

بررسی آماری الگوهای توزیع رسایی در مرز هجای زبان فارسی

محرم اسلامی
meslami@znu.ac.ir
دانشکده زبان شناسی،
دانشگاه زنجان

بهرام وزیرنژاد
bahram@sharif.edu
آزمایشگاه پردازش گفتار و زبان،
مرکز زبان ها و زبان شناسی، دانشگاه
صنعتی شریف

افشین رحیمی
a_rahimi@ce.sharif.edu
آزمایشگاه پردازش گفتار و زبان، مرکز
زبان ها و زبان شناسی، دانشگاه صنعتی
شریف

چکیده

در این مقاله محدودیت رسایی در مرز هجا در واژگان زبان فارسی مورد بررسی قرار گرفته است. محدودیت رسایی در مرز هجا بیانگر تمایل جهانی در جهت تغییر افتان رسایی همخوانها در مرز هجا می باشد. در این بررسی ساخت های چند هجایی که در مرز هجا دارای خوشه همخوانی هستند از واژگان زبانی فارسی استخراج شده و مطالعه شده اند. همخوانهای زبان فارسی به ۵ طبقه رسایی مختلف گروه بندی شده و ارتباط میزان تفاوت رسایی در مرز هجا و بسامد در واژگان مورد بررسی قرار گرفته است. جهت محاسبه بسامد نرمال از معیاراطلاعات مشترک نقطه ای^۱ استفاده شده است. از آنجا که خوش ساختی یک ساختار باعث افزایش بسامد آن به صورت همزمانی و در زمانی می گردد، می توان میزان خوش ساختی یک ساختار را از بسامد رخداد آن دریافت. نتایج به دست آمده نشانگر این است که نه تنها افتان بودن رسایی خوشه همخوانی در زبان فارسی به صورت مقوله ای عمل نکرده و ساختارها را به دو طبقه خوش ساخت و بدساخت تقسیم نمی کند بلکه اینگونه نیست که خوشه های با الگوی تغییر رسایی افتان دارای بسامد بیشتری باشند. همچنین استفاده از معیار اطلاعات مشترک نقطه ای نشان می دهد که تمایل به همخدادی در خوشه های با الگوی تغییر رسایی افتان

^۱ PMI Pointwise Mutual Information : $PMI(x, y) = \log_2 \frac{p(x, y)}{p(x)p(y)}$

بیشتر از الگوی تغییر رسایی خیزان نمی باشد. این نتایج نشان می دهد که محدودیت رسایی در مرز هجا در زبان فارسی اعمال نشده و بنابراین در بسامد واژه ها بروز نکرده است. با توجه به اینکه در زبان فارسی به دلیل عدم وجود خوشه ی همخوانی در پیش هسته تشخیص مرز هجا همواره به صورت قطعی انجام می گردد، می توان نتیجه گرفت که نقش الگوی افتان تغییر رسایی در خوشه های همخوانی مرز هجا ایجاد کلید شناختی در جهت تشخیص مرز هجاست و در زبان فارسی به علت قطعی بودن مرز هجا فارغ از نوع خوشه همخوانی این کلید شناختی استفاده نمی گردد.

۱. مقدمه

رسایی یکی از مشخصه های شناختی^۱ واج های زبان است که از نظر تولیدی نشانگر باز شدن بیشتر دهان، از نظر شنیداری نشانگر شدت بیشتر صوت و از نظر آکوستیک نشانگر دامنه بیشتر سیگنال صوتی است. رسایی مقوله ای نسبی است به این معنی که میزان رسایی یک واج نسبت به واج دیگری سنجیده می شود. بر اساس این نسبت واج های هر زبان را به صورت سلسله مراتبی دسته بندی می کنند. نسبت رسایی واج ها در زبان های مختلف به میزان کمی متغیر است به صورتی که می توان سلسله مراتب واج ها را از نظر میزان رسایی در مورد همه زبان ها با تغییراتی اندک جاری دانست. سلسله مراتب رسایی واج ها در برکوست^۲ (۲۰۱۰: ۱۴۷) آمده است. رسایی واکه های باز بیشترین و همخوان های انفجاری ترکیبی دارای کمترین رسایی می باشند.

در منابع، رسایی و نقش های آن به صورت های مختلفی تعریف شده است. مهم ترین تفاوت در این تعاریف قایل شدن نقش مشخصه ممیزه^۳ واجی برای رسایی در برخی منابع و عدم قایل شدن این نقش در برخی از منابع دیگر است. در صورتی که رسایی یک مشخصه ممیزه باشد که باعث تمایز واجی گردد می بایست مانند سایر مشخصه های ممیز مانند واکداری، فرمنت ها، نحوه تولید و ... دارای معادل تولیدی-آوایی باشد. در برخی از منابع مانند کلمنتز^۴ (۱۹۹۰) رسایی را مجموعه های از مشخصات آکوستیکی معرفی کرده اند. در برخی دیگر از منابع

^۱ Cognitive Cue

^۲ D. A. Burquest

^۳ Distinctive Feature

^۴ G. N. Clements

مانند هریس^۱ (۲۰۰۵) رسایی به عنوان سیگنال حامل^۲ و رساننده پیام زبانی شناخته شده و از خود پیام^۳ که دارای مشخصات واجی و ممیز است مجزا می گردد.

۲. اصل رسایی در مرز هجاها^۴

اصل رسایی و اصل بیشینگی پیش‌هسته به بررسی میزان رسایی دنباله‌ی واج‌ها در درون ساخت هجایی می‌پردازند. بررسی میزان رسایی واج‌ها در نقاط اتصال بین هجاها نشان می‌دهد که در زبان‌های مختلف تمایل به گذر از یک پس‌هسته رساتر از هجای اول به یک پیش‌هسته رساتر در هجای دوم بیشتر است. به عبارت دیگر میزان رسایی همخوان‌ها در مرز هجاها افتان است. در ونمان^۵ (۱۹۸۸) تمایل بسیار زیادی در زبان‌های مختلف در جهت نزدیکی به این قانون گزارش شده است. در مواردی که فاصله رسایی خیزان وجود دارد تمایل زبان‌ها به واکه‌افزایی گزارش شده است. تبدیل 'that way' به 'that-a-way' در زبان انگلیسی یک نمونه از واکه‌افزایی در جهت انطباق بیشتر با این قانون است. در نظریه بهینگی این قانون به صورت یک یا چند محدودیت تعریف می‌شود به این صورت که تفاوت رسایی پس‌هسته و پیش‌هسته بعد از آن از تفاوت t.w به عنوان بیشترین فاصله رسایی +7 خیزان و w.t به عنوان بیشترین فاصله افتان با فاصله 7- متغیر است. هر فاصله رسایی به عنوان یک محدودیت تعریف می‌شود (گوشکووا^۶، ۲۰۰۴).

در شکل ۱ میزان اختلاف مجاز رسایی در چند زبان مشاهده می‌گردد (گوشکووا، ۲۰۰۴). به عنوان مثال در زبان قزاق الگوی رسایی خیزان مجاز نیست اما الگوی رسایی flat و یا همه حالت‌های افتان مجاز هستند اما در زبان قرقیز می‌بایست حتما شیب افتان بودن رسایی از ۳- بیشتر باشد در غیر این صورت ساختار بدساخت خواهد بود.

¹ J. Harris

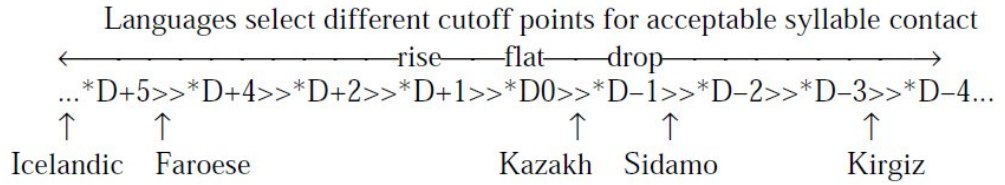
² Carrier

³ Message

⁴ Syllable Contact Law

⁵ T. Vennemann

⁶ M. Gouskova



شکل ۱ - میزان تفاوت رسایی مجاز در زبان های مختلف ارایه شده در (Gouskova, 2004). در این زبان ها میزان تفاوت رسایی در مرز هجا به صورت مقوله ای عمل کرده و ساختار را به دو مقوله خوش ساخت و بد ساخت تقسیم می کند.

۳. دادگان و روش بررسی

با استفاده از پیکره واژگان زیای زبان فارسی که دارای بیش از ۵۴۰۰۰ واژه زبان فارسی می باشد، ساختار هجایی با استفاده از آوانگاری موجود استخراج شد. با توجه به اینکه ساختار هجایی زبان فارسی فاقد خوشه همخوانی در پیش هسته می باشد و وجود یک همخوان پیش هسته الزامی می باشد هجابندی به صورت قطعی و به شکلی ساده انجام می گیرد. برای محاسبه رسایی ساختار سلسله مراتبی رسایی واج ها به صورت زیر در نظر گرفته شده اند.

Low-Vowels>High-Vowels >Liquids >Nasals >Fricatives >Affricates >Plosives

واج های زبان فارسی هر کدام در یکی از هفت گروه ذکر شده قرار گرفته اند.

[a e o] > [A i u] > [r l y] > [m n] > [z s f v s S h x Zh] > [tS dj] > [b d g q ? p t k]

از آنجا که واکه ها در گروه همخوانی حضور ندارند، گروه های همخوانی ذکر شده را به ترتیب رسایی امتیازبندی می کنیم. در جدول ۱ امتیاز رسایی گروه ها مشاهده می شود. با توجه به اینکه ۵ سطح رسایی در بین همخوان ها تعریف کرده ایم فاصله رسایی بین دو همخوان می تواند از +۴ تا -۴ متغیر باشد. فاصله +۴ به معنی فاصله رسایی خیزان بین دو همخوان در مرز هجاست که حالتی نشاندار محسوب می شود. واژه ای که دارای این فاصله رسایی باشد الگوی قالب افتان بودن رسایی در مرز هجا را نقض کرده است. همانطور که در جدول ۶ مشخص است در زبان فارسی نمونه هایی از تمامی فاصله های رسایی وجود دارد. بنا به بررسی های انجام شده در (Gouskova, 2004) هر زبان یک میزان خاص از تفاوت رسایی (Cutoff Point) در مرز هجا را اجازه می دهد(شکل ۱).

جدول ۱- سلسله مراتب رسایی گروه های واجی زبان فارسی به همراه عبارت مخفف و درجه رسایی آنها

ردیف	گروه همخوان	مخفف	تعداد همخوان	درجه رسایی
۱	Liquids	LI	۳	۵
۱	Nasals	NA	۲	۴
۳	Fricatives	FR	۹	۳
۴	Affricates	AF	۲	۲
۵	Plosives	PL	۸	۱

واژه‌های دارای مرز هجا با خوشه همخوانی از واژگان زبان فارسی استخراج شده و میزان تفاوت رسایی آنها در مرز بین دو هجا اندازه‌گیری شد.

استخراج بسامد ها و محاسبه معیار اطلاعات مشترک نقطه ای

به ازای هر خوشه همخوانی در مرز هجا مانند C1.C2 اختلاف رسایی با توجه به طبقه رسایی به دست آمده و تعداد رخداد آن خوشه هم در واژگان (بسامد واژگانی^۱) و هم با استفاده از بسامد رخداد هر واژه در پیکره (بسامد پیکره^۲) محاسبه شده است. دلیل محاسبه بسامد واژگانی و نیز بسامد در پیکره این است که ممکن است یک محدودیت در سطح واژگانی اعمال نشود اما در بسامد پیکره خود را نشان دهد. به عنوان مثال ممکن است در واژگان، واژه ها بدون توجه به محدودیت های خوش ساختی اعمال گردند اما بعدا در کاربرد کلمات خوش ساخت تر بیشتر استفاده شده و بنابراین بسامد پیکره ای بالا اما بسامد واژگانی پایین داشته باشند.

اطلاعات مشترک نقطه ای

برای محاسبه میزان تمایل به هم‌رخدادی رخدادهایی از دو متغیر تصادفی از معیار اطلاعات مشترک نقطه ای استفاده می گردد. این معیار نشانگر میزان هم‌رخدادی دو رخداد فارغ از میزان رخداد هر یک به صورت جداگانه می باشد. اطلاعات مشترک نقطه ای دو رخداد x و y از دو متغیر تصادفی X و Y به صورت زیر محاسبه می گردد.

$$PMI(x, y) = \log_2 \frac{p(x, y)}{p(x) * p(y)} \quad (1)$$

¹ Type Frequency

² Token Frequency

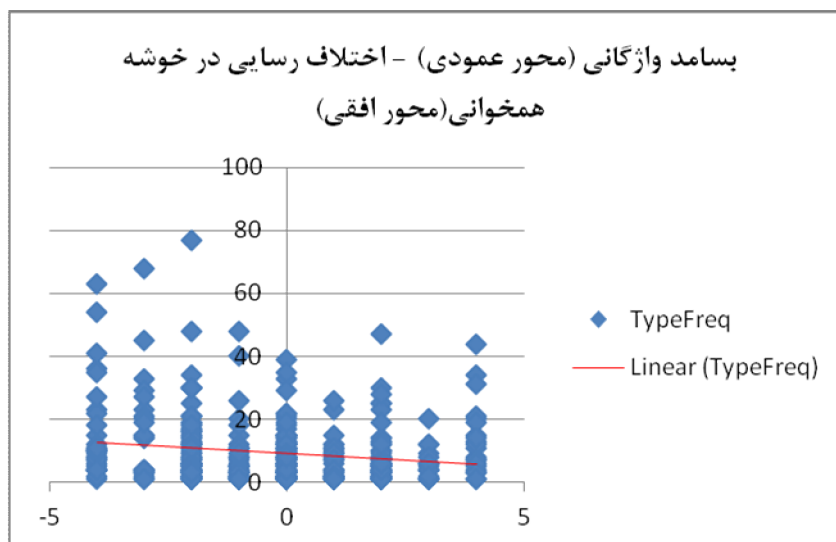
هر چه میزان اطلاعات مشترک نقطه ای دو متغیر بیشتر باشد به این معنی است که تمایل به رخ داد همزمان دو متغیر بیشتر است. این معیار جهت بررسی میزان تمایل دسته های مختلف رسایی در مرز بین دو هجا به کار رفته است. از این معیار در (مک گوان^۱، ۲۰۰۸) نیز جهت بررسی محدودیت رسایی در مرز هجا بهره گرفته شده است. برای در نظر گرفتن سطوح مختلف تاثیر محدودیت الگوی تغییر رسایی در مرز هجا بسامد و معیار اطلاعات مشترک نقطه ای در هر دو سطح واژگان و پیکره محاسبه شده اند. بسامد واژگانی در نمودار-۱ و بسامد پیکره ای در نمودار-۲ نشان داده شده است. همچنین با استفاده از این بسامد ها معیار اطلاعات مشترک نقطه ای برای هر دو حالت ممکن خوشه همخوانی به دست آمد. معیار اطلاعات مشترک نقطه ای نمایانگر میزان تمایل دو همخوان به رخداد در کنار یکدیگر در مرز هجا می باشد. هرچه معیار اطلاعات مشترک نقطه ای مثبت تر باشد نشانگر تمایل بیشتر دو همخوان به هم‌رخدادی است. مقادیر منفی نشانگر عدم تمایل دو همخوان به هم‌رخدادی است. در صورتی که محدودیت رسایی در مرز هجا برای مقادیر افتان بر زبان حکم فرما باشد انتظار می رود که برای الگوهای تغییر رسایی افتان مقدار معیار اطلاعات مشترک نقطه ای مثبت تر و برای الگوهای تغییر رسایی خیزان این معیار منفی تر باشد. این معیار یک بار با استفاده از بسامد واژگانی در نمودار-۳ و یک بار با استفاده از بسامد پیکره ای در نمودار-۴ نشان داده شده است.

۴. نتایج

در نمودار-۱ بسامد واژگانی به ازای اختلاف رسایی دو همخوان در خوشه همخوانی مشاهده می شود. خط قرمز نشانگر تمایل کلی تغییرات با استفاده از درون یابی^۲ به دست آمده است. همانطور که دیده می شود تفاوت بارزی بین بسامد واژگانی خوشه های با الگوی تغییر رسایی افتان و خوشه های با الگوی تغییر رسایی خیزان مشاهده نمی گردد هر چند تمایل قالب زبانها به پیروی از الگوی افت رسایی در مرز هجاست.

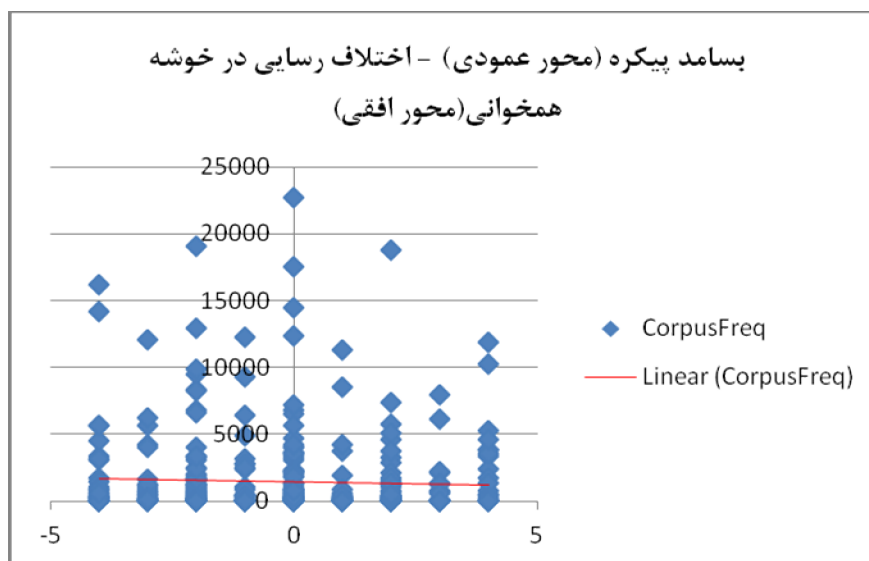
¹ K. B. McGowan

² Interpolation



نمودار ۱ - بسامد واژگانی به ازای اختلاف رسایی دو همخوان در همخوان های C1 و C2 در مرز هجا

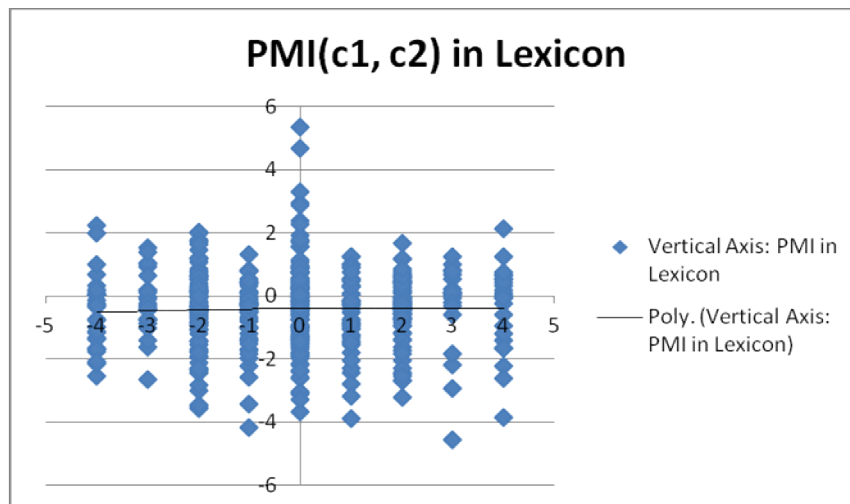
در نمودار ۲- بسامد پیکره ای به ازای اختلاف رسایی دو همخوان در خوشه همخوانی نشان داده شده است. همانند بسامد واژگانی در این نمودار نیز ارتباط مشخصی بین اختلاف رسایی دو همخوان و بسامد رخداد آنها در پیکره مشاهده نمی گردد. هر چند به میزان ناچیزی تمایل به پیروی از اصل افت رسایی در مرز هجا قابل تشخیص است.



نمودار ۲ - - بسامد در پیکره به ازای اختلاف رسایی دو همخوان در همخوان های C1 و C2 در مرز هجا

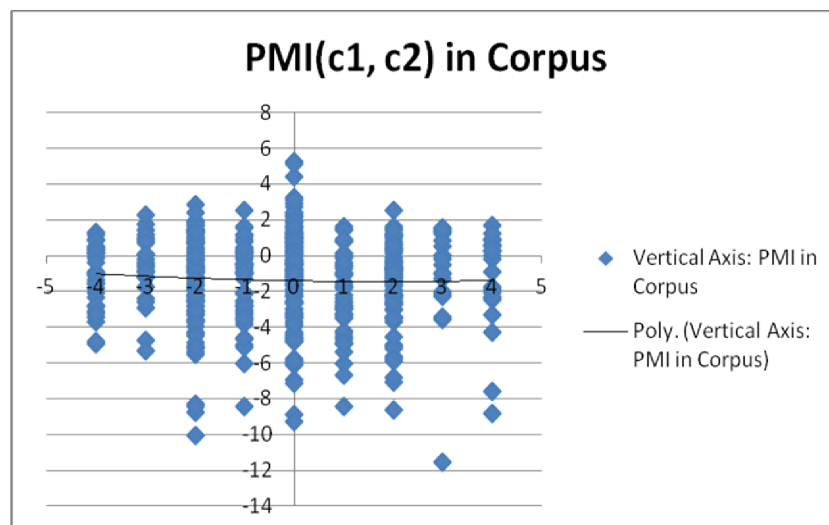
در نمودار ۳- معیار اطلاعات مشترک نقطه ای با استفاده از بسامد واژگانی مشاهده می گردد. این معیار تقریباً در همه جا نزدیک به صفر است. همانطور که پیش تر گفته شد مقدار صفر برای معیار اطلاعات مشترک نقطه ای

بین دو متغیر نشانگر استقلال رخداد دو متغیر نسبت به یکدیگر می باشد، که به این معنی است که ارتباط مستقیمی بین بسامد رخداد واژگانی و الگوی رسایی وجود ندارد.



نمودار ۳ - اطلاعات مشترک نقطه ای بین دو متغیر میزان اختلاف رسایی بین همخوان اول و دوم در خوشه همخوانی بین دو هجا و بسامد واژگانی

در نمودار ۴ معیار اطلاعات مشترک نقطه ای برای دو متغیر بسامد در پیکره و میزان اختلاف رسایی بین دو همخوان در خوشه همخوانی در مرز هجا مشاهده می گردد. از آنجا که اطلاعات مشترک نقطه ای تقریباً در همه جا نزدیک به صفر است می توان نتیجه گرفت که رابطه ای بین افتان یا خیزان بودن الگوی تغییر رسایی در مرز هجا و بسامد رخداد آن وجود ندارد.



نمودار ۴ - اطلاعات مشترک نقطه ای بین دو متغیر میزان اختلاف رسایی بین همخوان اول و دوم در خوشه همخوانی بین دو هجا در ساختار و بسامد در

پیکره

۵. بحث و نتیجه گیری

نتایج به دست آمده نشانگر این است که بسامد رخداد ساختارهایی با الگوی رسایی خیزان و افتان در مرز هجا در زبان فارسی تفاوتی با یکدیگر ندارند. همچنین نتایج نشان می دهد که اطلاعات مشترک نقطه ای بین دو متغیر بسامد و میزان اختلاف رسایی در خوشه همخوانی CC مرز هجا تقریبا برابر با صفر است. که نشان می دهد که این دو متغیر مستقل از یکدیگر بوده و هیچ گونه ارتباطی بین الگوی تغییر رسایی در مرز هجا و بسامد خوشه همخوانی وجود ندارد. این موضوع هم در سطح واژگان با استفاده از بسامد واژگانی و هم در سطح پیکره با استفاده از بسامد در پیکره به دست آمده است.

محدودیت های خوش ساختی باعث می شود تا گونه هایی که کمتر خوش ساخت هستند، به مرور زمان خوش ساخت تر شده و به مرور بسامد گونه های خوش ساخت در واژگان بالاتر رود. هم چنین انتظار می رود واژه های خوش ساخت تر بیشتر به کار رفته و بسامد پیکره ای بالاتری داشته باشند. بنابراین در صورتی که محدودیتی در سطح زبان اعمال گردد انتظار است که بسامد رخداد آن ساختارهایی که محدودیت را نقض نکرده اند به مرور بیشتر گردد. با توجه به اینکه نتایج به دست آمده نشانگر عدم وجود رابطه بین الگوی تغییر رسایی در خوشه های همخوانی در مرز هجا و بسامد می باشد می توان نتیجه گرفت که محدودیت افتان بودن الگوی رسایی در مرز هجا در زبان فارسی اعمال نمی گردد.

با توجه به اینکه در برخی از زبان ها مانند زبان های قرقیزی و قزاقی که گونه هایی از زبان ترکی هستند محدودیت رسایی در مرز هجا اعمال می شود می بایست عدم اعمال این محدودیت در زبان فارسی دلیلی داشته باشد. در زبان فارسی نه تنها این محدودیت به صورت مقوله ای ساختارها را به خوش ساخت و بدساخت تبدیل نمی کند بلکه این محدودیت، برخلاف زبان انگلیسی در فارسی خوش ساختی طیفی ساختارها (gradual wellformedness) را نیز ایجاد نمی کند. دلیل این امر را می توان قطعی بودن تشخیص مرز هجا در زبان فارسی دانست. در صورتی که در زبانی پیش هسته و پس هسته هر دو بتوانند دارای خوشه همخوانی باشند لازم می شود تا محدودیتی ایجاد گردد تا تشخیص مرز هجا توسط شنونده میسر گردد. به عنوان مثال ساختار CVCCV را در نظر بگیرید. این ساختار را می توان به صورت های مختلفی تقطیع هجایی نمود. این صورت ها شامل (CV,CCV),(CVC,CV),(CVCC,V) می باشد. با توجه به اینکه زبان فارسی اجازه وجود یک هجای تک واکه ای و خوشه همخوانی در پیش هسته را نمی دهد نیاز به محدودیتی برای مشخص نمودن مرز هجا ندارد و تقطیع هجایی مجاز به صورت قطعی (CVC,CV) خواهد بود. اما در زبان هایی مانند انگلیسی که

ساختار هجایی پیچیده تری دارند و تشخیص مرز هجا گاه توسط گویشوران زبان نیز محل تردید است، نیاز به محدودیت هایی وجود دارد تا حالات تقطیع هجایی را کمینه کرده و تشخیص مرز هجا را آسان تر نمایند. هر چه حالات ممکن تقطیع در زبانی بیشتر باشد نیاز به محدودیت های قوی تری برای مشخص نمودن مرز هجا می باشد. بنابراین زبان فارسی به علت قطعی بودن مرز هجاها نیازی به اعمال محدودیت رسایی در مرز هجا نخواهد داشت. در صورتی که در نظریه بهینگی محدودیت ها را جهانی فرض کنیم، می بایست اولویت این محدودیت را برای زبان فارسی بسیار کم و نزدیک به صفر در نظر بگیریم و یا محدودیت را به صورت پارامتریک در نظر بگیریم تا در مورد زبان هایی با پارامتر مرز هجای قطعی اولویت بسیار پایینی به آن اختصاص داده شود. در صورتی که اعمال محدودیت در مرز هجا در چند زبان با مرز هجای قطعی (مانند فارسی و فرانسوی) و چند زبان با مرز هجای غیر قطعی (مانند انگلیسی و قزاق) به روش های آماری اشاره شده در این مقاله انجام گردد، می توان فرضیه ارایه شده در این مقاله را به صورت رده شناختی نیز بررسی نمود.

۶. مراجع

Prince, A., and P. Smolensky (1993). *Optimality theory: Constraint interaction in generative grammar*. Wiley.

Blevins, J. (1995). The Syllable in phonological theory. In *The handbook of phonological Theory*, ed. John A. Goldsmith. 206-244. Cambridge, Mass, and Oxford, UK:Blackwell.

Burquest, D. A. (1993). *Phonological analysis: A functional approach*. Dallas, TX: Summer Institute of Linguistics.

Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In *Papers in laboratory phonology I: Between the grammar and the physics of speech*. New York: Cambridge University Press. 283-333

Flemming, E. (1996). Evidence for Constraints on Contrast: The Dispersion Theory of Contrast. *UCLA Working Papers in Phonology I*. 86-106

Gouskova, M. (2004). *Relational hierarchies in Optimality Theory: the case of syllable contact*. *Phonology* 21:201-250. Cambridge Univ Press.

Harris, J. (2005). The phonology of being understood: Further arguments against sonority. *Lingua* , 55(10).

Kager, R. (1999). *Optimality theory*. Cambridge University Press.

Kahn, D. (1976). *Syllable-based Generalisations in English Phonology*.

Lindblom, B. (1996). *Systemic constraints and adaptive change in the formation of sound structure*. Edinburgh: Evolution of Human Language.

M. Eslami, M. Bijankhan (2003). vâžegozini va vâjârâyi. *Proceedings of 2th Conference of Word-formation and Terminology.*, Tehran, Iran: Academy of Persian language and literature 1.

McGowan, K. (2008). *Gradient Lexical Reflexes of Syllable Contact Law*.

Vennemann, T. (1988). *Preference laws for syllable structure and the explanation of sound change: With special reference to German, Germanic, Italian, and Latin*. Berlin: Mouton de Gruyter.