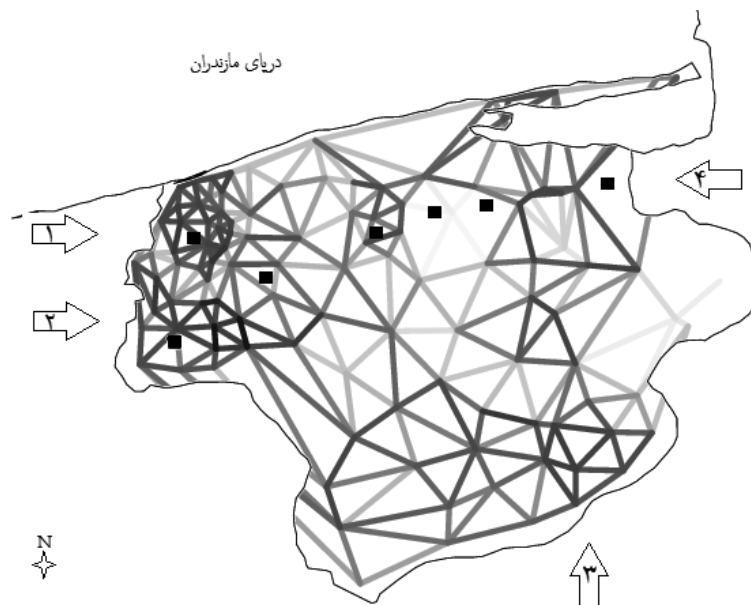
<sup>1</sup> network map

بدست می‌دهد، بویژه زمانی که تراکم سایت‌ها بالا است. نقشه شبکه‌ای تصویر (2)، چهار ناحیه کانونی<sup>1</sup> را نشان می‌دهد: نخست ناحیه کانونی شمال باختری حول نواحی شهری ساری و قائمشهر؛ دو دیگر در شمال باختری حول ناحیه شهری جویبار؛ سه دیگر در شمال خاوری حول نواحی شهری بهشهر و گلوگاه، و سرانجام در ناحیه کوهستانی جنوب‌خاوری (نجفیان و همکاران، 1393). در ناحیه کانونی خطوط تیره شبکه‌ای همگن می‌سازند، یعنی زبانگونه‌ها شبیه‌اند. با فاصله از کانون، در ناحیه پیرامونی، به تدریج از شباهت‌ها کاسته شده و خطوط کمرنگ‌تر می‌شوند تا مرزهای زبانی به شکل نوارهای سفید آشکار شود. بیشتر نواحی کانونی حول نواحی شهری شکل گرفته‌اند؛ بررسی معناداری ارتباط بین نواحی کانونی و نواحی شهری شایسته پژوهشی دیگر در حوزه زبان‌شناسی اجتماعی است.



نمودار (1) خوشه‌نگار 113 سایت مورد پژوهش با الگوریتم وارد 5 گروه گویشی اصلی بدست می‌دهد (نجفیان و همکاران، 1393).

<sup>1</sup> central representative area



تصویر 2) نقشه بازنمودی شبکه‌ای چهار ناحیه کانونی شمال باختری، باختری، شمال خاوری، و ناحیه کوهستانی جنوب خاوری را آشکار می‌سازد (آلفای کرنباخ=0/81).

### 2-3. تحلیل بسامد آوایی

نخستین بار هوپنبروورس و هوپنبروورس (1984) روش‌های بسامد حروف<sup>1</sup> و بسامد آوایی<sup>2</sup> را معرفی کردند. این روش‌ها پیکره‌بنیان هستند. در روش بسامد حروف برای مقایسه دو گویش میزان فراوانی حروف را در پیکره مربوطه دو گویش می‌سنجند. از آنجا که همه پیکره‌ها یک اندازه نیستند بسامدها به درصد تبدیل می‌شود. شاخص فاصله دو گویش برابر است با مجموع تفاوت‌های بسامد حروف آن دو گویش. از معایب این روش نخست این است که همیشه هجی متفاوت کلمات متناظر با تلفظ‌های مختلف نیست و بالعکس؛ مثلاً *academie* در هلندی در مقابل *academy* در فریزی املائی متفاوت اما تلفظ یکسان دارند و بالعکس *we* انگلیسی در مقابل *we* هلندی املائی یکسان و تلفظ متفاوت دارند. نرین و هیرینگا (1988) نیز از این روش بهره جستند.

بنابراین برادران هوپنبروورس (1:2001) روشی آواشناختی را اتخاذ کردند. در روش بسامد آوایی، در هر پیکره فراوانی آواها مشخص می‌شود و سپس مانند روش بسامد حروف شاخص فاصله زبان‌ها از یکدیگر محاسبه می‌گردد. در این بررسی پیکره زبانی از داستان *خورشید و باد شمال*<sup>3</sup> که در جزوه انجمن بین‌المللی آوانگاری (1949)<sup>4</sup> با استفاده از علائم آوانگار بین‌المللی (IPA) به 51 زبان آوانویسی شده ایجاد گردیده است. در هر متن بسامد آواها مشخص شد. چون همه متن‌ها یک اندازه نبودند قاعدتاً بسامد نسبی آنها محاسبه شد. شاخص فاصله زبانی بین دو زبان برابر است با مجموع قدر مطلق تفاضل بین بسامد نسبی هر یک از آواهای مربوطه. نرین و هیرینگا (2001) این روش را بسط دادند و نسبت نویسه‌های هر کلمه را نیز وارد محاسبات نمودند.

<sup>1</sup> letter frequency method

<sup>2</sup> phone frequency method

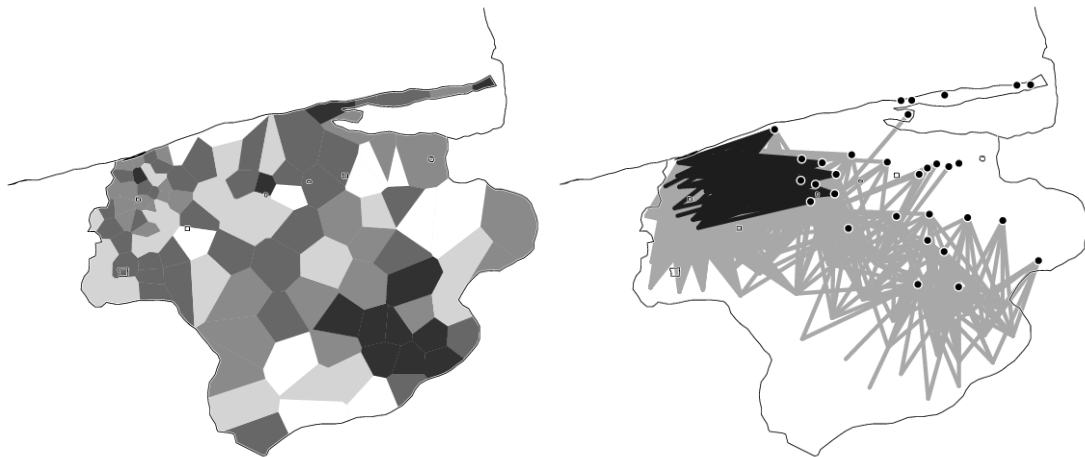
<sup>3</sup> The North Wind and the Sun

<sup>4</sup> The Principles of the International Phonetic Association



در چارچوب این روش، ملایی پاشایی (1393: 74-88) ابتدا بسامد مطلق هر واکه در کل پیکره را محاسبه کرد. همانگونه که در بخش دوم اشاره شد، در این پژوهش پیکره زبانی مشکل است از آوانگاشت معادل 62 مدخل واژگانی (جمعا 7006 معادل) در 113 سایت قلمرو پژوهش. سپس بسامد نسبی هر واکه در هر منطقه گویشی محاسبه و بر مبنای درصد بهنجار شد. این یافته‌ها در وهله نخست علت تمایزات منطقه‌ای را بر اساس متغیر دستگاه واکه‌ای را تبیین می‌کند. نیز این بسامد یکی از معیارهای اعتبارسنجی یافته‌های تحلیل انبوهه دامنه شمالی البرز با اعمال الگوریتم لونشتین است. همچنین ضریب تأثیر هر واکه در ایجاد یک منطقه گویشی برابر است با قدر مطلق میانگین بسامد نسبی هر آوا نسبت به بسامد آن در کل پیکره. در پیکره زبانی پژوهش حدود 40٪ از آواها در هر یک از منطقه‌های گویشی چهارگانه واکه است، و در هیچ یک از این مناطق جایگزینی واکه با همخوان رخ نداده است؛ بنابراین مقایسه بسامد واکه‌ها با یکدیگر روا است. در پژوهش پیش‌رو بر اساس پایگاه دادگان برای نظام واکه‌های مازندرانی در این منطقه سه واکه پیشین [a, i, u]، سه واکه پسین [ɒ, ɔ, u]، و دو واکه مرکزی [ɐ و ə] بازشناخته‌ایم که تقابل [I] با [i] یادآور تقابل کشش واکه‌ای از دوره میانه زبان‌های ایرانی است. در اینجا بسامد نسبی واکه‌ها در نقشه‌های تفسیری پرتوی نمایانده می‌شود تا از این طریق نواحی گویشی مشخص شود. در نقشه پرتوی بین هر دو سایت یک خط رسم می‌شود که تیرگی بیشتر این خط<sup>1</sup> واصل متضمن بسامد بیشتر است. برای خواندن نقشه‌ها از نقشه راهنمای تصویر (1) کمک بگیرید.

واکه /I/ واکه‌ای نزدیک به بسته، نزدیک به پیشین، گسترده، و کوتاه است که در 235 مدخل گویشی از پیکره همچون /dɪr/himə/ /niyaz/ /nina/ /ʃirin/ /dɪznən/ /xɪʃ/ /miʃ/ /ʃatpɪyər/ و /baldɪn/ تظاهر پیدا کرده است. بسامد مطلق این واکه با فراوانی بیشتر در جنوب و نیز در نواحی گذار<sup>1</sup> همراه است (ر.ک. تصویر (3) راست)؛ همه مناطق گویشی - جز اندک تفاوت مثبت در شمال باختری و منفی در خاور - بسامد نسبی بهنجار نسبتاً یکسانی دارد (ر.ک. تصویر (3) چپ). از این روی بین واکه‌های پیکره زبانی این واکه با کمترین ضریب بر تفاوت‌های گویشی منطقه‌ای - یعنی 0/955 - سبب تمیز گویشی خاوری-باختری منطقه می‌شود.



تصویر (3) نقشه باز نمودی بسامد مطلق (راست) و نقشه تفسیری بسامد نسبی بهنجار (چپ) واکه /I/

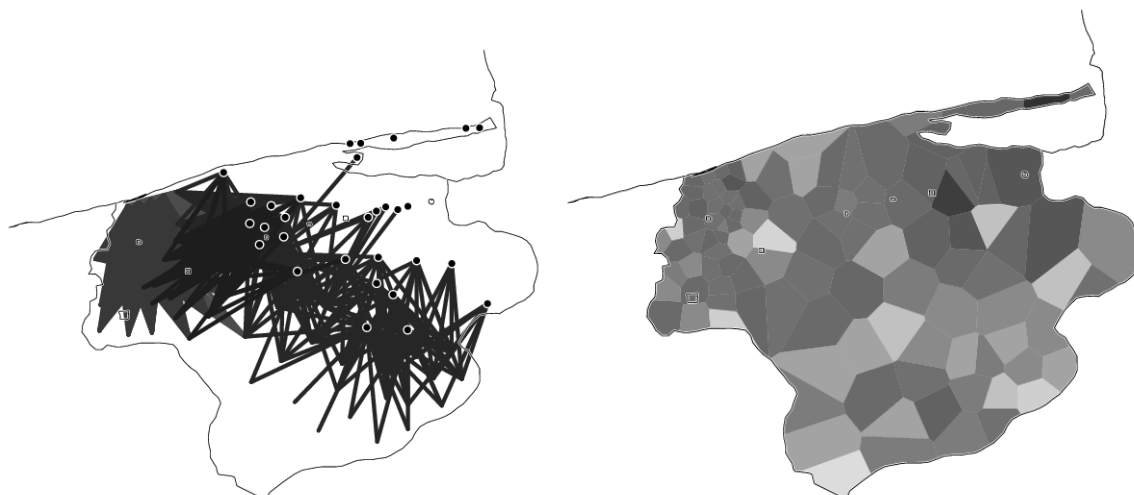
<sup>1</sup> طبق مشاهدات انجام شده، در مناطق مرزی گویش‌ها، هنگامی که یک متغیر زبانی جای خود را به متغیر دیگر می‌دهد، این گذار آنی (relic transition) نیست بلکه تدریجی (gradual transition) است. بسامد متغیرهای زبانی در نقاط مختلف این منطقه گذار (transition zone) نامنظم است و لذا الگوی مشخصی قابل تشخیص نیست.

واکه /i/ واکه‌ای پیشین، بسته، گسترده، و کشیده است که در 1000 مدخل گویشی از پیکره همچون /dɒyi/ /pandir/ /dir/ /fi:/ /nazdic/ /miʃ/ /mardi/ /ɟati/ /rica/ /saɲin/ /çəmɒri/ /ʔomi/ پراکنده است (ر.ک. تصویر (4) راست)؛ بسامد نسبی بهنجار این واکه از خاور به سوی باختر، بویژه در منطقه گویشی دوم (باختری) افزایش قابل توجهی دارد (ر.ک. تصویر (4) چپ). این واکه در منطقه گویشی جنوب‌مرکز بیشتر از دیگر مناطق دچار کشش واکه‌ای می‌شود (مانند /i:/). بدین ترتیب از بین واکه‌های پیکره زبانی این واکه با بیشترین ضریب بر تفاوت‌های گویشی منطقه‌ای - یعنی 2/340 - باعث تفاوت گویشی از خاور به سوی باختر و متمایز ساختن منطقه باختری شده است.



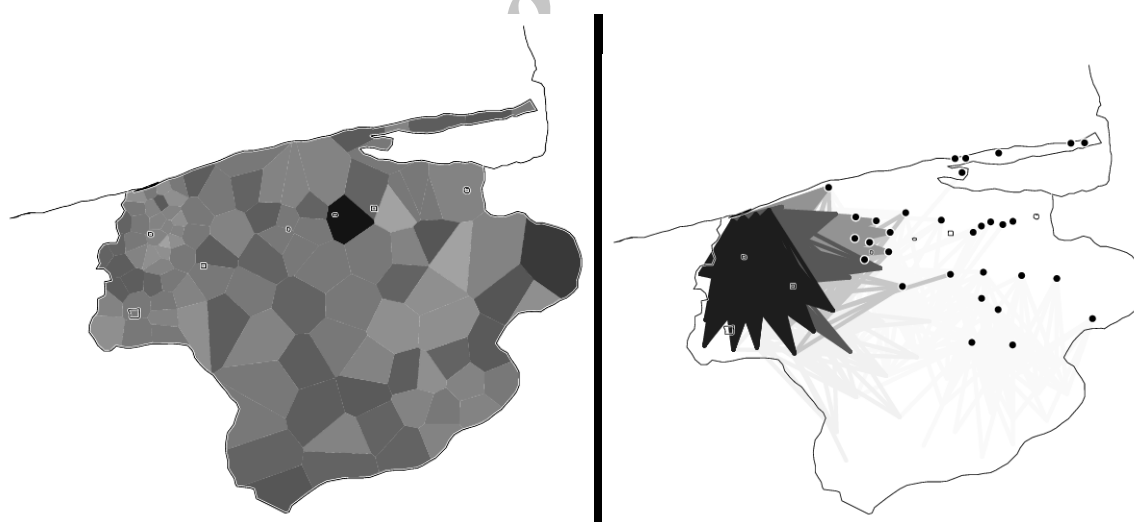
تصویر 4) نقشه بازنمودی بسامد مطلق (راست) و نقشه تفسیری بسامد نسبی بهنجار (چپ) واکه /i/

واکه /e/ واکه‌ای مرکزی، نزدیک به باز، نیم گسترده، و کوتاه است که در 1110 مدخل گویشی از پیکره همچون /dɛccə/ /tɛc/ /ɟɛft/ /dɛtər/ /tɛlɒ/ /ɣəzɛlɒn/ /cɛrc/ /sɛrə/ /cɛlə/ /cutər/ /ʔɒbɛssən/ /ɟɛl/ تظاهر پیدا کرده است. بسامد مطلق این واکه در سایت‌های شمالی بویژه خاوری بیشتر است (ر.ک. تصویر (5) راست)؛ همانگونه که تصویر (5) چپ نشان می‌دهد، بسامد نسبی بهنجار این واکه به ترتیب در منطقه‌های گویشی خاور و شمال باختر بیشتر است، و سپس سبب تمایز منطقه‌های گویشی جنوب‌مرکز و باختر می‌شود. از بین واکه‌های پیکره زبانی این واکه ششمین ضریب بر تفاوت‌های گویشی منطقه‌ای - یعنی 1/412 - را دارا است.



تصویر 5) نقشه بازنمودی بسامد مطلق (راست) و نقشه تفسیری بسامد نسبی بهنجار (چپ) واکه /a/.

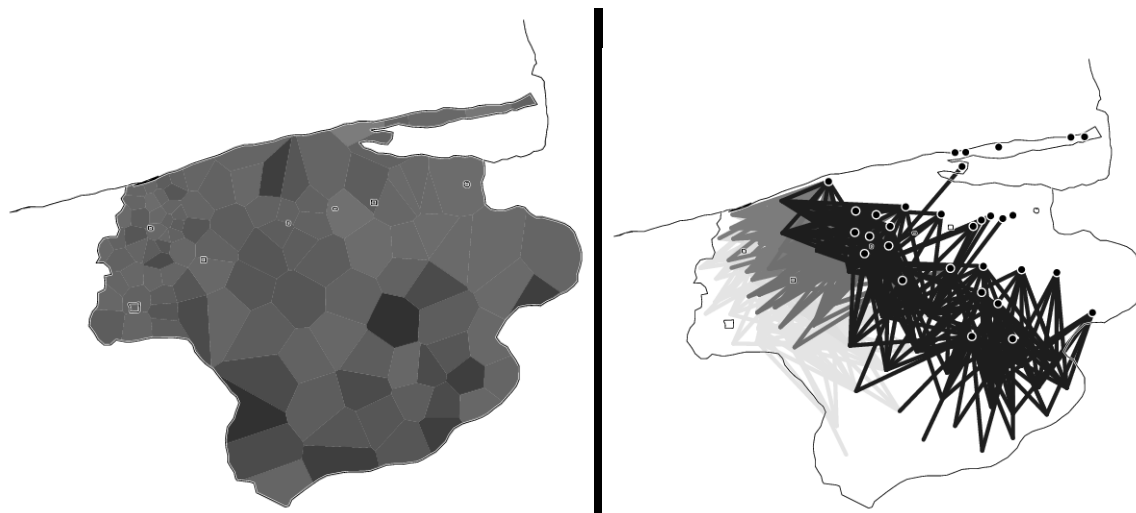
واکه /a/ واکه‌ای پیشین، باز، گسترده، و کوتاه است که در 1830 مدخل گویشی از پیکره همچون /valɤ/ مطلق این واکه در بین سایت‌ها نسبتاً همگن و اندکی پراکنده است (ر.ک. تصویر (6) راست)؛ از بین واکه‌های پیکره زبانی این واکه با سومین ضریب بر تفاوت‌های گویشی منطقه‌ای - یعنی 1/551- سبب تفکیک منطقه گویشی شمال‌باختر است.



تصویر 6) نقشه بازنمودی بسامد مطلق (راست) و نقشه تفسیری بسامد نسبی بهنجار (چپ) واکه /a/.

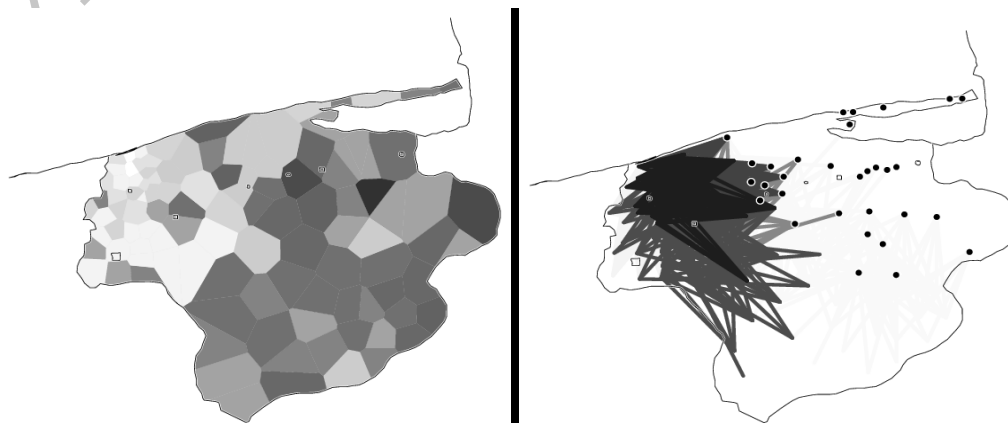
واکه /ə/ واکه‌ای خنثی، میانی، مرکزی، و کوتاه است که در 2797 مدخل گویشی از پیکره در هسته و پایان هجای معادل منطقه‌ای مدخل‌های واژگانی همچون /vaçə/ /xɒxər/ /pəsər/ /cərc/ /vɒrəf/ /cələ/ /sətɒrə/ در بین سایت‌های خاوری کمتر و در سایت‌های مرکزی و جنوبی بیشتر است (ر.ک. تصویر (7) راست). بسامد نسبی بهنجار این واکه، با هشتمین ضریب بر تفاوت‌های گویشی منطقه‌ای یعنی 1/026، به علت هماهنگی واکه‌ای مانند

افراشتگی واکه در /nəməʊz/ → /nɒməʊz/، افراشتگی واکه همراه با مرکزی‌شدگی /dəhɔn/ → /dɔhun/ و پسین‌شدن هسته /ʔəftɔb/ ← /ʔɔftɔb/ ← /ʔɒftɔb/ متمایز شدن منطقه گویشی جنوب‌مرکز گشته است. در ادبیات پژوهش‌های زبان‌شناختی واکه‌های /ə/، /ɐ/ و /e/ در نظام آوایی مازندرانی مورد بحث است. بر اساس داده‌های پژوهش حاضر تمایز بین /ə/ و /ɐ/ همچون پایان واژه‌های /sɐ/ عدد سه و /sə/ سیب، ممیز معناست.



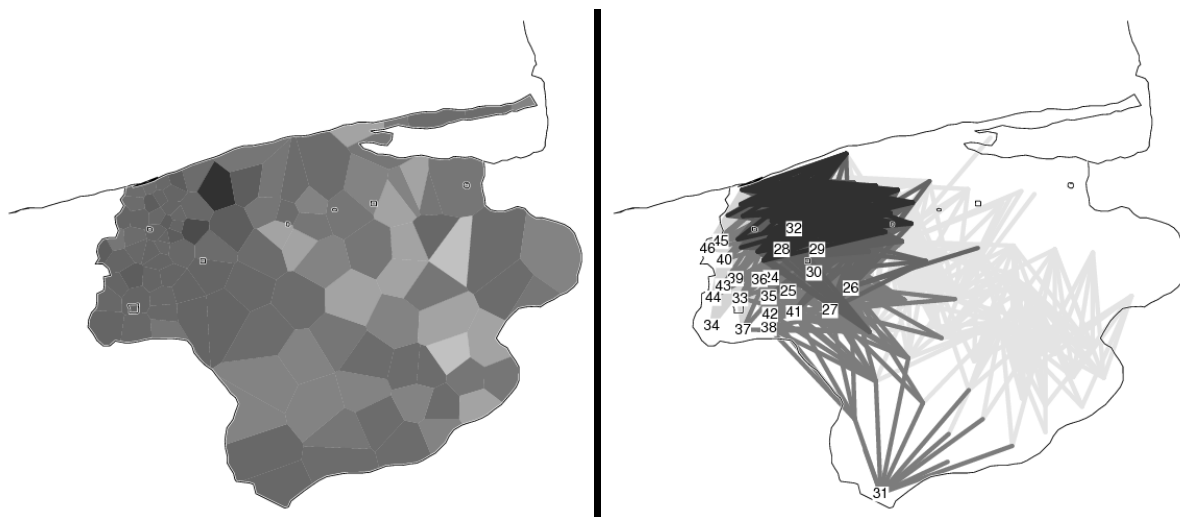
تصویر 7) نقشه بازنمودی بسامد مطلق (راست) و نقشه تفسیری بسامد نسبی بهنجار (چپ) واکه /ə/

واکه /ɔ/ واکه‌ای پسین، نیم باز، نیم گرد، و کوتاه است که در 516 مدخل گویشی از پیکره در معادل منطقه‌ای مدخل‌های واژگانی همچون /ɔrad/ /ʔɔ:s/ /ʔɔftɔb/ /xɔnə/ /ʔɔlɔm/ /bɔdɔm/ /jɔ/ /ʔɔ:/ /ʔɔ/ همچون /zəvɔn/ و /dəhɔn/ تظاهر پیدا کرده است. بسامد مطلق این واکه در بین سایت‌های خاوری کمتر و در سایت‌های مرکزی و جنوبی بیشتر است (ر.ک. تصویر 8) (راست). بسامد نسبی بهنجار این واکه در منطقه گویشی شمال باختر بسیار کم، در باختر کم، و در خاور بیشترین است (ر.ک. تصویر 8) (چپ). بین واکه‌های پیکره زبانی این واکه با چهارمین ضریب بر تفاوت‌های گویشی منطقه‌ای - یعنی 1/522 - سبب تمیز گویشی شمال‌باختر از جنوب‌مرکز و خاوری می‌شود.



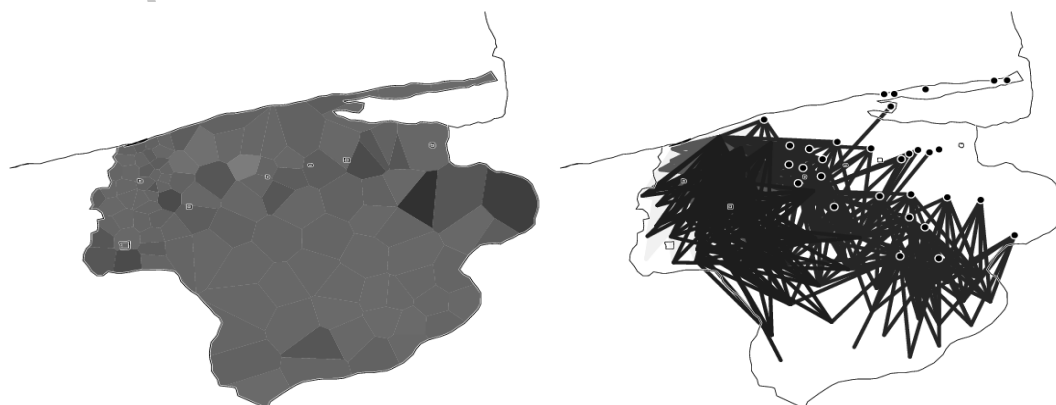
تصویر 8) نقشه بازنمودی بسامد مطلق (راست) و نقشه تفسیری بسامد نسبی بهنجار (چپ) واکه /ɔ/

واکه /u/ واکه‌ای پسین، بسته، گرد، و کشیده است که در 1131 مدخل گویشی از پیکره در معادل‌هایی همچون /çu:/ و /fu/ /ɟu/ /mɒhuri/ /ru:z/ /ʔamu/ /zəvun/ /cutər/ /xun/ /dɒhun/ /nun/ /ɟu:/ تظاهر پیدا کرده است. بسامد مطلق این واکه در بین سایت‌های خاوری کمتر و در سایت‌های مرکزی و باختری بیشتر است (ر.ک. تصویر (9) راست). بیشترین بسامد نسبی بهنجار این واکه به ترتیب در منطقه‌های گویشی شمال‌باختر، باختر، جنوب‌مرکز و خاور است (ر.ک. تصویر (9) چپ). بدین ترتیب بین واکه‌های پیکره زبانی این واکه با پنجمین ضریب بر تفاوت‌های گویشی منطقه‌ای - یعنی 1/509 - سبب تمیز گویشی با محور شمال‌باختر جنوب‌خاوری می‌شود.



تصویر (9) نقشه بازنمودی بسامد مطلق (راست) و نقشه تفسیری بسامد نسبی بهنجار (چپ) واکه /u/

واکه /ɒ/ واکه‌ای پسین، باز، گرد، و کشیده است که در 2736 مدخل گویشی از پیکره همچون [ɸəɒm]، /mɒ/ و /dɒr/ /vɒ/ /təɒ/ /dɒyi/ /nəɒvz/ /bərɒr/ /xɒl/ /vɒrəɸ/ /ciɟɒ/ /mɒr/ تظاهر پیدا کرده است. بسامد مطلق این واکه با فراوانی بیشتر در شمال و نیز در نواحی گذار همراه است (ر.ک. تصویر (10) راست). همه مناطق گویشی - جز اندک تفاوت مثبت در شمال و منفی در خاور - بسامد نسبی بهنجار نسبتاً یکسانی دارد (ر.ک. تصویر (10) چپ). از این روی بین واکه‌های پیکره زبانی این واکه با هفتمین ضریب بر تفاوت‌های گویشی منطقه‌ای - یعنی 1/068 - سبب تمیز گویشی شمال-جنوب منطقه می‌شود.



تصویر (10) نقشه بازنمودی بسامد مطلق (راست) و نقشه تفسیری بسامد نسبی بهنجار (چپ) واکه /ɒ/

#### 4. یافته‌های پژوهش و اعتبار سنجی

در مقاله حاضر روش‌هایی برای یافتن تنوعات زبانی و نمایش پراکندگی زبانی بررسی شده است که از طریق سنجش همزمانی، محاسبه مشابهت‌ها و تحلیل انبوه تفاوت‌های زبانی ارائه شده، محقق می‌شود. یافته‌های پژوهش کنونی از برخی جهات با نتایج روش‌های سنتی همخوانی دارد و از جهاتی دیگر چشم‌اندازی نو را صورت می‌بندد. در واقع این بررسی‌ها اولین تلاش با رویکردی کمی و رایانشی برای بازشناسی گونه‌های زبان مازندرانی و بازنمایی چگونگی تأثیرگذاری جغرافیا بر گونه‌گونی زبانی است. گویش‌سنجی زبان مازندرانی و یافتن گروه‌های گویشی اصلی بر اساس داده‌های طرح ملی اطلس زبانی ایران دارای اهمیت بسزایی است؛ هم از این جهت که گروه‌های گویشی اصلی چشم‌انداز پیوستار گویشی را بدست می‌دهند، هم از آن جهت که بتوان روش‌شناسی رایانشی اخیر را با روش‌شناسی سنتی و نیز یافته‌های آن دو را با هم مقایسه کرد.

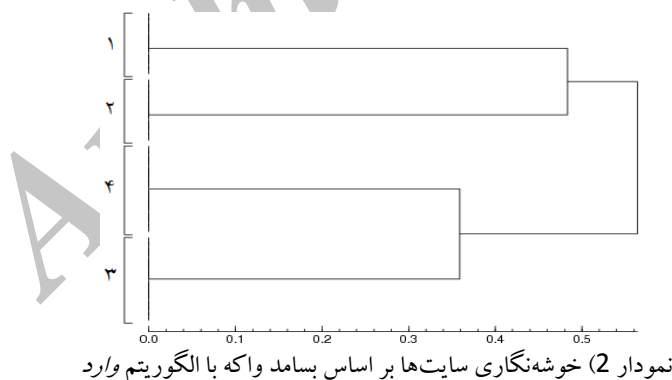
در روش تحلیل انبوه به دلیل قابلیت پردازش پیکره حجیم‌تر، می‌توان ضعف آماری روش‌های سنتی و تأثیر گویش فردی را کم‌رنگ‌تر ساخت. بررسی کل‌گرایانه و تحلیل انبوه دادگان در روش تحلیل انبوه ضمن تعیین جهت این تغییرات، بعلاوه میزان آن، و مهمتر، پیوستار گویشی و مناطق گذار را با استفاده از ترسیم نقشه‌های پرتوی بازنمایانده است. وانگهی، در نقشه پرتوی محلی تفاوت‌های محلی سایت‌های مجاور پررنگ‌تر می‌شود و بدین سان الگوهای کلی قلمرو پژوهش ممکن است پنهان بماند، لذا نقشه پرتوی جامع مکمل آن است. همانگونه که دیدیم، روش‌های تحلیل انبوه رایج در گویش‌شناسی ضمن آنکه مطالعات روش‌مند زبان‌شناختی و جامعه‌شناختی را بهبود می‌بخشد، جامع‌تر و مطمئن‌تر از روش سنتی است. تحلیل‌های انبوه به کار رفته در آن نتایجی بی‌طرفانه و بدون سوگیری عامل انسانی بدست می‌دهد که از طریق روش‌های سنتی قابل دست‌یابی نیست.

روش بسامد واکه روشی پیکره‌بنیاد است که نقش یک پارامتر زبانی را در کنار دیگر پارامترها بررسی می‌کند و وزن تأثیر آنرا نیز می‌سنجد. همچنین اطلاعات بدست آمده می‌تواند مبنای بررسی‌های رده‌شناختی قرار گیرد. نقشه‌های بازنمودی نماینده پراکنش جغرافیایی بسامد مطلق واکه‌ها به همراه نقشه‌های تفسیری نشان‌دهنده بسامد نسبی بهنجار واکه‌ها در بخش قبل چگونگی و سبب تظاهر منطقه‌های گویشی مختلف را بر اساس متغیر فراوانی واکه تبیین می‌کنند. به علاوه، ضریب تأثیر محاسبه شده و نیز وزن هر واکه را در بروز تفاوت‌های منطقه‌ای تفسیر می‌کند. روش بسامد مشخصه از آنجا که داده‌های جامع‌تر یک پیکره زبانی و همین‌طور یک واحد زبانی کامل را مورد بررسی قرار می‌دهد مقبول می‌نماید. وانگهی، باید در نظر داشت که دادگان پژوهش همه ساخت زبانی مورد نظر را در برگیرد. مثلاً در نمونه موردی ما همه واکه‌های زبان مازندرانی به تناسب باید در پیکره لحاظ شود؛ این کار دشوار می‌تواند با در نظر گرفتن واژگان پایه یا میزان تداول آنها لحاظ شود.

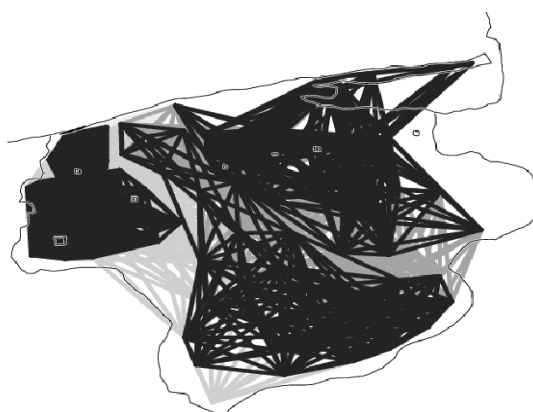
جدول (4) جمع‌بندی بررسی مقابله‌ای روش‌های گویش‌سنجی

خوانش نقشه	رسم نقشه	نوع نقشه	شیوه تحلیل	جمع‌آوری دادگان	
آسان و گویا تقسیم‌بندی قاطع	نسبتاً آسان نیازمند انتخاب فردی در تقسیم‌بندی اولیه نقشه پایه	تفسیری	انبوهه پیکره‌بنیان	بستگی به حجم آن دارد ولی نمی‌تواند محدود باشد	<b>بسامد واکه</b>
نسبتاً آسان و گویا نیازمند اطلاعات اولیه گویش‌شناختی	نسبتاً آسان نیازمند امکانات نرم‌افزاری خاص	تفسیری (محلّی و جامع)	انبوهه پیکره‌بنیان	اندکی دشوار باید تا اندازه ممکن گسترده باشد	<b>پرتوی</b>

خوشه‌نگار نمودار (1) را دوباره بر اساس بسامد واکه‌ها رسم می‌کنیم. هدف از خوشه‌نگاری تمیز گروه‌های گویشی اصلی است. در نمودار (2) زبان‌گونه‌های چهارگانه به صورت سلسله‌مراتبی در گروه‌های گویشی بازنمود یافته‌اند. مکان‌نگاری جغرافیایی این گروه‌ها نقشه بازنمودی تصویر (11) را به دست می‌دهد. همانگونه که پیداست این نقشه نیز گروه‌های گویشی چهارگانه را باز می‌نمایاند که منطبق با گروه‌های گویشی مورد نظر در نمودار (1) است و نواحی گذار به صورت نوارهای خاکستری نمایان گشته است. به کوتاه سخن، نقشه‌های بازنمودی و تفسیری حاصل از تحلیل بسامد واکه تظاهر چهار منطقه گویشی شمال باختری، باختری، شمال خاوری، و ناحیه کوهستانی جنوب خاوری را پشتیبانی می‌کند و نشان می‌دهد که واکه‌های پیشین /i/ و /a/، در مقابل واکه خنثی مرکزی /ə/، بیشترین ضریب تأثیر را در گوناگونی‌های منطقه‌ای دارند. نمودار (2) خوشه‌نگاری سایت‌ها بر اساس بسامد واکه با الگوریتم وارد تظاهر چهار گروه گویشی اصلی را تأیید می‌کند.



هر دو این روش‌ها نشان می‌دهند که گونه‌های مازندرانی قلمرو پژوهش به صورت قاطع و جزمی از هم جدا نمی‌شوند، بلکه یک پیوستار گویشی را تشکیل می‌دهند. به علاوه، اگر نقشه‌های گویشی بالا را با هم مقایسه کنیم، بارزترین تمایز بین گویش‌های مازندرانی تمایز بین خاور و باختر است. این یافته با داشته‌های سنتی همسو است. در قلمرو مکانی پژوهش، چند ناحیه مرکزی و نواحی پیرامونی و مناطق گذار وجود دارد که به اندازه بسیاری با هم همپوشانی دارند.



تصویر 11) نقشه‌های بازنمودی پرتوی چهار منطقه گویشی شمال باختری، باختری، خاوری، و ناحیه کوهستانی جنوب را تأیید می‌کند (آلفای کرنباخ=0/84)

در این پژوهش آلفای کرنباخ برابر 0/84 بدست آمده است، بنابراین تحلیل گوناگونی‌های گویشی حاصل از بسامد واکه‌ها روا و قابل اعتماد است. نیز در بررسی نجفیان و همکاران (1393) آلفای کرنباخ نزدیک و برابر 0/81 بوده است. در نتیجه یافته‌های هر دو پژوهش قابل مقایسه است و همپوشانی بالای نتایج به دست آمده بیانگر پایایی تحقیق و روایی آن است.

#### کتابنامه

- اسدپور، هیوا (1390). اطلس زبانی ریان‌های استان آذربایجان غربی: رویکردی رده‌شناختی-ادراکی در زبان‌شناسی جغرافیایی، پایان‌نامه کارشناسی ارشد، تهران: دانشگاه آزاد اسلامی واحد علوم و تحقیقات.
- پرمون، یدالله (1385). طرح ملی اطلس زبانی ایران: کتابچه جامع تدوین اطلس (مستندسازی، پایگاه داده‌ها، نقشه)، تهران: سازمان میراث فرهنگی، صنایع دستی و گردشگری؛ پژوهشکده زبان و گویش.
- جعفری، عباس (۱۳۶۸). کوهها و کوه‌نامه ایران، تهران: وارث.
- رستم‌بیک تفرشی، آتوسا (1393). اطلس گویشی و اندازه‌گیری فاصله‌های گویشی در استان همدان، جستارهای زبانی: (انتشار آنلاین 1393/06/22) [http://lrr.modares.ac.ir/article\\_11299\\_0.html](http://lrr.modares.ac.ir/article_11299_0.html)
- ملایی‌پاشایی، سیف‌اله (1391). تدوین اطلس گویشی استان گلستان- غرب، گزارش طرح پژوهشی، گرگان: دانشگاه پیام‌نور استان گلستان.
- ملایی‌پاشایی، سیف‌اله (1393). گویش سنجی رایانشی دامنه شمالی البرز مرکزی بر پایه الگوریتم لونیشتین: تدوین یک اطلس زبان‌شناختی، رساله دکتری تخصصی، تهران: دانشگاه پیام‌نور.
- نجفیان، آرزو، طیبه موسوی‌میانگانه، بلقیس روشن، و سیف‌اله ملایی‌پاشایی (1393). بازشناسی زبان‌شناختی و جغرافیایی گونه‌های زبان مازندرانی در حد فاصل خلیج گرگان تا مرکز مازندران با اعمال روش‌های گویش‌سنجی بر تفاوت‌های آوایی، جستارهای زبانی: (انتشار آنلاین 1393/06/09) [http://lrr.modares.ac.ir/article\\_11277\\_0.html](http://lrr.modares.ac.ir/article_11277_0.html)



- Heeringa, W. (2004). Measuring dialect pronunciation differences using Levenshtein Distance, PhD thesis, University of Groningen.
- Heeringa, w & k, Johnson & C, Gooskens (2009). Measuring Norwegian dialect distances using acoustic features, in *Speech Communication* 51: 167-183.
- Hoppenbrouwers, C. and Hoppenbrouwers, G. (1988). De featurefrequentiemethode en de classificatie van Nederlandse dialecten. *TABU: Bulletin voor taalwetenschap*, 18(2):51-92.
- Hoppenbrouwers, C. and Hoppenbrouwers, G. (2001). De indeling van de Nederlandse streektalen. Dialecten van 156 steden en dorpen geklasseerd volgens de FFM. Koninklijke Van Gorcum B.V., Assen.
- IPA (1999). *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.
- Jain, A. K. & R. C. Dubes (1988). *Algorithms for Clustering Data*. Englewood Cliffs, New Jersey: Prentice Hall.
- Leinonen, Therese (2010). An acoustic analysis of vowel pronunciation in Swedish dialects, University of Groningen Dissertations in Linguistics 83.
- Nerbonne, J. & W. Kretzschmar (2003). Introducing computational techniques in dialectometry. *Computers and the Humanities*, 37:245-255.
- Nerbonne, J. & W. Heeringa (1998). Computationale vergelijking en classificatie van dialecten. *Taal en Tongval, Tijdschrift voor Dialectologie*, 50(2):164-193.
- Nerbonne, J. & W. Heeringa (2001). Computational comparison and classification of dialects. *Dialectologia et Geolinguistica. Journal of the International Society for Dialectology and Geolinguistics*, 9:69-83.
- Osenova, P. & W. Heeringa & J. Nerbonne (2010). A quantitative analysis of Bulgarian dialect pronunciation. In *Zeitschrift für Slavische philologie*, Vol. 66, No. 2, p. 425-458.
- Palander, M. & L. L. Opas-Hanninen & F. Tweedie (2003). Neighbours or enemies? Computing variants causing differences in transitional dialects, *Computers and Hummanities* 37: 359-372.
- Patriarca, M. and E. Heinsalu (2009). Influence of geography on language competition. In *Physica A: statistical mechanics and its applications*. Volume 388 (2): pp 174-186.
- Prokic, J. (2007) Identifying linguistic structure in a quantitative analysis of dialect pronunciation. In proceedings of the ACL student research workshop, Prague, June 2007:61-66.
- Weijnen, A. (1946). The borders between the Brabant dialects cross east-north. In *lectures on North Brabant dialect issues*. Royal Dutch Academy of Sciences, April 12, 1944, Volume VIII.
- Weijnen, A. (1946).
- Wieling, M. (2012). A quantitative approach to social and geographical dialect variation, PhD Dissertation in Linguistics. The University of Groningen, Groningen.

## پیوست‌ها

پیوست 1) فهرست سایت‌های مشخص شده در نقشه راهنما (تصویر 2)

1	پاشا کلا	16	استخرپشت	31	میستان	46	آزادبن	61	یکه توک
2	برارده	17	کچب محله	32	دون چال	47	ترک محله	62	سفیدچاه
3	تیر کلا	18	سفید کوه	33	لاکدشت	48	ماندی محله	63	قره تپه
4	مقام	19	لایی پاسند	34	میان ملک	49	بیزیکی	64	زاغمرز
5	عیسی کنده	21	ارزک	35	تلنار	50	فوتم	65	ارزت
6	خارمیان	22	بندبن	36	آباده	51	سید کلا	66	یعقوب‌لنگه
7	دامیر	22	سورک	37	زرین کلا	52	حاجی کلا	67	حسین آباد
8	جره سر	23	چلمردی	38	خرده لاریم	53	غریب محله	68	زینه‌وند
9	دلمرز	24	نیم چاه	39	چوباغ	54	محمد آباد	69	آسیاب‌سر
10	تجن لته	25	سیاوش کلا	40	عشیره ده	55	شیرداری	70	کوهستان
11	بزمین آباد	26	دغان‌لو	41	شورکا	56	چالکده	71	گرچی محله
12	طبقه	27	لاک تراشان	42	چهارطاق‌بن	57	قلعه پایان	72	کنت
13	لاک دشت	28	شوراب	43	قوشچی محله	58	کرمانی محله	73	اندورات
14	نبی آباد	29	کمیشان	44	سفرخیل	59	یارسم		
15	گالش کلا	30	چرخ کتی	45	تالش محله	60	کیاسر		

پیوست 2) مدخل‌های واژگانی پرسشنامه

پدر	مادر	برادر	خواهر	پسر	دختر	بچه	عمو	دایی
عمه	خاله	زن	شوهر	داماد	عروس	پدر پدر	پدر مادری	خورشید
ماه	ستاره	شب	روز	باد	آتش	خاک	باران	برف
سنگ	چشم	دهان	زبان	خون	گاو	بز	بره	میش
گرگ	قاطر	سگ	مرغ	خروس	کبوتر	چوب	برگ	گل
درخت	جو	بادام	یوغ	خانه	اجاق	هیزم	نان	پنیر
نماز	نر	ماده	شیرین	دور	نزدیک	انسان آبستن	حیوان آبستن	