شنوایی‌شناسی ـ دوره ۲۳، شماره ۳، ۱۳۹۳، ۳۸-۳۰

مقاله پژوهشی

محاسبهٔ عدد کوهورت برای فهرست کلمات دوهجایی پربسامد زبان فارسی

سهیلا شایان مهر[۱]، جمیله فتاحی[۱]، سید علی‌اکبر طاهایی[۲]، شهره جلایی[۳]

[۱]ـ گروه شنوایی‌شناسی، دانشکده توانبخشی، دانشگاه علوم پزشکی تهران، ایران

[۲]ـ گروه شنوایی‌شناسی، دانشکده علوم توانبخشی، دانشگاه علوم پزشکی ایران، تهران، ایران

[۳]ـ آمار زیستی، دانشکده توانبخشی، دانشگاه علوم پزشکی تهران، ایران

چکیده

زمینه و هدف: تعداد رقیب‌های ممکن برای هر کلمه که ابتدای آنها با هم مشابه است «عدد کوهورت» نام دارد. علی‌رغم اهمیت تعداد و ویژگی‌های رقیب‌ها در شناسایی کلمه، در هیچ یک از آزمون های ساخته شده برای زبان فارسی عامل عدد کوهورت در انتخاب کلمه در نظر گرفته نشده است. هدف مطالعه کنونی، معرفی نقش عدد کوهورت در شناخت کلمه و سپس محاسبهٔ عدد کوهورت برای کلمات دوهجایی پربسامد زبان فارسی بود.

روش بررسی: پس از تهیهٔ پیکرهٔ واژه‌های پر بسامد زبان فارسی و استخراج کلمات دوهجایی آنها، عدد کوهورت هر یک از کلمه‌های دوهجایی از فرهنگ فارسی عمید محاسبه شد. به این ترتیب یک فهرست کامل از کلمات دوهجایی پر بسامد مشتمل بر ۴۱۲۱ واژه به همراه مقدار کوهورت برای هر یک از کلمات به‌دست آمد.

یافته‌ها: مقادیر کوهورت کلمات محدوده‌ای از صفر تا ۸۷ را شامل شدند. نیمی از کلمات مورد بررسی عدد کوهورت بالای ۱۴ و نیمی عدد کوهورت زیر ۱۴ داشتند.

نتیجه‌گیری: عدد کوهورت بر سرعت و دقت تصمیم‌گیری درکی کلمه تأثیر دارد. کلمات فارسی از نظر متغیر کوهورت وزن یکسانی ندارند. از این رو برای طراحی مواد آزمونی کنترل شده‌تر در ساخت انواع مختلف آزمون‌های شنوایی، می‌توان عامل عدد کوهورت را نیز در کنار سایر عوامل مؤثر در نظر داشت.

واژگان کلیدی: درک گفتار، شناخت کلمه، مدل کوهورت شناخت کلمه، عدد کوهورت، زبان فارسی

(دریافت مقاله: ۹۲/۲/۱۷، پذیرش: ۹۲/۶/۲۶)

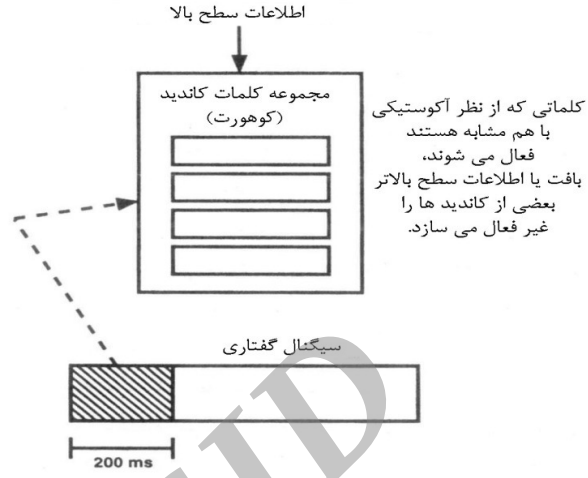
مقدمه

گفتار یک سیگنال پیچیده و متغیر است که به‌صورت توالی‌ای از واحدهای معنی‌دار زبانی درک می‌شود. به‌منظور درک یک جمله شنونده باید کلمات منفرد را شناسایی کند و به معنی هر یک از کلمات دست یابد و آنها را برای ایجاد معنی درست با هم ترکیب کند(۱و۲). با شنیدن بخش ابتدایی کلمه تعدادی جایگزین لغوی رقیب برای فعال شدن با هم به رقابت می‌پردازند و در نهایت با پردازش‌های بالا به پایین صورت گرفته یکی از رقیب‌ها به‌طور هم‌زمان برنده می‌شود(۳و۴).

یکی از مدل‌های نظری درک گفتار، مدل کوهورت است. در این مدل، برای هر کلمهٔ هدف می‌توان تعداد رقیب‌های ممکن برای آن کلمه را محاسبه کرد و آن مقدار، عدد کوهورت (Cohort size) یا تراکم هم‌جواران (the neighborhood density) نامیده می‌شود.

در نظریهٔ کوهورت دو مرحله برای تشخیص کلمهٔ گفتاری به این شرح پیشنهاد می‌شود مرحلهٔ مستقل و دیگری مرحلهٔ تعاملی. در مرحلهٔ اول (مستقل) اطلاعات فیزیکی‌ـآوایی مربوط به ابتدای کلمهٔ درون‌داد برای تمام کلماتی که اطلاعات ابتدای آنها با هم یکسان هستند در حافظه برانگیخته می‌شوند. کلمات برانگیخته شده بر مبنای اطلاعات مربوط به آغاز کلمه، یک گروه یا اصطلاحاً کوهورت تشکیل می‌دهند. برای مثال، اگر گوینده کلمه‌ای که با هجای «سر» شروع می‌شود به‌کار ببرد در حافظهٔ شنونده کلماتی که با هجای «سر» شروع می‌شوند برانگیخته

نویسنده مسئول: تهران، خیابان انقلاب، بعد از پیچ شمیران، دانشکده توانبخشی، دانشگاه علوم پزشکی تهران، گروه شنوایی‌شناسی، کد پستی: ۱۱۴۸۹۶۵۱۴۱، تلفن: ۰۲۱-۷۷۵۳۰۶۳۶، E-mail: jfatahi@tums.ac.ir



شکل ۱ـ مدل کوهورت تشخیص کلمه(۱۳)

می‌شوند مانند سرباز، سرما، سردی، سرشیر، سرداب، سرور، سرریز، سردرد، و سرگرم. هنگامی که یک گروه برانگیخته می‌شود کلمات نامناسب از طریق اطلاعات بالا به پایین از قبیل قواعد نحوی یا معنایی غیرفعال می‌شوند (به‌صورت داوطلب‌هایی از گروه حذف می‌شوند). تمامی منابع احتمالی اطلاعات شامل منابع اطلاعاتی سطح بالاتر، در فرایند انتخاب کلمهٔ مناسب از گروه دخالت دارند. برای مثال، اگر کلمهٔ «سرباز» برای آن بافت گفتار مناسب نباشد از گروه حذف خواهد شد. این همان مرحلهٔ دوم تشخیص کلمه یعنی مرحلهٔ تعاملی است. شکل ۱ نمایی طرحواره‌ای از مدل کوهورت را نشان می‌دهد.

Marslen-Wilson (۱۹۹۰) اثر بسامد رقیب‌ها بر شناخت کلمات را بررسی کرد و دریافت که افزایش فعال‌سازی برای یک کلمهٔ پربسامد بیشتر از آن برای یک کلمهٔ کم‌بسامد است(۵). نتایج مطالعات Luce و همکاران (۱۹۸۹ و ۱۹۹۰) حاکی از آن است که عدد کوهورت (تعداد رقیب‌ها) بر دورهٔ زمانی شناخت کلمه اثر می‌گذارد. تعداد رقیب‌ها (عدد کوهورت) و ویژگی‌های رقیب‌ها (مانند بسامد) در تشخیص کلمه مهم است. برای مثال، ما کمتر می‌توانیم کلمات کم‌بسامدی را که هم‌جواران زیاد و پربسامد دارند نسبت به کلمات دارای هم‌جواران کمتر و کم‌بسامدتر شناسایی کنیم. Luce و همکاران (۱۹۹۰) عقیده داشتند عدد کوهورت که آنها آن را تراکم هم‌جواران نامیدند بر تصمیم‌گیری اثر می‌گذارد. کلماتی با هم‌جواران بیشتر دیرتر شناسایی می‌شوند و به‌دلیل رقابت، خطاهای بیشتری تولید می‌کنند(۵).

Theunissen و همکاران (۲۰۰۹) در مروری نظام‌مند به اثر متغیرهای مختلف بر طراحی و تفسیر آزمون‌های جمله در نویز، اشاره می‌کند که نوع و بافت مواد آزمون بر عملکرد افراد اثرگذار است. در این مرور جامع به اهمیت نقش بسامد کلمه، تعداد کلمات مشابه از نظر آوایی (تراکم هم‌جواران یا عدد کوهورت) و بسامد رقیب‌های کلمه در شناخت کلمه اشاره شده است. همچنین در مطالعات نشان داده شده است که افراد کم‌شنوا برای اینکه بتوانند درصد یکسانی از کلمات دشوار از نظر لغوی (کلمات کم‌بسامد و با تعداد همجواران زیاد و پربسامد) و کلمات ساده از نظر لغوی (کلمات پربسامد و با تعداد همجواران کم و کم‌بسامد) را تشخیص دهند به شدت ۵/۸– ۳/۱ دسی‌بل بیشتری نیاز دارند(۶).

درک گفتار با واحدهای زبانی مختلفی از جمله واج‌ها، هجاها، کلمات یا جملات ارزیابی می‌شود. مواد آزمونی گوناگونی برای اندازه‌گیری درک گفتار با این واحدها طراحی شده‌اند(۷و۸). تاکنون مواد کلمه‌ای مختلفی برای آزمون‌های گوناگون در فارسی طراحی شده است، ولی در هیچ یک از آنها عامل عدد کوهورت برای انتخاب کلمات درنظر گرفته نشده است. همواره امتیاز درک بعضی کلمات پایین‌تر از برخی دیگر است. از دلایل آن می‌توان به عدم کنترل میزان آشنایی، بسامد و عدد کوهورت کلمه اشاره کرد. می‌توان گفت هرچه عدد کوهورت کلمه‌ای بیشتر باشد، یا به عبارتی تعداد کلمات مشابه بیشتری برای فعال شدن با هم رقابت کنند، آن کلمه غیرقابل پیش‌بینی‌تر می‌شود و با شنیدن ابتدای کلمه به‌راحتی نمی‌توان کل کلمه را حدس زد. هنگام انتخاب کلمات برای اهداف مختلف در نظر گرفتن قابلیت پیش‌بینی و حدس کلمه مهم است. برای مثال، اگر در آزمون بازشناسی گفتار (speech recognition threshold) فهرست کلمات از این نظر متوازن و متعادل نشده باشند و فهرست صرفاً از کلمات کم‌کوهورت تشکیل شده باشد احتمالاً امتیاز حاصل از این فهرست با فهرست دیگری که از کلمات پرکوهورت تشکیل شده است

متفاوت خواهد بود و این اختلاف امتیاز به اشتباه به‌عنوان تفاوت عملکرد افراد تلقی خواهد شد.

با توجه به اینکه تاکنون چنین محاسبه‌ای برای کلمات زبان فارسی صورت نگرفته است این فهرست در طراحی مواد آزمونی دقیق‌تر و کنترل شده‌تر می‌تواند مفید باشد. از این رو این مطالعه با هدف محاسبهٔ تعداد رقیب‌های با هجای اول مشابه در کلمات دوهجایی پربسامد زبان فارسی انجام شد.

روش بررسی

در مطالعهٔ حاضر که از نوع توصیفی‌ـاکتشافی بود، پس از تهیهٔ پیکرهٔ واژه‌های پربسامد زبان فارسی، کلمات دوهجایی آنها استخراج و عدد کوهورت هریک از کلمه‌های دوهجایی از فرهنگ فارسی عمید محاسبه شد. به این منظور از پیکرهٔ زبانی گردآوری شده توسط عاصی (۱۹۹۷) در پژوهشگاه علوم انسانی و مطالعات فرهنگی استفاده شد. وی از سال ۱۳۷۲ با جمع‌آوری اطلاعات از منابع زیر به ساخت پایگاه داده‌ای زبان فارسی در اینترنت پرداخته‌اند. نمونه‌های وارد شده که به‌عنوان داده‌ها در حافظهٔ رایانه ذخیره شده‌اند شامل متون زبانی و متون مهم نظم و نثر ادبیات معاصر ایران، کتب و مقالات تخصصی، گفتار پیوستهٔ ضبط شده، متن روزنامه‌ها و غیره است که به پایگاه درون‌داد شده‌اند. مجموع متن‌های گردآوری شده نزدیک به یکصد میلیون واژه است که تاکنون تنها ۶۰ میلیون واژه از آن به درون پایگاه داده وارد شده است. قابل ذکر است که این کار به‌صورت فعالیتی همیشگی و با افزودن منابع تازه دنبال خواهد شد. در این پایگاه داده‌ها به شکل‌ها و قالب‌های گوناگون نظیر فهرست‌های واژه‌نما و بسامدی ذخیره شده‌اند. پایگاه داده‌های زبان فارسی مجموعه‌ای است از متون مختلف فارسی که بخشی از آن دارای نشانه‌گذاری‌هایی از جمله شناسنامه متن، برچسب‌های دستوری، آوایی، ریشه‌ای و معنایی است(۹). بنابراین امکان تهیهٔ فهرست‌های بسامدی واژه‌ها از این پایگاه داده وجود داشت. به این ترتیب فهرست بسامدی واژه‌ها که شامل ۱۴۰۰۰ واژهٔ مختلف بود و واژه‌ها به دو صورت الفبایی و بسامدی در فهرست مرتب شده بودند برای این مطالعه تهیه شد. پس از تهیه و بررسی فهرست ۱۴۰۰۰ واژه‌ای کلمات پربسامد، تمام کلمات دوهجایی، به‌جز اسامی خاص، قیدها و حروف ندا، از بین آنها استخراج شدند. در این مرحله یک فهرست کلمات دوهجایی پربسامد مشتمل بر ۴۱۲۱ واژه به‌دست آمد. گام بعدی محاسبهٔ عدد کوهورت یا تراکم هم‌جواران برای فهرست کلمات دوهجایی بود. با توجه به اینکه هجای اول در تولید کوهورت اهمیت بسیاری دارد محاسبهٔ عدد کوهورت از روی هجای اول کلمات صورت گرفت؛ به این صورت که با استفاده از فرهنگ فارسی دو جلدی عمید، تمام کلماتی که هجای اول مشابه با کلمهٔ هدف داشتند شمارش و به‌عنوان عدد کوهورت آن کلمه در نظر گرفته شد. به‌منظور شمارش عدد کوهورت، کلمات بیگانه و غیرفارسی و نیز کلمات قدیمی و مستعمل که در فارسی معاصر شناخته شده نیستند و به کار نمی‌روند در محاسبه منظور نشدند. برای مثال، برای کلمهٔ «دفتر» که یکی از کلمات فهرست ۴۱۲۱ واژهٔ دوهجایی است در فرهنگ فارسی فقط کلمهٔ «دفعه» دارای هجای اول «دف» است. بنابراین عدد کوهورت کلمهٔ «دفتر» یک در نظر گرفته می‌شود. جدول ۱ مثالی از نحوهٔ محاسبهٔ عدد کوهورت برای چند کلمه را نشان می‌دهد. این روند برای تمام ۴۱۲۱ کلمهٔ دهجایی انجام شد و مقادیر کوهورت هر یک از کلمات به‌دست آمد.

یافته‌ها

با محاسبهٔ مقادیر کوهورت، کلمات دوهجایی محدودهٔ وسیعی از مقادیر کوهورت را نشان دادند. کمترین مقدار عدد کوهورت برای کلمات دوهجایی پربسامد زبان فارسی، یعنی صفر، به‌معنی فاقد رقیب بودن کلمات فوق و بیشترین میزان آن، یعنی ۸۷، به‌معنی وجود ۸۷ کلمه با هجای اول مشابه به‌دست آمد؛ یعنی برای کلمات بررسی شده عدد کوهورت در محدودهٔ بین صفر تا ۸۷ بود. نمودار ۱ کلمات با مقادیر کوهورت مختلف را نشان می‌دهد.

نتایج نشان می‌دهد که از بین ۴۱۲۱ کلمهٔ دوهجایی بررسی شده، ۵۰ درصد کلمات عدد کوهورتی پایین‌تر از ۱۴ و ۵۰ درصد دیگر عدد کوهورت بالاتر از ۱۴ داشتند. همچنین عدد

جدول ۱ـ مثالی از محاسبه عدد کوهورت برای کلمات

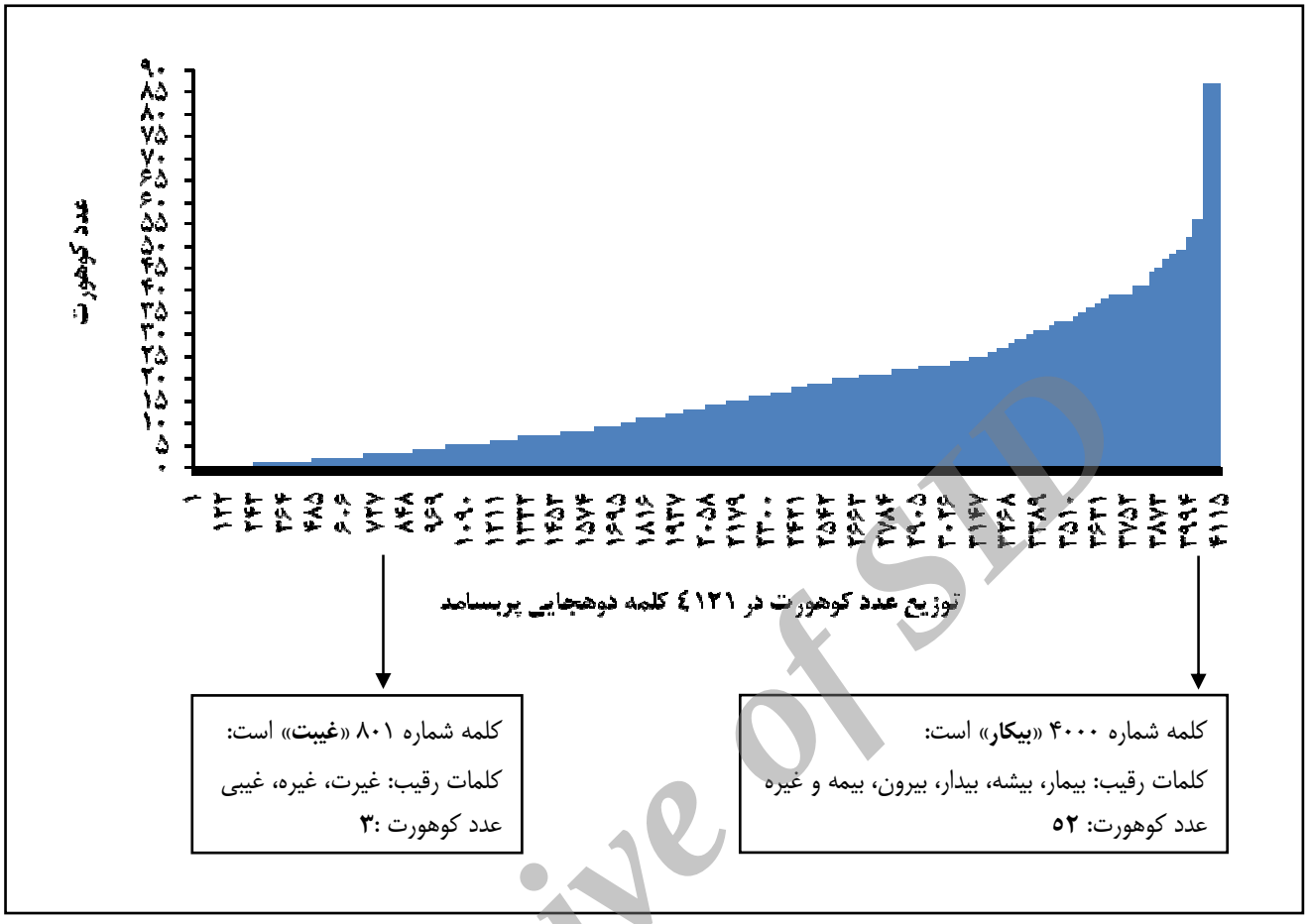
کلمه دوهجایی	بسامد کلمه	کلمات رقیب	عدد کوهورت
آفتاب	۲۳۰۱	–	۰
صحرا	۸۱۶	صحنه	۱
دکمه	۳۰۳	دکتر ـ دکان	۲
خروس	۲۶۰	خنک ـ خدا ـ خصوص ـ خروش ـ خروج ـ خطوط ـ خلود ـ خمار ـ خطور ـ خلوص	۱۰
پاییز	۱۴۳۴	پابند ـ پابوس ـ پاپی ـ پاپیچ ـ پاتوق ـ پاتیل ـ پاچه ـ پاچید ـ پادار ـ پاداش ـ پادرد ـ پادو ـ پاراف ـ پاریس ـ پارو ـ پاره ـ پاساژ ـ پاسور ـ پاشید ـ پاکی ـ پاکت ـ پاگیر ـ پالان ـ پالش ـ پالیز ـ پایان ـ پایه ـ پایین ـ پاسخ ـ پابست	۳۰

کوهورت یک بیشترین تعداد کلمات را به‌خود اختصاص می‌داد. به‌عبارتی دیگر، ۲۳۹ کلمه، معادل ۵/۷۹ درصد، دارای عدد کوهورت یک بودند. پس از آن، عدد کوهورت صفر رتبهٔ دوم را داشت که شامل ۲۳۷ کلمه، معادل ۵/۷۵ درصد بود. یعنی در بین ۴۱۲۱ کلمهٔ دوهجایی پربسامد بررسی شده، در ۵/۷۵ درصد آنها با شنیدن هجای اول هیچ کلمهٔ دیگری با آنها رقابت نمی‌کند و بنابراین به‌راحتی حدس زده می‌شوند. همچنین کلمات پربسامدی وجود داشتند که دارای عدد کوهورت بسیار پایین بودند و عکس آن. برای مثال، در کلمهٔ پرکاربرد «بچه» با بسامد ۸۹۶۲، عدد کوهورت صفر بود. مثال دیگر کلمهٔ کم‌کاربرد «آژیر» با بسامد پایین ۸۴ و با بیشترین مقدار عدد کوهورت، یعنی ۸۷، بود. جدول ۲ نمونه‌ای از کلماتی است که ارتباط عکس بین بسامد و عدد کوهورت آنها وجود دارد.

یافتهٔ کاربردی دیگر این است که بیشترین مقدار کوهورت (۸۷) متعلق به هجای اول «آ» بود. به‌عبارت دیگر، برای هجای اول «آ» ۸۷ کلمهٔ رقیب وجود دارد که با شنیدن این هجای اول همهٔ این ۸۷ کلمه برای فعال شدن با هم به رقابت می‌پردازند. در کلمات بررسی شده ۷۰ کلمه عدد کوهورت ۸۷ داشتند. ارتباط بین تعداد کلمات دوهجایی با عدد کوهورت متناظر آن را در نمودار ۲ به‌خوبی می‌توانید مشاهده کنید.

بحث

در مطالعهٔ انجام شده، پس از استخراج واژگان دوهجایی از فهرست کلمات پربسامد زبان فارسی، عدد کوهورت برای واژگان دوهجایی پربسامد محاسبه شده است. علت انتخاب واژگان پربسامد این است که برای اهداف بسیاری، از جمله در ساخت آزمون‌های مختلف، از کلمات رایج و پرکاربرد آن زبان استفاده می‌شود و معمولاً کلمات نادر یا بیگانه مورد استفاده قرار نمی‌گیرند. از طرفی برای شناخت کلمه، کوهورتی (مجموعه‌ای) از کلمات احتمالی با کلمهٔ هدف به رقابت می‌پردازند. نقش بسامد کلمه در اینجا برجسته می‌شود. در مدل کوهورت، در مرحلهٔ اولیهٔ دستیابی لغوی، بسامد کلمه بر شدت فعال‌سازی کلمات داوطلب اثر می‌گذارد؛ به این صورت که برای کلمات پربسامدتر میزان بهرهٔ فعال‌سازی بیشتر است (کلمات داوطلب پربسامدتر با بهرهٔ بیشتری فعال می‌شوند). بسامد تنها یکی از عوامل تأثیرگذار بر شناخت کلمه است. عامل مهم دیگر میزان آشنایی کلمه است. هرچه کلمه برای شنونده آشناتر باشد راحت‌تر و سریع‌تر شناسایی می‌شود(۵). از دیگر عوامل مهم که کمتر مورد توجه واقع می‌شود تراکم هم‌جواران یا عدد کوهورت کلمه است. با داشتن مقادیر کوهورت کلمات، می‌توان کلمات غیر قابل پیش‌بینی‌تر را انتخاب و جدا کرد و به مقاصد گوناگون از آنها بهره برد. البته بسامد رقیب‌ها نیز بر شناخت کلمهٔ هدف اثرگذار است و کلمهٔ پربسامدی که تنها



نمودار ۱ـ روند تغییرات عدد کوهورت

رقیب‌های کم‌بسامد دارد سریع‌تر شناسایی می‌شود و عکس آن. اهمیت و نقش پیش‌بینی کلمه با شنیدن ابتدای کلمه، کمک به درک گفتار در شرایط شنیداری دشوار، که به‌طور روزمره رخ می‌دهد، است(۱۲-۱۰). وقتی سیگنال گفتاری با نویز زمینه تخریب می‌شود علائم بافتی اهمیت ویژه‌ای می‌یابند. هرچه علائم بافتی در گفتار بیشتر باشد شنونده کمتر به ویژگی‌های آکوستیکی دقیق صدا اتکا می‌کند. مفهوم کاربردی مدل کوهورت این است که در ساخت آزمون‌های مختلف، به‌ویژه آزمون‌های جمله در حضور نویز، دشواری لغوی کلمه مورد استفاده بر دشواری و دقت آزمون اثر خواهد گذاشت(۶). پیش‌بینی کلمه در درک موفق زبان، حیاتی و مهم است. برای ارزیابی درک گفتار در نویز بهتر است کلمات مورد استفاده غیر قابل پیش‌بینی‌تر باشند (دارای عدد کوهورت بالاتر باشند) تا با کاهش علائم بافتی توجه فرد به کلمه‌ای که واقعاً در حضور نویز شنیده است متمرکز شود نه آن چیزی که با شنیدن ابتدای کلمه حدس زده است.

همان‌طورکه در بالا اشاره شد و در نمودارهای ۱ و ۲ نیز دیده می‌شود ۵۰ درصد کلمات دوهجایی پربسامد زبان فارسی عدد کوهورت پایین‌تر از ۱۴ دارند. همچنین بیشترین تعداد کلمات مورد بررسی دارای عدد کوهورت به‌ترتیب یک و صفر هستند. در ۴۱۲۱ کلمهٔ دوهجایی پربسامد مورد بررسی در مجموع حدود ۱۱/۶ درصد کلمات، کمترین مقادیر کوهورت یعنی اعداد یک و صفر را به‌خود اختصاص داده‌اند. اگر این یافته را به کل کلمات فارسی تعمیم دهیم می‌توانیم نتیجه بگیریم اکثر کلمات فارسی دارای عدد کوهورت پایین هستند و تعداد کلمات با عدد کوهورت بالا یا به

جدول ۲ـ نمونه‌ای از کلمات نشان‌دهنده عدم ارتباط بین بسامد و مقدار کوهورت

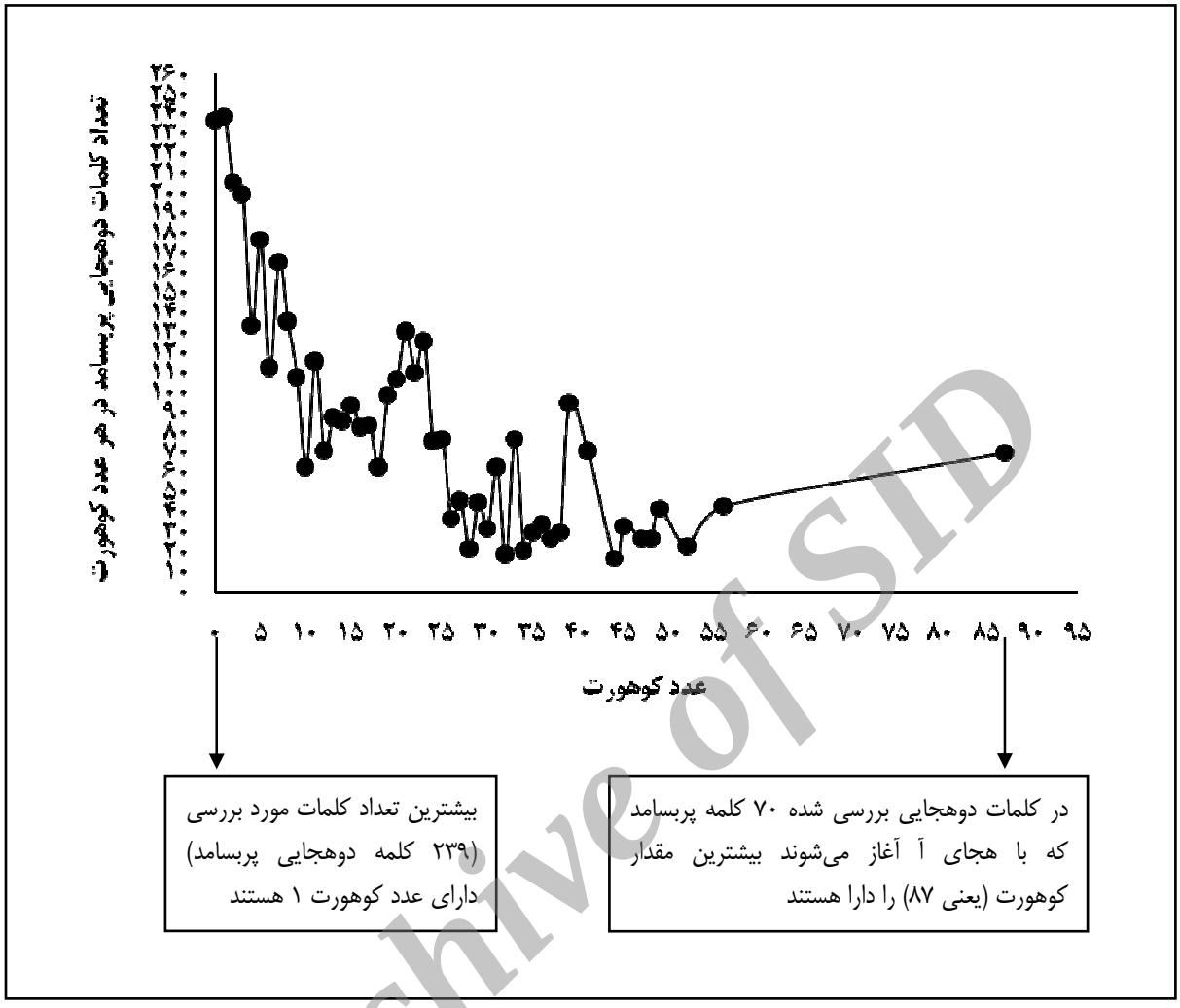
کلمات پربسامد و کم‌کوهورت			کلمات کم‌بسامد و پرکوهورت		
کلمه	بسامد	کوهورت	کلمه	بسامد	کوهورت
عنوان	۴۷۱۷۱	۱	آژیر	۸۴	۸۷
بسیار	۲۹۱۸۷	۰	ناقص	۱۳۲	۵۶
فوتبال	۱۶۱۵۳	۱	عیان	۱۴۱	۴۹
دختر	۸۴۴۰	۰	واله	۳۳	۴۷
فعال	۷۶۴۸	۰	پرپشت	۹۴	۳۱

عبارتی کلمات غیر قابل پیش‌بینی‌تر محدودتر است. بنابراین برای انتخاب کلماتی که قابلیت حدس آنها کمتر باشد باید دقت نظر بیشتری به خرج داد.

در کلمات مورد بررسی بیشترین مقدار کوهورت ۸۷ بود که متعلق به هجای اول «آ» است. از این یافته می‌توان چنین استنباط کرد که کلمات دارای هجای اول «آ» به‌دلیل رقابت فراوان موجود به سختی قابل حدس زدن هستند. مفهوم عدد کوهورت ۸۷ این است که در فرهنگ فارسی عمید ۸۷ کلمهٔ دوهجایی وجود دارد که با هجای «آ» شروع می‌شوند و به نوعی رقیب هم محسوب می‌شوند. از طرف دیگر در کلمات دوهجایی پربسامد بررسی شده ۷۰ کلمه دارای عدد کوهورت ۸۷ هستند که با هجای «آ» شروع می‌شوند. می‌توان نتیجه گرفت که از بین ۸۷ کلمهٔ دوهجایی که هجای آغازین «آ» دارند ۷۰ کلمهٔ آنها پربسامد بوده و در فهرست واژگان پربسامد زبان فارسی قرار می‌گیرند. پس از هجای اول «آ»، هجاهای اول «نا، بی عَ اَ، وا، با، دا، سر و غیره» به‌ترتیب بیشترین مقادیر کوهورت را به‌خود اختصاص داده‌اند. بنابراین کلماتی که هجای اول آنها از هجاهای مذکور باشد نسبت به سایر کلمات کم‌کوهورت رقابت بیشتری ایجاد کرده و قابلیت پیشگویی کمتری خواهند داشت.

یافتهٔ قابل تأمل دیگر این است که بین بسامد کلمه و عدد کوهورت آن ارتباط مثبتی وجود ندارد. اغلب این باور در اذهان وجود دارد که کلمات رایج و پرتکرار راحت‌تر شناسایی می‌شوند ولی در این پژوهش یافت شد که بسامد کلمهٔ ارتباطی به قابلیت شناسایی کلمه ندارد و ممکن است کلمه‌ای بسیار رایج و پربسامد باشد ولی به‌دلیل وجود رقیب‌های فراوان و داشتن عدد کوهورت بالا، دیرتر و سخت‌تر شناخته شود. علت احتمالی این یافته مربوط به ساختار کلمه است. کلمات مرکب که هجای اول آنها به‌تنهایی یک کلمهٔ معنی‌دار است، صرف نظر از میزان بسامدشان، مقادیر کوهورت بالایی دارند. برای مثال، کلمهٔ «دست‌خط»، بسامد بسیار کم(۳۰) و عدد کوهورت بالا(۲۱) دارد یا مثلاً کلمهٔ «پرپشت» بسامد ۹۴ ولی عدد کوهورت ۳۱ دارد. علاوه بر این، همان‌طور که قبلاً نیز اشاره شد، کلماتی که هجای آغازین آنها «آ» است بیشترین مقدار کوهورت را دارند گرچه برخی از این کلمات بسامد بسیار کمی دارند.

محاسبهٔ کوهورت در مطالعهٔ کنونی براساس هجای اول کلمات دوهجایی بود. مطالعات فراوانی از این فرضیه حمایت می‌کنند که ابتدای کلمه (۱۵۰ میلی‌ثانیه اولیه)، به‌ویژه هجای اول، در تولید و محاسبهٔ کوهورت بسیار مهم است. شواهد رفتاری نشان داده است بخش‌های ابتدایی کلمه در شناخت کلمهٔ شنیده شده اهمیت دارند. شواهد حاصل از پتانسیل‌های دیررس شنوایی (Event Related Potentials: ERP) نیز نشان داده‌اند افراد به‌طور ترجیحی ابتدای کلمه را در مراحل درکی اولیه پردازش می‌کنند. در گفتار طبیعی، هجای ابتدایی کلمه در مقایسه با هجای میانی کلمه که از نظر آکوستیکی با آن تطبیق داده شده باشد موج n1 (اولین قلهٔ منفی) بزرگتری می‌دهد که این حالت word-onset negativity نام دارد(۱۳). بنابراین، ابتدای کلمه برای شناخت شنیداری آن مهم است. از آنجا که میزان توجه انتخابی، ابزاری برای تعیین میزان اطلاعات است، افراد باید به بخش‌های حاوی بیشترین اطلاعات در گفتار توجه بیشتر و مستقیم‌تری کنند. ابتدای کلمه نسبتاً غیرقابل‌پیش‌بینی است و بنابراین به‌شدت حاوی اطلاعات است و این فرضیه را که افراد به اجزای غیرقابل‌پیش‌بینی در گفتار توجه مستقیم می‌کنند مطرح می‌کند(۱۳). مطالعات ERP نشان می‌دهند کلماتی که از نظر



نمودار ۲ ـ تعداد کلمات دوهجایی پربسامد موجود به ازای هر مجموعه از اعداد کوهورت یکسان

بافتی منسجم (coherent) هستند ولی قابل پیش‌بینی نیستند، در مقایسه با کلمات قابل پیش‌بینی، n400 بزرگتری استخراج می‌کنند. نتایج مطالعات Astheimer و Sanders (۲۰۱۱) نشان داده است که افراد به‌طور انتخابی به ابتدای کلماتی که نمی‌توانند آنها را از روی بافت پیش‌بینی کنند توجه می‌کنند ولی در پاسخ به کلماتی که کاملاً قابل پیش‌بینی هستند (معادل با عدد کوهورت پایین یا صفر برای کلمه) افراد هیچ اثر توجهی نشان نمی‌دهند و حتی در بعضی موارد شناخت کلمهٔ شنیده شده بدون هیچ توجهی صورت می‌گیرد. یعنی برای کلماتی که رقیبی ندارند (عدد کوهورت پایین یا صفر) افراد با شنیدن ابتدای کلمه بدون هیچ توجهی می‌توانند کل کلمه را حدس بزنند. می‌توان گفت افراد به اجزای غیرقابل‌پیش‌بینی در گفتار توجه مستقیم می‌کنند(۱۳).

نتیجه‌گیری

عدد کوهورت کلمات دوهجایی پربسامد فارسی محاسبه شد. محدودهٔ وسیع عدد کوهورت در کلمات مورد بررسی نشان می‌دهد کلمات از نظر عدد کوهورت ارزش متفاوتی دارند. آگاهی از مقادیر کوهورت هنگام گزینش کلمات برای ساخت آزمون‌هایی

مانند آزمون بازشناسی گفتار، انواع آزمون‌های ارزیابی درک گفتار در نویز، ERPs و آزمون‌های ارزیابی سیستم شنوایی مرکزی باعث افزایش اعتبار و صحت این آزمون‌ها می‌شود.

سپاسگزاری

این مقاله بخشی از پایان‌نامهٔ مصوب دانشگاه علوم پزشکی تهران به شماره قرارداد ۹۱/د/۲۶۰/۱۹۷۶ مورخ ۹۱/۶/۲۰ است. از جناب آقای دکتر مدرسی، آقای دکتر عاصی و همکاران زبان‌شناس در پژوهشگاه علوم انسانی و مطالعات فرهنگی برای کمک به انجام این پژوهش سپاسگزاری می‌شود.

REFERENCES

1. Davis MH, Johnsrude IS. Hearing speech sounds: top-down influences on the interface between audition and speech perception. Hear Res. 2007;229(1-2):132-47.
2. Dumay N, Content A. Searching for syllabic coding units in speech perception. J Mem Lang. 2012;66(4):680-94.
3. Lecumberri MLG, Cooke M, Cutler A. Non-native speech perception in adverse conditions: A review. Speech Commun. 2010;52(11-12):864-86.
4. Calabrese A. Auditory representations and phonological illusions: a linguist's perspective on the neuropsychological bases of speech perception. J Neurolinguistics. 2012;25(5):355-81.
5. Harley TA. The psychology of language: from data to theory. 3rd ed. New York: Psychology Press; 2008.
6. Theunissen M, Swanepoel de W, Hanekom J. Sentence recognition in noise: variables in compilation and interpretation of tests. Int J Audiol. 2009;48(11):743-57.
7. Thibodeau LM. Speech audiometry. In: Roeser RJ, Valente M, Hosford-Dunn H, editors. Audiology Diagnosis. 2nd ed. New Yourk: Theime Medical Publishers, Inc; 2007.p. 288-311.
8. McArdle R, Hnath-Chisolm T. Speech audiometry. In: Katz J, Medwetsky L, Burkard R, Hood L, editors. Handbook of clinical audiology. 6th ed. Baltimore: Lippincot Williams & Wilkins; 2009.p.64-79.
9. Assi SM. "Farsi Linguistic Database (FLDB)". International Journal of Lexicography. 1997;10(3):5.
10. Conway CM, Bauernschmidt A, Huang SS, Pisoni DB. Implicit statistical learning in language processing: word predictability is the key. Cognition. 2010;114(3):356-71.
11. Gahl S, Yao Y, Johnson K. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. J Mem Lang. 2012;66(4):789-806
12. Kim D, Stephens JD, Pitt MA. How does context play a part in splitting words apart? Production and perception of word boundaries in casual speech. J Mem Lang. 2012;66(4):509-29.
13. Astheimer LB, Sanders LD. Predictability affects early perceptual processing of word onsets in continuous speech. Neuropsychologia. 2011;49(12):3512-6.

Audiol. 2014;23(3):30-38.

Research Article

Calculation of Cohort size for the list of Persian high-frequency spondee words

Soheila Shayanmehr[1], Jamileh Fatahi[1], Seyed Aliakbar Tahaie[2], Shohreh Jalaie[3]

[1]- Department of Audiology, School of Rehabilitation, Tehran University of Medical Sciences, Iran
[2]- Department of Audiology, Faculty of Rehabilitation Sciences, Iran University of Medical Sciences, Tehran, Iran
[3]- Biostatistics, School of Rehabilitation, Tehran University of Medical Sciences, Iran

Received: 7 May 2013, accepted: 17 September 2013

Abstract

Background and Aim: Setting of candidates for a word with similar beginnings is known as the Cohort size. Despite the importance of the number and properties of candidates in word recognition, so far, in none of the tests made for Persian language, the Cohort size is considered. The purpose of current study was the introduction of importance of Cohort size in word recognition and calculation of Cohort size for the list of Persian high-frequency spondee words.

Methods: The spondee words extracted from high-frequency Persian word store. Then, total spondee words with same first syllable in Amid Persian dictionary recorded and Cohort size calculated for each spondee word. Thus, the list of high-frequency spondee words with their Cohort size composed of 4121 words obtained.

Results: The Cohort sizes of word had a wide range from 0 to 87. In the half of the words, the Cohort sizes were less than 14 and in the rest were more than it.

Conclusion: The Cohort size affects the time course and precision of decision making about words. Persian words are not equal in Cohort size. For having more controlled test materials to develop and design different types of auditory tests, it is possible to consider the Cohort size of words along other effective factors.

Keywords: Speech perception, word recognition, Cohort model of word recognition, Cohort size, Persian language

Please cite this paper as: Shayanmehr S, Fatahi J, Tahaie SA, Jalaie S. Calculation of Cohort size for the list of Persian high-frequency spondee words. Audiol. 2014;23(3):30-8. Persian.

Corresponding author: Department of Audiology, School of Rehabilitation, Tehran University of Medical Sciences, Pich-e-Shemiran, Enghelab Ave., Tehran, 1148965141, Iran. Tel: 009821-77530636, E-mail: jfatahi@tums.ac.ir