فصلنامه تازه های علوم شناختی
سال۱۶، شماره ۲، تابستان ۱۳۹۳

Advances in Cognitive Science
Vol. 16, No. 2, 2014

آنالیز پی: توزیع و خنثی شدن سرنخ‌های صوتی-شناختی در واژگان زبان فارسی معیار

افشین رحیمی
دانشجوی کارشناسی ارشد زبان‌شناسی رایانشی، مرکز زبان‌ها و زبان‌شناسی، دانشگاه صنعتی شریف
بهرام وزیرنژاد*
استادیار، گروه زبان‌شناسی رایانشی، آزمایشگاه پردازش گفتار و زبان، مرکز زبان‌ها و زبان‌شناسی، دانشگاه صنعتی شریف
محرم اسلامی
دانشیار، دانشکده زبان شناسی، دانشگاه زنجان

***نشانی تماس**: دانشگاه صنعتی شریف، گروه زبان‌شناسی رایانشی، آزمایشگاه پردازش گفتار و زبان، ساختمان ابن سینا، طبقه چهارم
رایانامه: bahram@sharif.edu

هدف: سرنخ‌های صوتی-شناختی نقش مؤثری در شکل‌گیری ساختار زبان دارند. میزان تمایزدهندگی این سرنخ‌ها در سیگنال صوتی باعث می‌شود تا بتوان از زبان به عنوان یک ابزار ارتباطی به نحو مؤثری استفاده کرد. **روش**: در این مقاله با استفاده از آنالیز پی میزان نقش‌آفرینی ویژگی‌های واجی شامل شیوهٔ تولید، محل تولید و واک‌داری در بافت‌های پیش‌واکه C1، پیش‌هم‌خوانی C2 و جایگاه پایانی C3 در هجای C1VC2C3 بررسی شده است. **نتیجه‌گیری**: نتایج نشان می‌دهد که میزان تمایزدهندگی این ویژگی‌ها در محیط پیش‌واکه بیشتر از محیط پیش‌هم‌خوان بوده و در جایگاه پایانی کمی بیشتر از جایگاه پیش‌هم‌خوانی است. همچنین این مطالعه حاکی از آن است که میزان تمایز ویژگی شیوهٔ تولید از محل تولید و واک‌داری در زبان فارسی بیشتر است، به طوری که در محیط پیش‌هم‌خوان تقریبا هیچ‌گاه ویژگی واک‌داری تمایز ایجاد نمی‌کند. نتایج این مطالعه کاملا بر نظریه‌های جواز ادراکی و توزیع تقابل منطبق است.

کلیدواژه‌ها: *سرنخ‌های صوتی-شناختی، آنالیز پی، جواز ادراکی، توزیع مشخصه‌های تقابل‌دهنده، زبان فارسی*

P-Analysis: Dispersion and Neutralization of Acoustic Cognitive Cues in

Afshin Rahimi
MSc student in Computational Linguistics, Sharif University of Technology
Bahram Vazirnezhad*
Assistant Professor, Division for Computational Linguistics, Speech and Language Processing Laboratory, Language and Linguistics Center, Sharif University of Technology
Moharram Eslami
Associate Professor, Linguistics Department, Zanjan University

Corresponding Author:
Email: bahram@sharif.edu

Introduction: Cognitive acoustic cues have an important role in shaping the phonological structure of language as a means to optimal communication. In this paper, we introduced P-trac procedure in order to track dispersion of contrasts in different contexts in lexicon. The results of applying P-trac procedure to the case of dispersion of contrasts in preconsonantal contexts and in consonantal positions of CVCC sequences in Persian provide Evidence in favor of phonetic basis of the dispersion of contrasts argued by Licensing by Cue hypothesis and the Dispersion Theory of Contrast. The P-trac procedure is proved to be very effective in revealing the dispersion of contrasts in lexicon especially when providing means to compare the dispersion of contrasts in various contexts.

Keywords: *Cognitive Acoustic Cues, P-trac procedure, Licensing by Cue, Dispersion of Contrasting Features, Persian.*

مقدمه

به طور سنتی محدودیت‌های واجی دلیل ایجاد الگوهای واج‌شناختی شناخته شده است. اما نتیجهٔ مطالعات جدید در زمینهٔ نقش عوامل آواشناختی (مانند عوامل تولیدی و ادراکی) در پدیده‌های واج‌شناختی بسیار ثمربخش بوده است. واج‌شناسی مبتنی بر آواشناسی حاصل رهیافت‌های واج‌شناختی مبتنی بر فیزیک صوت است (۱-٤). در این ارتباط در ادامه دو نظریهٔ جواز ادراکی[1] و توزیع تقابل[2] معرفی می‌شود.

نظریهٔ "جواز ادراکی" (٤). مبتنی بر نقش آواشناسی در واج‌شناسی است. بر اساس این نظریه، رخداد یک دنبالهٔ آوایی[3] ارتباط بسیار نزدیکی با میزان جواز ادراکی‌ئی داردکه یک بافت در اختیار آن آوا قرار می‌دهد. تعریف جواز ادراکی عبارت است از وضوح و مقاوم بودن گفتار در شرایط نویزی (٤). هر میزان که ویژگی واجی F متعلق به آوای S در بافت C از نظر شنیداری مقاوم‌تر باشد، احتمال اینکه آوای S بر اساس مقادیر مختلف ویژگی F تمایز واجی ایجاد کند بیشتر و در مقابل هر چه در یک بافت جواز ادراکی یک ویژگی کمتر باشد، احتمال خنثا شدن تمایز واجی براساس آن ویژگی در آن بافت کمتر می‌شود. البته کوچتف (٥) با استفاده از تقابل کامی‌شدگی ساده[4] در هم‌خوان‌های انفجاری تیغه‌ای[5] در مقابل تبیین کاملا آواشناختی شواهدی از پدیده‌های واج‌شناختی ارایه کرده است.

"توزیع تقابل" نظریهٔ دیگری است برای تبیین الگوهای واج‌شناختی که طراح آن فلیمنگ (١،٦،٧) است. این نظریه مبتنی بر اثرگذاری کارکردهای ارتباطی زبان بر شکل‌گیری الگوهای واج‌شناختی است. بر اساس این نظریه، توزیع تقابل‌های واج‌شناختی بر پایهٔ سه هدف ارتباطی شکل گرفته‌اند:

- بیشینه کردن تعداد تقابل‌ها؛
- بیشینه کردن تمایز ادراکی تقابل‌ها؛
- کمینه کردن تلاش در جهت تولید تقابل‌ها.

تقابل بیشتر باعث می‌شود تعداد واژگان در زبان افزایش یابد و ارتباطات با استفاده از واژگان بیشتر پیچیده‌تر و گسترده‌تر شود. هر چه تمایز ادراکی ویژگی‌های ممیز بیشتر باشد، سیگنال گفتار نسبت به نوفه مقاوم‌تر است و بهتر شنیده می‌شود. کمینه بودن تلاش در جهت تولید ویژگی‌های ممیزکمک می‌کند تا با صرف انرژی کمتر اطلاعات بیشتری در ارتباط منتقل شود. تأثیر کارکرد ارتباطی زبان بر شکل‌گیری پدیده‌های واج‌شناختی در پژوهش‌های متعدد مطرح (٨-١٠) و البته ایرادهایی نیز به این نظریه‌ها وارد شده است. برای بررسی بیشتر این ایرادها می‌توان به (١٣-١١) مراجعه کرد.

ایدهٔ مشترک دو نظریهٔ جواز ادراکی و توزیع تقابل، بررسی نقش مؤثر عوامل ادراکی بر الگوها و پدیده‌های واج‌شناختی است. سرنخ‌های ادراکی هم‌خوان‌ها کاندیدهای بسیار مناسبی برای پژوهش در زمینهٔ نقش عوامل آواشناختی در واج‌شناسی، به خصوص نقش سرنخ‌های ادراکی در توزیع ویژگی‌های ممیز واجی، هستند. هم‌خوان‌ها در بافت‌های مختلف میزان تمایزات واجی متفاوتی از خود بروز می‌دهند. میزان تمایز واجی که هم‌خوان‌ها در محیط پیش‌واکه نشان می‌دهند، بسیار بیشتر از میزان این تمایزات در محیط پیش‌هم‌خوانی در هجای CVVV است. تبیین واج‌آرایی زبان بر اساس رسایی، هر چند پیش‌بینی‌های درستی می‌کند اما در همهٔ موارد قادر به توضیح الگوهای واج‌شناختی نیست. در مقابل، واج‌شناسی مبتنی بر آواشناسی می‌کوشد تا با مطالعهٔ تأثیرپذیری سرنخ‌های ادراکی از بافت، نظریهٔ واج‌آرایی زبان را تبیین کند (١٧-١٤، ١). مطالعات انجام شده شواهد جهانی را در مقابل تبیین واج‌شناختی بر اساس رسایی و به سود تبیین مبتنی بر میزان قوی بودن سرنخ‌های ادراکی در یک بافت نشان می‌دهد. در این بخش از مقاله، بر اساس مقالهٔ مهم١٦ سرنخ‌های ادراکی، محل تولید، شیوهٔ تولید و واک‌داری هم‌خوان‌ها معرفی می‌شوند.

سرنخ‌های ادراکی محل تولید با ویژگی‌های آکوستیکی مانند گذر سازه‌ها[6]، رهش هم‌خوان‌ها[7]، الگوهای قطب-صفر خیشومی‌ها[8] و نوفه سایشی‌ها مرتبط‌اند. این سرنخ‌ها در بافت گذر به واکهٔ مجاور قابلیت شنیداری بیشتری دارند. این امر به خصوص در سرنخ‌های ادراکی ویژگی محل تولید هم‌خوان‌های انفجاری بیشتر دیده می‌شود؛ به این معنا که سرنخ‌های ادراکی مرتبط با ویژگی محل تولید، در هم‌خوان‌های انفجاری در محیط‌های گرفته پس از خود ضعیف‌ترند. هم‌خوان‌های

1. Licensing by Cue Hypothesis
2. Dispersion Theory of Contrasts
3. Segment sequence
4. Plain-Platalized
5. Coronal
6. Formant Transition
7. Stop Relsease Burst
8. Pole Zero Patterns

سایشی بر خلاف هم‌خوان‌های انفجاری سرنخ‌هایی دارند که محل تولید آنها را مستقل از بافت پسین قابل بازیابی می‌کند. کناری‌ها نیز دارای سرنخ‌های ادراکی هستند که می‌توانند حتا در کل یک هجا پخش شوند (۱۶). به دلیل حساسیت بیشتر سرنخ‌های ادراکی ویژگی محل تولید درهم‌خوان‌های انفجاری، کلیهٔ زبان‌ها به صورت جهانی محدودیت‌های بیشتری بر واج‌آرایی و محل قرار گرفتن این هم‌خوان‌ها در دنبالهٔ آوایی اعمال می‌کنند.

از نظر شیوهٔ تولید، هم‌خوان‌ها را با استفاده از میزان ارتعاش در سیگنال گفتار، به عنوان یک کلید شناختی می‌توان متمایز کرد. برای مثال، یک تغییر ناگهانی در سیگنال نشانهٔ یک هم‌خوان انفجاری است. ارتعاش کامل در طول آوا همراه با نوفهٔ سایشی نشانگر یک هم‌خوان سایشی است. افت دامنهٔ سیگنال هم‌خوان‌های خیشومی معمولا" کمتر از هم‌خوان‌های سایشی است. هم‌خوان‌های خیشومی همچنین از الگوهای قطب-صفر به عنوان کلید شیوهٔ تولید خیشومی بهره می‌برند. سرنخ‌های ادراکی مرتبط با ویژگی شیوهٔ تولید، قوی‌تر و در مقابل نوفه مقاوم‌تر از سرنخ‌های شناختی محل تولید و واک‌داری هستند (۱۶). سرنخ‌های ادراکی مرتبط با واک‌داری شامل منظم بودن، زمان آغاز واک[1] و دمش[2] است.

مطالعهٔ سرنخ‌های ادراکی نشان داده است که سرنخ‌های مرتبط با واک‌داری ضعیف‌تر از سرنخ‌های ادراکی مرتبط با ویژگی محل تولید و این دو نیز ضعیف‌تر از سرنخ‌های شناختی مرتبط با ویژگی شیوهٔ تولید هستند. این موضوع به خصوص در محیط پیش‌هم‌خوانی صادق است. بنابر نظریهٔ جواز ادراکی در صورتی که سرنخ‌های ادراکی یک ویژگی واجی در یک بافت ضعیف باشند، احتمال اینکه یک آوا در آن بافت بر اساس مقادیر مختلف آن ویژگی متمایز شود، کمتر است و در این حالت ویژگی ممیز اصطلاحاً خنثی[3] می‌شود.

هدف این پژوهش، بررسی نظریهٔ جواز ادراکی (٤) در زمینهٔ خنثا شدن تمایزات[4] هم‌خوان‌ها در بافت‌های پیش‌هم‌خوانی است. بر اساس این نظریه پیش‌بینی می‌شود که دنباله‌های آوایی تمایز بیشتر را با استفاده از ویژگی‌هایی نشان دهند که در یک بافت ادراکی قوی هستند و به عکس تمایز کمتر را بر اساس ویژگی‌هایی نشان دهند که در یک بافت سرنخ‌های ادراکی ضعیفی داشته باشند. به دلیل تمایز بیشتر ویژگی‌های شیوهٔ تولید نسبت به محل تولید و ویژگی واک‌داری، اگر نظریهٔ جواز ادراکی درست باشد، می‌بایست میزان تمایز واجی ایجاد شده با استفاده از ویژگی شیوهٔ تولید بیشتر از محل تولید و میزان تمایز واجی ایجاد شده با استفاده از ویژگی محل تولید نیز بیشتر از ویژگی واک‌داری باشد.

هدف دیگر این مطالعه، معرفی آنالیز پی است که یک تحلیل آماری است و توزیع ویژگی‌های ممیز را از بافت‌های مختلف دنباله‌های آوایی استخراج می‌کند. فرض این است که اگر یک کلید ادراکی در یک بافت ضعیف است، به صورت درزمانی و به تدریج تمایزات واجی ایجاد شده با استفاده از آن ویژگی به دلیل تغییرات واجی کاهش یافته و این کاهش با استفاده از آنالیز پی می‌تواند دریافت شود. آنالیز پی می‌تواند توزیع تمایزات واجی را بر اساس ویژگی‌های ممیز واژگان به صورت وابسته به بافت استخراج و برای مطالعهٔ میزان قوی بودن سرنخ‌های ادراکی در بافت‌های مختلف راه‌کارهای اثربخش ارایه کند.

هدف آنالیز پی استخراج توزیع ویژگی‌های ممیز از واژگان بر اساس بافت است. فرض آنالیز پی این است که اگر عوامل ادراکی بر الگوهای واج‌شناختی مؤثرند، واژگان می‌بایست به صورت درزمانی با استفاده از تغییرات واجی از نظر شنیداری رو به بهینه شدن داشته باشند، به طوری که تعداد زیادی از تمایزات واجی ایجاد شده در واژگان در بافتی باشد که برای آن تمایز سرنخ‌های ادراکی قوی وجود دارد. برای مثال، در هجای CVCC سرنخ‌های ادراکی واک‌داری در بافت پیش‌هم‌خوانی ضعیف‌تر از بافت‌های پیش‌واکه‌ای هستند. بنابراین بسامد تمایزات موجود با مشخصهٔ واک‌داری می‌بایست در بافت‌های پیش‌واکه بیش از بافت‌های پیش‌هم‌خوان باشد. برای یافتن توزیع ویژگی‌های ممیز تعریف جفت‌کمینهٔ مشخصه‌ای[5] و جفت‌کمینهٔ دنبالهٔ واجی[6] مورد نیاز است. یک جفت-کمینه شامل دو دنبالهٔ آوایی است که در همهٔ

1. Voice Onset Time (VOT)
2. Aspiration noise
3. Neutralize
4. Contrast Neutralization
5. Featural Minimal Pair
6. Sequence Minimal Pair

ویژگی‌های ممیز(به جز یک ویژگی) با هم اشتراک دارند. برای مثال، واج‌های /b/ و /p/ یک جفت‌کمینهٔ مشخصه‌ای را تشکیل می‌دهند، زیرا آنها در همهٔ ویژگی‌های واج‌شناختی (به جز واک‌داری) مشترک‌اند. همچنین /b/ و /d/ نیز یک جفت‌کمینهٔ مشخصه‌ای را تشکیل می‌دهند که فقط در ویژگی محل تولید تفاوت دارند. دنبالهٔ آوایی/b/ و /t/ یک جفت‌کمینهٔ مشخصه‌ای نیستند، چرا که علاوه بر ویژگی واک‌داری، در ویژگی محل تولید نیز متفاوت‌اند. یک جفت‌کمینه دنبالهٔ واجی به دو دنبالهٔ آوایی اطلاق می‌شود که فقط در یک آوا با یکدیگر تفاوت دارند و این دو آوای متفاوت نیز تشکیل یک جفت‌کمینهٔ مشخصه‌ای می‌دهند. مثلا، دنباله‌های /band/ و /pand/ یک جفت‌کمینهٔ دنبالهٔ واجی را تشکیل می‌دهند، زیرا فقط در /b/ و /p/ تفاوت دارند و این دو آوا نیز جفت‌کمینهٔ مشخصه‌ای هستند. به همین صورت /band/ و /dand/ نیز یک جفت‌کمینهٔ دنبالهٔ واجی‌اند اما /band/ و /tand/ نیستند، چون /b/ و /t/ یک جفت‌کمینهٔ مشخصه‌ای را تشکیل نمی‌دهند. تعریف جفت‌کمینهٔ مشخصه‌ای با تعریف جفت‌کمینه‌ای ،که معمولا" در مطالعات واج‌شناختی مطرح می‌شود، متفاوت است. جفت‌کمینه در واج‌شناختی به صورت دو دنباله از دنباله‌های آوایی تعریف می‌شود که فقط در یک آوا/واج با هم تفاوت دارند. از جفت‌های کمینه معمولا برای استخراج فهرست واج‌های یک زبان ناشناخته استفاده می‌شود، در حالی که ویژگی جفت‌کمینه در شناخت ویژگی‌های ممیز کاربرد دارد. جدول ۱ تمامی جفت‌های کمینهٔ مشخصه‌ای ممکن، ویژگی ممیز مربوطه و سرنخ‌های صوتی آن را در زبان فارسی معیار نشان می‌دهد.

آنالیز پی با استخراج جفت‌های کمینهٔ مشخصه‌ای با توجه به موضوع و اهداف مطالعه از واژگان آغاز می‌شود. مثلا، برای استخراج توزیع ویژگی‌ممیز شیوهٔ تولید، محل تولید و واک‌داری در بافت‌های مختلف در هجاهای CVCC واژگان زبان فارسی می‌بایست تمامی

جدول ۱ – جفت‌های کمینهٔ مشخصه‌ای زبان فارسی

ویژگی ممیز	تعداد جفت‌های کمینهٔ مشخصه‌ای	جفت‌های کمینهٔ مشخصه‌ای	سرنخ‌های صوتی
شیوهٔ تولید	۷۰	(b, m) (b, w) (b, v) (č, t) (č , s) (č , š) (d, n) (d, r) (d, z) (d, l) (d, ǰ) (d, ž) (f, p) (g, y) (ǰ, n) (ǰ, r) (ǰ, z) (ǰ, l) (ǰ, d) (ǰ, ž) (k, x) (l, z) (l, r) (l, d) (l, n) (l, ž) (l, ǰ) (m, w) (m, b) (m, v) (n, z) (n, r) (n, l) (n, d) (n, ǰ) (n, ž) (p, f) (G, y) (r, d) (r, z) (r, ǰ) (r, n) (r, ž) (r, l) (s, t) (s, č) (š, t) (š, č) (t, s) (t, š) (t, č) (v, b) (v, m) (v, w) (w, b) (w, v) (w, m) (x, k) (y, g) (y, G) (z, r) (z, n) (z, l) (z, d) (z, ǰ) (ž, r) (ž, n) (ž, l) (ž, d) (ž, ǰ)	درجهٔ تغییر سیگنال[1]
محل تولید	۵۰	(ʔ, b) (ʔ, G) (ʔ, d) (ʔ, g) (b, ʔ) (b, G) (b, d) (b, g) (d, G) (d, ʔ) (d, b) (d, g) (f, s) (f, š) (f, x) (f, h) (g, ʔ) (g, d) (g, b) (h, s) (h, x) (h, f) (h, š) (k, t) (k, p) (m, n) (n, m) (p, t) (p, k) (G, d) (G, b) (G, ʔ) (s, f) (s, x) (s, h) (š, x) (š, f) (š, h) (t, k) (t, p) (v, z) (v, ž) (w, y) (x, s) (x, f) (x, š) (x, h) (y, w) (z, v) (ž, v)	گذر فرمانتی رهش انفجاری قطب-صفر خیشومی نوفهٔ سایشی
واک-داری	۲۰	(b, p) (č , ǰ) (d, t) (f, v) (g, k) (ǰ, č) (k, G) (k, g) (p, b) (G, k) (s, z) (s, ž) (š, ž) (š, z) (t, d) (v, f) (z, š) (z, s) (ž, š) (ž, s)	منظم بودن زمان آغاز واک دیرش دمش[2]

1. duration, 2. degree of signal attenuation

⬚	$Context_1$	$Context_2$	$Context_i$
Place	$freq$	$freq$	...
Manner	$freq$	$freq$	...
Voice	$freq$	$freq$	...

شکل ۱- نمایی از ماتریس ویژگی-بافت منتج از آنالیز پی

جفت‌های کمینهٔ دنباله به صورت CVCC از واژگان استخراج شود. به عنوان مثالی دیگر اگر هدف مطالعه بررسی توزیع ویژگی‌های ممیز در محیط پیش‌هم‌خوانی در دنباله‌های آوایی در هجای CVCC زبان فارسی باشد، می‌بایست تمامی جفت‌های کمینهٔ دنبالهٔ مشخصه‌ها به صورت C_2C_3 از دنباله‌های $C_1VC_2C_3$ استخراج شوند. در زیر الگوریتم اصلی آنالیز پی ارایه شده است.

الف) بر اساس اهداف مطالعه تمامی جفت‌دنبالهٔ کمینه ویژگی‌ها را استخراج کن.

ب) برای هر جفت‌دنبالهٔ کمینه، ویژگی بافت و ممیز را مشخص و نشانه‌گذاری کن.

ج) بسامد رخداد همهٔ دوتایی‌های به شکل context, contrasting feature را محاسبه کن.

در پایان آنالیز پی یک ماتریس ویژگی-بافت ایجاد می‌شود که مقدار هر درایهٔ آن برابر با میزان رخداد یک تمایز در یک بافت است. شکل ۱ این ماتریس را نشان می‌دهد.

از آنالیز پی می‌توان هم‌زمان در یافتن توزیع یک یا چند ویژگی ممیز در یک یا چند بافت استفاده کرد؛ مثلا برای یافتن توزیع تمامی ویژگی‌های ممیز در بافت پیش از هم‌خوان انفجاری واک‌دار لبی /b/ در تمامی خوشه‌های هم‌خوانی به شکل C_2b (چند ویژگی ممیز در یک بافت) و یا برای استخراج توزیع ویژگی ممیز واک‌داری (یک ویژگی ممیز) در تمامی بافت‌های ممکن (یک ویژگی ممیز در چندین بافت). الگوریتم‌های مشابه آنالیز پی در پژوهش‌های دیگری معرفی شده‌اند تا ویژگی‌های ممیز را بیابند اما این الگوریتم‌ها برای اهدافی غیر از یافتن میزان بهینه بودن واژگان (از نظر توزیع ویژگی‌های ممیز در بافت‌های مختلف) به کار رفته‌اند (۱۸).

روش

همانند واژگان CELEX که پایگاه داده‌ای مختص چند زبان ژرمنی مانند انگلیسی، فرانسه، هلندی و ... است (۱۹) ، واژگان FLexicon (۲۰) برای زبان فارسی

```
for each consonant in C2 position of C2C3 consonant cluster
    for each C21 in C2 position
        for each C22 in C2 position
            If C21 and C22 are featural minimal pair
                feature <- Contrastive Feature of C21 and C22
                freq ← min(frequency of C21C3 , frequency of C22C3)
            frequencies(C3, feature) += freq
```

شکل ۲- الگوریتم آنالیزپی

تدوین شده است. این دادگان شامل ٥٤ هزار و ٤٠٩ واژهٔ زبان فارسی و نگارش واجی آنهاست. به دلیل سادگی ساختار هجایی زبان فارسی و ممنوع بودن آغازه‌های ترکیبی در این زبان هجابندی واژگان به صورت قطعی و به‌سادگی انجام می‌پذیرد. هجاهای زبان فارسی به اشکال CV, CVC, CVCC است.

استفاده از آنالیز پی برای استخراج توزیع ویژگی‌های ممیز در بافت پیش‌هم‌خوانی در هجاهای CVCC واژگان زبان فارسی

در این آزمایش توزیع ویژگی‌های ممیز در بافت پیش‌هم‌خوانی با استفاده از آنالیز پی مطالعه می‌شود. هدف این آزمایش، بررسی میزان هم‌خوانی توزیع ویژگی‌های ممیز در محیط پیش‌هم‌خوانی با میزان جواز ادراکی بافت برای آن ویژگی‌هاست. بر اساس بررسی‌های رایت (١٦)، سرنخ‌های ادراکی در بافت پیش‌هم‌خوانی از ویژگی شیوهٔ تولید به محل تولید و از ویژگی محل تولید به ویژگی واک‌داری کاهش می‌یابد. بنابراین بنا بر نظریهٔ جواز ادراکی و بر اساس تغییرات درزمانی می‌بایست بین میزان قدرت سرنخ‌های ادراکی و بسامد تمایزات ایجاد شده در ماتریس خروجی آنالیز پی یک ارتباط مستقیم مشاهده شود (شکل ۲).

بر اساس روال آنالیز پی همهٔ ٥٤ هزار و ٤٠٩ واژهٔ موجود در واژگان FLexicon بر اساس قواعد هجابندی زبان فارسی هجابندی شدند. هر هجا در زبان فارسی با یک هم‌خوان آغاز می‌شود و تا یک هم‌خوان قبل از واکهٔ بعدی ادامه می‌یابد. پس از هجابندی ٢٦٨ خوشه، هم‌خوانی C_2C_3 از هجاهای $C_1VC_2C_3$ به دست آمد و بسامد (بسامد نوع[1]) هر خوشهٔ هم‌خوانی

1. Type Frequency

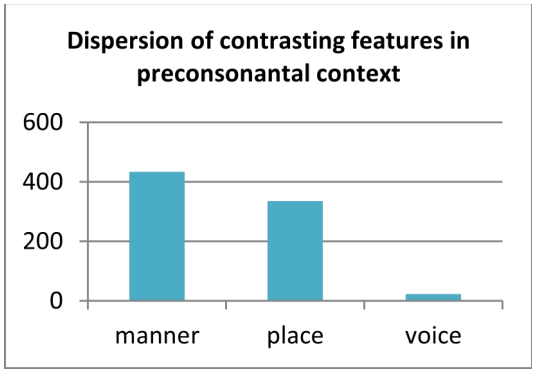
با استفاده از شمارش محاسبه شد. برای هر هم‌خوان در موقعیت C_3 در خوشه C_2C_3 تمامی جفت‌کمینه مشخصه‌ها در موقعیت C_3 استخراج شد. ویژگی‌های استفاده شده برای استخراج جفت‌کمینه‌ها، شیوهٔ تولید، محل تولید و واک‌داری بودند.

توزیع ویژگی‌های ممیز با شمارش تعداد دفعاتی که هر ویژگی در جایگاه C_2 و در بافت پیش‌هم‌خوانی $(_C_3)$ در خوشه‌های هم‌خوانی C_2C_3 تمایز ایجاد کرده بود استخراج شد. به عنوان مثال، برای همهٔ خوشه‌های هم‌خوانی به شکل C_2r تمامی جفت دنباله کمینه ویژگی‌های موجود استخراج و ویژگی‌های ممیز آنها شمرده شد. برای هر ویژگی ممیز بسامد تمایز در بافت پیش‌هم‌خوانی C_3_ (در اینجا /r_/) محاسبه شد. الگوریتم آنالیز پی به صورت سودوکو در شکل ۳ آمده است.

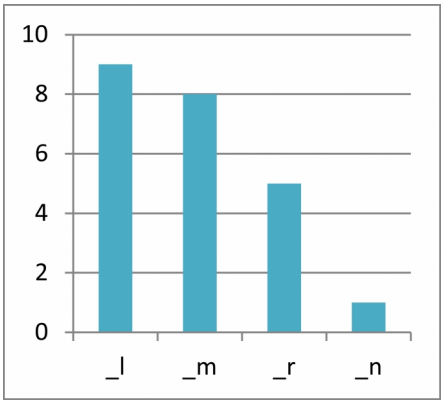
در پایان آنالیز پی برای همهٔ بافت‌ها و همهٔ ویژگی‌ها، تعداد دفعاتی که یک ویژگی در آن بافت تمایز ایجاد کرده محاسبه و در یک ماتریس ذخیره می‌شود.

استفاده از آنالیز پی برای مقایسهٔ ویژگی‌های ممیز در جایگاه‌های هم‌خوانی هجای CVCC در واژگان زبان فارسی

در اینجا از آنالیز پی برای یافتن توزیع ویژگی‌های شیوهٔ تولید محل تولید و واک‌داری در جایگاه‌های هم‌خوانی C_1, C_2 و C_3 در هجای CVCC استفاده شده است. هدف از آنالیز پی در این آزمایش سنجش ارتباط توزیع ویژگی‌ها در بافت‌های مختلف با میزان جواز ادراکی آن بافت‌ها برای ویژگی‌ها بر اساس دو نظریهٔ جواز ادراکی و توزیع ویژگی‌های ممیز است. در اینجا برای هجابندی و یافتن جفت‌های کمینه، دوباره نگارش واجی واژگان در آنالیز پی به کار رفت. همهٔ هجاهای به شکل CVCC به همراه بسامد رخداد آنها در واژگان (بسامد نوع) از واژگان استخراج شد. با توجه به الگوریتم آنالیز پی همهٔ جفت‌دنباله‌های کمینه ویژگی یافته شد. برای هر بافت هم‌خوانی در جایگاه‌های C_1, C_2 و C_3 در دنباله‌های $C_1VC_2C_3$ مشخصهٔ ممیز استخراج و بسامد آن شمرده شد. همانند آزمایش قبل برای هر جفت دنباله کمینه مشخصه‌ها *frequencies[context, feature]* به تعداد کمینه، بسامد نوع اعضای یک جفت کمینه اضافه شد. مثلا" برای /band/ و /pand/ با بسامد نوع ۲۰۰ و ۳۰۰،



شکل ۳- توزیع ویژگی‌های ممیز در بافت پیش‌هم‌خوانی در هجای *CVCC* واژگان زبان فارسی



شکل ٤- بسامد بافت‌های هم‌خوانی (/C_/) که تمایز واک-داری پیش از آنها رخ داده است.

بسامد *frequencies[_and, voice]* به ۲۰۰ افزوده شد؛ یعنی این دو دنباله در این بافت با استفاده از ویژگی واک‌داری ۲۰۰ بار تمایز ایجاد کرده‌اند.

یافته‌ها

نمودار ۱ نتایج آنالیز پی را روی خوشه‌های هم‌خوانی C_2C_3 از هجای *C1VC2C3* استخراج شده از واژگان زبان فارسی نشان می‌دهد (بخش ۲,۲). هم‌خوان‌های موجود در جایگاه C_3 بافت پیش‌هم‌خوانی را برای هم‌خوان‌های جایگاه C_2 فراهم می‌کنند. تمایزات ایجاد شده در موقعیت C_2 باعث تمایز معنایی می‌شوند. ۷۹۱ جفت دنباله کمینه ویژگی یافت شد که از این تعداد ٤۳۳ جفت در ویژگی شیوهٔ تولید، ۳۵۵ جفت در ویژگی محل تولید و ۲۳ جفت در ویژگی واک‌داری تمایز داشتند. بر اساس الگوریتم آنالیز پی، بسامد هر تمایز برابر با کمینه بسامد نوع[1] هر عضو از یک جفت دنباله کمینه ویژگی در واژگان واژگان است. برای مثال،

1. Minimal type frequency

جدول ۲ – جفت دنباله کمینه‌های موجود در زبان فارسی بر اساس تمایز ویژگی واک‌داری در بافت پیش‌هم‌خوانی هجای CVCC واژگان زبان فارسی

جفت دنباله‌های کمینه	تعداد تمایز	بافت پیش‌هم‌خوانی
(/?asl/, /?azl/)	9	**_l**
(/fasl/, /fazl/)		
(/rasm/,/razm/)	8	**_m**
(/?osr/, /?ozr/)	5	
(/?oSr/,(/?ozr)		
(/nasr/, /nazr/)		**_r**
(/naSr/, /nazr/)		
(/satr/, /sadr/)		
(/hosn/, /hozn/)	1	**_n**

در صورتی که دنبالهٔ /bl/ دارای فرکانس نوع ۲۰۰ (یعنی ۲۰۰ هجای CVbl متمایز در واژگان موجود است) و دنبالهٔ /pl/ دارای فرکانس نوع ۳۰۰ باشد، ویژگی واک‌داری در بافت پیش‌هم‌خوانی /_l/ ۲۰۰ تمایز (مقدار کمینهٔ دو بسامد) ایجاد کرده است، بنابر این frequencies[l, voice] ۲۰۰ واحد افزایش می‌یابد.

با اینکه نتایج در شکل۳به صورت تجمعی نمایش داده شده، آنالیز پی توزیع مشخصات را به ازای هر بافت هم‌خوانی و هر ویژگی به دست می‌دهد. برای پی بردن به این موضوع که تمایزات واک‌داری در جایگاه C_2 در کدام بافت‌های پیش‌هم‌خوانی به کار رفته‌اند از خروجی آنالیز پی استفاده شد. دلیل این توجه ویژه، تعداد اندک تمایزات واک‌داری در بافت پیش‌هم‌خوانی است. شکل ٤ همهٔ بافت‌های پیش‌هم‌خوانی (که تمایز واک‌داری جهت ایجاد معانی جدید در واژگان زبان فارسی به کار رفته است) را نشان می‌دهد. همان طور که شکل ٤ نشان می‌دهد، فقط بافت‌هایی که در آنها تمایز واک‌داری به کار رفته پیش از هم‌خوان‌های روان و خیشومی است. در سایر بافت‌ها ویژگی واک‌داری دچار خنثا شدگی شده و مقادیر مختلف آن تمایز معنا ایجاد نمی‌کنند. مثلا، در زبان فارسی جفت‌کمینهٔ */sb/* و */zb/* در هیچ یک از هجاهای */CV_b/* وجود ندارد.

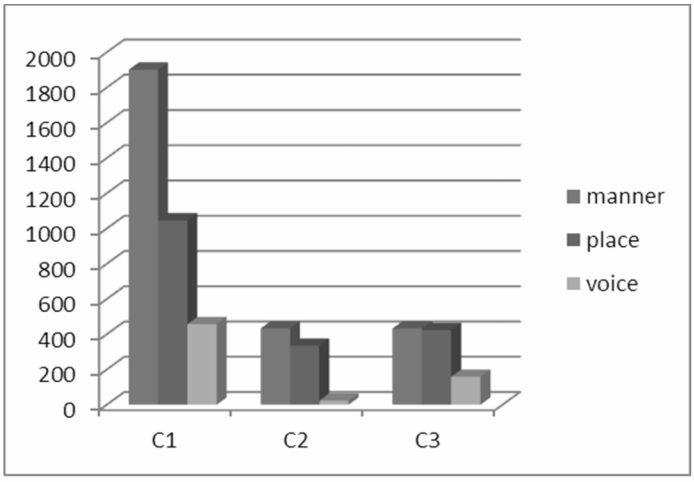
جدول ۲ جفت دنباله کمینه ویژگی‌هایی را که در ویژگی واک‌داری با یکدیگر تفاوت دارند نشان می‌دهد. نکتهٔ بسیار جالب این است که این معدود تمایزات را واژگان قرضی عربی ایجاد کرده‌اند. شایان ذکر است که این کلمات در زبان عربی غالباً با درج واکهٔ بین خوشهٔ دو هم‌خوانی تولید می‌شوند.

در شکل ۵ توزیع ویژگی‌های شیوهٔ تولید، محل تولید و واک‌داری در جایگاه‌های مختلف هم‌خوانی نشان داده شده است (بخش ۲,۳). همان طور که مشاهده می‌شود، تمایزات ایجاد شده با استفاده از ویژگی واک‌داری در بافت پیش‌هم‌خوانی C_2 کمینه و در بافت پیش‌واکه‌ای /_VCC/ بیشینه است.

همان طور که در شکل ۵ نیز مشاهده می‌شود، بسامد کلی تمایزات در جایگاه C_1 و در بافت پیش‌واکه‌ای بسیار بیشتر از سایر بافت‌هاست (٣٤٠٤ تمایز). همچنین تعداد کل تمایزات در جایگاه C_2 در بافت پیش‌هم‌خوانی کمینه (۷۹۱ تمایز) و در جایگاه C_3 در بافت پس‌هم‌خوانی پایانه متوسط (۱۰۱۵ تمایز) است. تعداد تمایزات از ویژگی شیوهٔ تولید به محل تولید و از ویژگی محل تولید به واک‌داری کاهش می‌یابد.

بحث و نتیجه‌گیری

نتایج آنالیز پی روی توزیع تمایزات ویژگی‌ها در بافت پیش‌هم‌خوانی، خوشه‌های هم‌خوانی هجای CVCC را



شکل ٥– توزیع ویژگی‌های ممیز در بافت‌ها و جایگاه‌های مختلف هم‌خوانی در هجای CVCC در واژگان زبان فارسی

نشان می‌دهد که توزیع تمایز ویژگی‌ها کاملا بر پیش‌بینی‌های نظریهٔ جواز ادراکی و نظریهٔ توزیع تمایز منطبق است. بسامد تمایزات ایجاد شده در بافت پیش‌هم‌خوانی برای ویژگی‌های شیوهٔ تولید، محل تولید و واک‌داری به ترتیب کاهش یافته است (شکل ۳). بر اساس نتایج تحقیقات رایت (۱۶)، سرنخ‌های ادراکی ویژگی شیوهٔ تولید قوی‌تر از ویژگی محل تولید و سرنخ‌های ادراکی ویژگی محل تولید قوی‌تر از ویژگی واک‌داری است. قوی بودن سرنخ‌های ادراکی در بافت پیش‌هم‌خوانی کاملاً با بسامد تمایزات ایجاد شده همسو است. هر چه یک ویژگی واجی سرنخ‌های ادراکی قوی‌تری در یک بافت داشته باشد از آن ویژگی در آن بافت جهت ایجاد تمایز معنایی در واژگان بیشتر استفاده شده است. این شواهد آماری که نتیجهٔ آنالیز پی است از تبیین واج‌شناختی بر اساس آواشناسی پشتیبانی می‌کند و بر نظریه‌های جواز ادراکی و توزیع تمایز نیز کاملا منطبق و تأییدی بر آنهاست.

ویژگی واک‌داری نسبت به شیوهٔ و محل تولید ضعیف‌ترین سرنخ‌های ادراکی را در بافت پیش‌هم‌خوانی دارد. نتایج نشان می‌دهد که از ۷۹۳ تمایز ایجاد شده در بافت پیش‌هم‌خوانی فقط اساس ۲۳ تمایز ویژگی واک‌داری است. این ۲۳ تمایز در بافت پیش از هم‌خوان‌های رسا (روان‌ها و خیشومی‌ها) ایجاد شده‌اند (شکل ٤). این نتایج تأییدی بر قوی‌تر بودن سرنخ‌های ادراکی ویژگی واک‌داری در بافت پیش از هم‌خوان‌های رسا از هم‌خوان‌های نارساست. در زبان فارسی ویژگی واک‌داری در بافت پیش‌هم‌خوانی گرفته (مانند هم‌خوان‌های انفجاری، سایشی و انسایشی) دچار خنثاشدگی شده و تمایز معنایی ایجاد نمی‌کند.

شاهد دیگری که بر اساس خروجی آنالیز پی مشخص شده این است که تمامی جفت‌کمینه‌هایی که با استفاده از تمایز در ویژگی واک‌داری در بافت پیش‌هم‌خوانی وجود دارند، بخشی از واژگان قرض گرفته شده از زبان‌های دیگر و به خصوص زبان عربی هستند (جدول ۲). بسیار جالب است که در بین واژگان غیرقرضی حتا یک تمایز واک‌داری مشاهده نشده است. نتایج به دست آمده از آنالیز پی بر هجاهای $C_1VC_2C_3$ حاکی از آن است که تمامی ویژگی‌ها در بافت پیش‌واکه‌ای نسبت به بافت انتهای هجا (جایگاه C_3) و این بافت نسبت به بافت پیش‌هم‌خوانی (جایگاه C_2) تمایز بیشتری را نشان می‌دهند. توزیع تمایزات در این سه بافت کاملا با قدرت سرنخ‌های ادراکی در این بافت‌ها مطابقت دارد. قدرت سرنخ‌های ادراکی در بافت پیش‌واکه‌ای بیشترین، در بافت آخر پایانه متوسط و در بافت پیش‌هم‌خوانی کمترین است (۱۶). الگوی توزیع تمایزات بر الگوی قدرت سرنخ‌های ادراکی کاملا" منطبق است. این شواهد تأییدی بر نقش عوامل آوایی مانند قدرت سرنخ‌های ادراکی بر الگوها و پدیده‌های واج‌شناختی است.

با توجه به نقش مؤثر سرنخ‌های شنیداری-درکی در شکل‌دهی به ساختار زبان، در این مقاله آنالیز پی در جهت یافتن الگوی توزیع ویژگی‌های ممیز در بافت‌های مختلف واژگان زبان معرفی شد. آنالیز پی در دو مورد توزیع مشخصات در بافت پیش‌هم‌خوانی و نیز توزیع مشخصات در تمامی جایگاه‌های هم‌خوانی هجای CVCC در واژگان زبان فارسی انجام شد. نتایج به دست آمده از آنالیز پی بر نظریه‌های تبیین واج‌شناختی بر اساس آواشناسی (مانند نظریهٔ جواز ادراکی و توزیع تمایزات) منطبق است و شواهدی در تحکیم این نظریه‌ها ارایه می‌کند. همچنین آنالیز پی نشان داد که می‌تواند در استخراج الگوی توزیع ویژگی‌های ممیز به خصوص در بافت‌های مختلف بسیار روشن‌گر باشد.

دریافت: ۹۱/۱۲/۷ ; پذیرش: ۹۲/۱۲/۱۵

منابع

1. Flemming E. *Auditory features in phonology [dissertation]*. Los Angeles: UCLA; 1995.
2. Jun J. *Perceptual and articulatory factors in place assimilation: An optimality theoretic approach [dissertation]*. Los Angeles: UCLA 1995.
3. Silverman D. *Phasing and recoverability [dissertation]*. London: Routledge; 1997.
4. Steriade D. *Phonetics in phonology: the case of laryngeal neutralization*. Unpublished manuscript. Ms, UCLA;1997.
5. Kochetov A. Testing Licensing by Cue: A case of Russian palatalized coronals. *Phonetica* 2006;63(2-3):113-48.
6. Flemming E. Contrast and perceptual distinctiveness. *Phonetically Based Phonology* 2004; 232-76.
7. Flemming E. *The role of distinctiveness*

constraints in phonology. Unpublished manuscript, MIT; 2006.
8. Zipf GK. Human behavior and the principle of least effort. *An Introduction to Human Ecology.*Cambridge, Massachuset: Addison Wesley; 1949.
9. Martinet A. Function, structure and sound change. *Journal of the International Linguistic Association* 1952;8(1):1-32.
10. Lindblom B. Economy of speech gestures. *The production of speech* 1983;217-245.
11. Labov W. *Principles of language change: Internal factors*. Oxford: Blackwell; 1994.
12. Ohala JJ. The origin of sound patterns in vocal tract constraints. *The production of speech* 1983;189-216.
13. Trask RL, Trask R. *Historical linguistics*. Arnold London: New York; 1996.
14. Kawasaki H, & Ohala JJ. Acoustic basis for universal constraints on sound sequences. *The Journal of the Acoustical Society of America* 1980;68(S1):S33-S33.
15. Kirchner R. Contrastiveness and faithfulness. *Phonology* 1997;14(1):83-111.
16. Wright R. A review of perceptual cues and cue robustness. *Phonetically based phonology* 2004;34-57.
17. Ohala JJ. The segment: primitive or derived. *Papers in Laboratory Phonology II: Gesture, Segment Prosody* 1992;166-183.
18. Archangeli D. Aspects of underspecification theory. *Phonology* 1988;5(02):183-207.
19. Baayen RH, Piepenbrock R, van HR. The {CELEX} lexical data base on {CD-ROM}.1993.
20. Eslami M, Atashgah MSH, Lamjiri A, Zandi T. Persian generative lexicon. Processing of 1st workshop on Persian and computer; 2004.[Persian].