

## حشو در زبان با رویکرد نظریه اطلاعات

اردوان میرزایی\*

### مقدمه

این مقاله به بررسی مختصر یکی از مشکلات زبان با استفاده از ریاضی پرداخته و کاربرد ریاضی (نظریه اطلاعات) در زبان را نشان می‌دهد. اهمیت ریاضی و کاربرد آن در علوم مختلف بر کسی پوشیده نیست. دانشمندان به این نتیجه رسیده‌اند که هرگاه علمی از ریاضی به صورت بهتر و کارا تر استفاده کند پیشرفتش اصولیتر، مستحکمتر و سریعتر خواهد بود. اصولاً، چنانچه بتوان مقادیر کیفی را به کمی تبدیل کرد، در این صورت، علم ریاضی، که عموماً بر مقادیر کمی حکمفرمایی می‌کند، به کار می‌آید و با نظریه‌های مستحکم خود نتایج درخور توجه، گاه غیرعادی، به دست می‌دهد. کاربرد ریاضی نه تنها فی‌نفسه برای خود ریاضی مفید و قابل توجه است، بلکه برای علوم دیگر نیز راهگشا خواهد بود. از این رو،

چکیده: این مقاله ابتدا به نمونه مکالمه دو نفر پرداخته و برخی از مشکلاتی را که از این گفتگو، بین آن دو ظهور کرده بازگو می‌کند. این مشکلات به دو بخش شکلی و ماهوی تقسیم می‌شود. بخش مشکلات شکلی (ظاهری) به دو قسمت فیزیکی بدن و قسمت زبان و ساختار آن می‌پردازد. و در اینجاست که مشکل انتقال اطلاعات در زبان مطرح می‌شود. به لحاظ ارتباط تنگاتنگی که بین نظریه اطلاعات (در ریاضی) و نظریه بهینگی (در زبان) دیده می‌شود زبان به مجموعه‌ای از محدودیتها تعریف شده و نقش محدودیتها در زبان به اختصار آمده است. سپس به ارائه مختصر تئوری اطلاعات<sup>۱</sup> اشاره شده و کانال انتقال اطلاعات از مبدأ به مقصد بیان شده است. در این مقاله دو مفهوم آنتروپی<sup>۲</sup> و حشو<sup>۳</sup> مطرح می‌شود. محاسبه آنتروپی و حشو در زبان انگلیسی و مثالی از یک کلمه انگلیسی آورده شده است. پس از آن به محاسبه آنتروپی و حشو در زبان فارسی پرداخته می‌شود و این نتیجه به دست می‌آید که آنتروپی زبان فارسی روی حرف اول کلمه ۵ بیت/نماد و زوائد آن حدود ۸۰ درصد است.

\*عضو هیئت علمی دانشگاه پیام‌نور، سازمان مرکزی.

1. Information Theory

2. Antropy

3. Redundancy

کلیدواژه: تئوری اطلاعات، نظریه بهینگی، آنتروپی، حشو، ساختار زبان، دستور زبان، بیت (Bit).

قسمت شکلی (یا ظاهری) است. در قسمت ماهوی چگونگی تنظیم مطالب در ذهن گوینده (شنونده) و موضوعاتی از این دست مطرح است. در این مقاله به بخش اول (قسمت ماهوی) وارد نمی‌شویم و به قسمتی از بخش دوم (قسمت شکلی یا ظاهری) می‌پردازیم. بخش ظاهری، یا بهتر بگوییم شکلی، خود به دو قسمت می‌شود. قسمت اول به موارد فیزیکی، مانند مکان مکالمه، زمان مکالمه، عمل نمودن صحیح یا غلط دستگاه عصبی طرفین و اعضای مانند دهان و گوش، که این انتقال را انجام می‌دهند، مربوط می‌شود. در این مقاله به قسمت دوم، که موارد غیر فیزیکی یعنی زبان و ساختار زبان است، توجه می‌کنیم.

#### زبان به منزله مجموعه‌ای از قواعد<sup>۵</sup>

در انقلاب دوم در تاریخ زبان‌شناسی، که همان پیدایش زبان‌شناسی نوین، یعنی زبان‌شناسی سوسوری و ساخت‌گرایی اروپایی است، زبان (language) به عنوان نظامی از نشانه‌ها (signs) تعریف شد. به بیانی دیگر، در این رویکرد بین نشانه‌ها شبکه‌ای از روابط متصور بود و ارزش هر نشانه در این شبکه تعیین می‌شد. اگرچه واج (phoneme) به منزله آوای ممیز را ابتدا بودوئن دوکورتی،<sup>۶</sup> زبان‌شناس لهستانی تبار، مقیم روسیه تزاری، مطرح کرد؛<sup>۷</sup> تروبتسکوی با بهره‌گیری از مفاهیم ارزش مطرح‌شده از سوی سوسور اقدام به معرفی انواع تقابلها بین واجها کرد و بدین ترتیب واج‌شناسی (phonology) به مثابه شاخه‌ای از زبان‌شناسی در مکتب پراگ پا به عرصه وجود گذاشت. در واج‌شناسی مکتب پراگ تعیین مشخصه‌های آوایی (phonetic features) تشکیل‌دهنده واجها توسط تروبتسکوی و یاکوبسن<sup>۷</sup> پیگیری شد و در همین مقطع زمانی تعریف

۴. نک: دبیرمقدم، ۱۳۸۳.

5. Baudouin de Courtenay

6. Terobestscoy

7. Yacobscon

لازم است مختصری از نظریه اطلاعات گفته شود. ابتدا دو نفر را در نظر بگیرید که می‌خواهند با هم گفتگو کنند. برای انتقال مطالب از یک ذهن به ذهن دیگر ناچار به استفاده از زبان (گفتاری - نوشتاری) هستیم. اطلاعاتی در ذهن وجود دارد، ابتدا این اطلاعات به صورتی خاص (ذاتی و اکتسابی) در ذهن (گوینده) مرتب می‌شود، سپس با ساختار زبان به آن شکل داده می‌شود و با دستور مغز به سیستم تکلم ادا می‌شود (با فرض اینکه عیناً همین صدا به شنونده منتقل می‌شود و مشکلی در تکلم گوینده و شنوایی شنونده وجود نداشته باشد بحث را ادامه می‌دهیم). صدا به شنونده می‌رسد و او با گوش و اعصاب که به مغز منتقل می‌شود، می‌شنود. اکنون شنونده مطالبی را در ذهن خود دارد. آیا این مطالب دقیقاً منظور گوینده است؟ آیا گوینده توانسته مطلب خود را طوری تنظیم و ادا نماید که شنونده دقیقاً همان منظور را دریابد؟ آیا گوش شنونده همان چیزی را که گوینده گفته به ذهن انتقال داده؟ آیا ذهن شنونده دقیقاً همان منظور ذهن گوینده را درک کرده است؟ آیا دقیقاً همان اطلاعاتی که در ذهن گوینده بوده، همان اطلاعات را ذهن شنونده در خود دارد؟ اکنون فرض کنید شنونده مطالب را درک کرده و در صدد پاسخگویی است این بحث در واقع به مباحث زیادی کشیده می‌شود. در این انتقال اطلاعات ذهن گوینده در تنظیم مطالب چقدر نقش دارد؟ نقش گفتار گوینده چقدر است؟ نقش ارتباط (فاصله - نوع ارتباط، یعنی تلفن، حضوری و...) چقدر است؟ گوش شنونده چقدر نقش دارد؟ گیرایی مطلب توسط ذهن شنونده چه میزان نقش دارد؟ شبکه اعصاب انتقال اطلاعات از ذهن گوینده به زبان او و شبکه اعصاب انتقال اطلاعات از گوش شنونده به ذهن او چقدر است؟ این انتقال اطلاعات از مبدأ گوینده به مقصد شنونده را به دو بخش می‌کنیم. بخش اول قسمت ماهوی و بخش دوم

گویش در زبانهای مختلف مربوط به نوع و چگونگی محدودیتهای حاکم بر آن است. در نظریه بهینگی (optimality theory) محدودیتهای دسته‌بندی و اولویت‌بندی شده‌اند و هر کدام اولویت داشته باشد در زبان نقش اساسیتری بازی می‌کند. اولویتها در زبانهای مختلف متفاوت است؛ بنابراین ساختارهای زبانها تا حدودی متفاوت خواهد بود، ولی اصول کلی حاکم بر آنها همچنان باقی است. به این ترتیب، نقش‌گوینده به محدودیتهایی که بر ساختار زبان حاکم است محدود می‌گردد. در ادامه، بحث را بر اساس نظریه اطلاعات، دنبال می‌کنیم و به بررسی اختیار گوینده در زبان می‌پردازیم.

#### رویکرد نظریه اطلاعات

نظریه اطلاعات به کمیّت، کدگذاری (نصب علائم) و انتقال اطلاعات از مبدأ به مقصد مربوط می‌شود. قسمتی از ایده اصلی این نظریه به توسط نی کوئیست<sup>۹</sup> (1924) (Beerbower & Jordan, 1960: 43/1184-1198) و هارتلی<sup>۱۰</sup> (۱۹۲۸) (Elliott, 1970: 3/319-335) فرمول‌بندی شد و بعد از جنگ جهانی دوم این نظریه پیشرفت قابل توجهی کرد. کار اصلی را شانن<sup>۱۱</sup> انجام داد (1949). برخی از کارهای شانن، به سرعت، علمی شد و کاربرد پیدا کرد. با توسعه تکنولوژی در سالهای اخیر، شاهد تغییراتی هستیم. این تغییرات سبب ارتقای سطح تئوری اطلاعات شد و این نظریه نیز کمک بسزایی در بخشی از توسعه تکنولوژی ایفا کرده است؛ تأثیرات این رشد را بر جامعه جهانی کنونی شاهد هستیم و در آینده شاهد پیشرفتهای بیشتری خواهیم بود.

اصول بنیادی این تئوری، تا حد زیادی، کار

واج و روالهای کشف واج در کانون توجه زبان‌شناسان ساخت‌گرای آمریکا قرار داشت (Twaddell, 1966: 55-80) واج‌شناسی زایشی معیار (Chomsky & Hall, 1968) بازگشتی بود به دستور پانینی (قرن ۴ ق. م) و مشخصاً، کشف قاعده‌های واجی زبانها و نحوه صورت‌بندی و فرموله نمودن آنها مورد توجه ویژه واج‌شناسان زایشی قرار گرفت.

#### زبان به منزله مجموعه‌ای از محدودیتهای<sup>۱۲</sup>

جان رابرت راس (1986) در رساله دکترای خود، که در ۱۹۸۶ به چاپ رسید، بحثی در باب محدودیتهای (constraints) را بر عملکرد گشتارها مطرح می‌کند. این محدودیتهای که گاهی شرط (condition) و گاهی اصل (principle) نیز خوانده شده است، همان هماهنگی زبانی‌اند و از این رو، متعلق به دستور همگانی (universal grammar) می‌باشند. کودکان این محدودیتهای و شرطها و اصلها را نمی‌آموزند؛ بلکه آنها بخشی از دانش زیستی و فطری آنان را تشکیل می‌دهند. تعدادی از این محدودیتهای در اثر (Ross, 1986) و چامسکی (1973) دیده می‌شوند (دبیرمقدم، ۱۳۸۳).

این محدودیتهای عملکرد گشتارها و مشخصاً گشتارهای حرکتی، مانند مجهول‌سازی، مبتداسازی، حرکت پرسشواژه و حرکت بند، موصولی‌سازی و ارتقاء را محدود می‌کنند. در نظریه حاکمیت و مرجع‌گزینی (Chomsky, 1981) گام مهمی به سوی کمینه‌گرایی (minimalism) برداشته شد و آن این بود که تمام گشتارهای حرکتی موارد خاصی از یک فرایند کلی به نام گشتار حرکت آلفا تلقی شدند. تحول دیگر این بود که در بخش نحو دیگر سخنی از گشتارهای اختیاری و اجباری در میان نبود، بلکه همه گشتارهای این بخش اجباری بودند.

در این نظریه زبان توسط محدودیتهای کنترل می‌شود و تفاوت زبانها در محدودیتهایی است که بر ساختار زبان تأثیر می‌گذارند. از این جهت، چگونگی

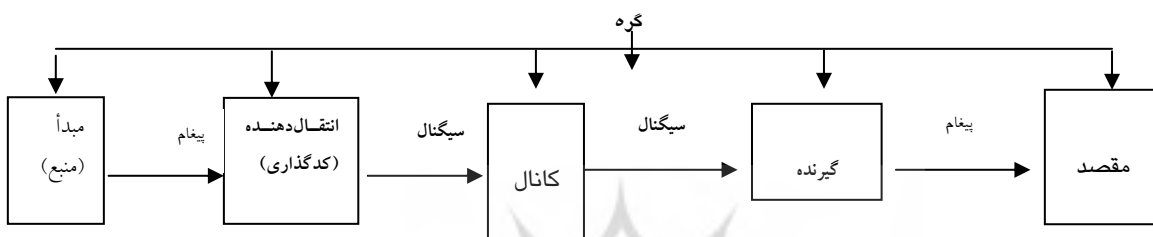
۸. نک: دبیرمقدم، ۱۳۸۳.

10. Hartley

9. Nyquist

11. Shannon

کار اصلی شانن طراحی سیستمهای مخابراتی بود که به کمک آنها بتوان اطلاعاتی معین را از نقطه‌ای به نقطه‌ای دیگر انتقال داد و در آنجا، به طور دقیق، بازسازی کرد. به کمک ریاضی می‌توان اطلاعات را، چه در مبدأ و چه در مقصد، تعریف کمی کرد و از طریق کانال، اطلاعات را منتقل کرد. نمونه‌ای کلی از یک سیستم ارتباطی در شکل ۱ نشان داده شده است.



شکل ۱. نمای کلی یک سیستم ارتباطی

$n$  تا از این وسایل می‌توانند  $n$  بیت اطلاعات ذخیره کنند، عده کل حالات ممکن  $2^n$  است. پس:

$$\text{بیت } n = -\log_2 \frac{1}{2^n}$$

برای مثال تعداد حروف الفبای انگلیسی ۲۶ است بنابراین:

$$\text{بیت } 4/7 = -\log_2 \frac{1}{26}$$

برای درک نظریه باید خانه‌هایی را تصور کنیم که هر کدام می‌توانند دو حالت به خود بگیرند. در باب حروف الفبای فارسی شکل ۲ را ببینید:



شکل ۲. در ۳۲ خانه ۵ بیت اطلاعات ذخیره می‌شود

اکنون دو نکته در خور ذکر است: یکی اینکه، می‌توان به جای مبنای ۲، که شانن آن را در نظر گرفت،

شانن و همکارانش در آزمایشگاههای تلفن بل در امریکا است. شانن در ۱۹۴۸ نظریه ریاضی مخابرات را منتشر کرد و به دنبال آن، کاربردهای زیادی در زمینه‌های کامپیوتری، به خصوص ICT<sup>۱۲</sup> و مخابرات، زمین‌شناسی، تلویزیون و رادیو، الکترونیک پیدا کرد (Shannon, 1984: 27/37-423; Guisau, 1977; Pelto, 1954: 62/501-511; Tasch, 1980; Beerbower..., 1960: 43/1184-1198; (Ferguson, 1980: 137/107-108; Idem, 1982; Idem, 1983

مغز گوینده که مبدأ یا منبع اصلی تولید اطلاعات است، اطلاعات (پیغام) را از طریق نرون عصبی به مجموعه انتقال‌دهنده، یعنی دهان، زبان و ...، به صورت کدگذاری شده، یعنی زبان فارسی (یا انگلیسی یا ...) با رعایت قواعد حاکم بر آن ارسال کرده و دستور انتقال را می‌دهد. زمانی که شروع به صحبت می‌کند، این پیغام به شکل سیگنالهای صوتی خارج شده و به مجموعه گیرنده، یعنی گوش شنونده می‌رسد. گوش شنونده، که با زبان فارسی (انگلیسی یا ...) و قواعد آن از طریق مغز آشناست، پیغام را به توسط سلسله اعصاب به مغز شنونده انتقال می‌دهد. مغز شنونده اطلاعات را دریافت می‌کند و به پردازش و بررسی آن می‌پردازد.

معیاری که شانن (1984:27/380) برای اطلاعات پیشنهاد کرد، بیت یا رقم در مبنای ۲ است. بنابراین، اگر وسیله‌ای مثل کلید داشته باشیم که دو حالت پایه‌ای روشن و خاموش دارد، داریم:

$$\text{بیت } 1 = -\log_2 \frac{1}{2}$$

آنترپپی در ترمودینامیک است. در ترمودینامیک، آنترپپی به مقدار اضمحلال سیستم گفته می‌شود. یکی از مفاهیم اساسی و مهم نظریه اطلاعات زوائد است، که به صورت زیر تعریف می‌شود ( Usher, 1984):

آنترپپی واقعی - حداکثر آنترپپی

$$Re = \% \frac{\text{حداکثر آنترپپی}}{\text{آنترپپی واقعی}}$$

حداکثر آنترپپی

اکنون سیستمی را در نظر می‌گیریم که دو حالت ممکن دارد و احتمال وقوع هر دو نیز برابر است؛ یعنی:

$$p_1 = p_2 = 0/5$$

پس داریم:

$$H = (-0/5 \log_2 0/5) + (-0/5 \log_2 0/5) = 1$$

اگر  $p_1 = 1, p_2 = 0$  (یا بر عکس) آنگاه  $H = 0$  و

اگر  $p_1 = 0/25, p_2 = 0/75$  (یا بر عکس) آنگاه

$$H = 0/8$$

بنابراین، آنترپپی سیستم وقتی ماکزیمم است که احتمال وقوع تمامی رخدادها با هم برابر باشد؛ و وقتی مینیمم است که احتمال وقوع یک رخداد برابر ۱ و احتمال وقوع سایر رخدادها صفر باشد.

### محاسبه آنترپپی زبان انگلیسی

زبان انگلیسی دارای ۲۶ حرف است و کل این حروف کلمات را تشکیل می‌دهند. ساختار زبان و دستور زبان انگلیسی کلمات را کنار هم قرار داده و جملات، که منظور نظر است، ساخته می‌شود. از آنجایی که ۲۶ حرف دارای احتمالات مساوی است، پس هر کدام به احتمال  $\frac{1}{26}$  ممکن است ظاهر شود. مقدار آنترپپی (H) برای حروف انگلیسی به صورت

از مبنای دیگری (مثلاً ۳ یا e یا ۱۰) استفاده کرد؛ دیگر اینکه، در هر خانه بیش از یک حرف نمی‌توان قرار داد و چون n حالت وجود دارد پس احتمال قرار گرفتن یکی از این حروف در خانه اول  $\frac{1}{n}$  است. برای حروف الفبای فارسی برای خانه اول  $\frac{1}{32}$  احتمال این وجود دارد که یکی از حروف در آن خانه قرار گیرد. رابطه‌ای نزدیک بین اطلاعات و عدم یقین وجود دارد. هر چه پاسخ نامعینتر باشد، با دادن پاسخ، اطلاعات بیشتری به دست می‌آید. اگر چند حالت ممکن وجود داشته باشد، مقدار اطلاعات زمانی ماکزیمم می‌شود که احتمال وقوع همه حالتها یکسان باشد. مجموعه‌ای از n اتفاق را با احتمال هر کدام  $p_1, p_2, \dots, p_n$  در نظر بگیرید. احتمال اینکه اتفاق اول روی دهد  $p_1$  است که می‌تواند در خانه اول جای گیرد چون n خانه وجود دارد و ممکن است اتفاق اولی در قسمتهای خانه ۴ تایی جای گیرد. پس  $n \times p_1$  زمان خانه‌ها توسط اتفاق اول اشغال می‌شوند و به این ترتیب  $n \times -\log_2 p_1$  بیت اطلاعات ذخیره می‌شود.

آنترپپی سیستم را با H نشان داده و به صورت زیر تعریف می‌شود:

$$H = -K \sum_{i=1}^n P_i \log_2 p_i \quad \text{بیت}$$

در رابطه فوق K به انتخاب واحد سنجش

اطلاعات بستگی دارد (Shannon, ..., 1949: 393; Guiasu, 1977: 1ff).

در اینجا  $K=1$  انتخاب می‌شود. پس داریم:

$$H = -\sum_{i=1}^n P_i \log_2 p_i \quad \text{بیت}$$

این معادله اساس نظریه اطلاعات است و

اطلاعات را به انتخاب و عدم یقین مربوط می‌کند.

کلمه آنترپپی در نظریه اطلاعات تقریباً همان مفهوم

زیر محاسبه می شود:

$$H = -\sum_{i=1}^{26} p_i \log_2 p_i$$

$$H = -(p(A)\log_2 p(A) + p(B)\log_2 p(B) + \dots + P(Z)\log_2 p(Z))$$

$$= -\left(\frac{1}{26} \log_2 \frac{1}{26} + \frac{1}{26} \log_2 \frac{1}{26} + \dots + \frac{1}{26} \log_2 \frac{1}{26}\right) = -(26 \left(\frac{1}{26} \log_2 \frac{1}{26}\right)) \approx 4/1$$

۲۶ مرتبه

است و بقیه پیغامها تکرار اولی بوده و زائد هستند. برای مثال، در یک سیستم دو دویی (مثلاً A=0 و B=1)، وقتی پیغام A را ۳ بار ارسال کنیم، به صورت کد گذارنده OOO، یا پیغام B را سه بار ارسال کنیم، که به صورت کدگذاری شده به شکل 111 می باشد، و اولین کد رسیده یعنی O (برای A) یا 1 (برای B) با معنی است و بقیه کدها تکرار پیغام قبلی و زائد است. در این صورت، زوائد به شکل زیر است:

واحد آنتروپی (H) اغلب با بیت / حرف یا بیت / سمبل نشان داده می شوند. به این ترتیب، آنتروپی زبان انگلیسی حدود ۱/۴ بیت / نماد است.

### زوائد در زبان انگلیسی

در نظریه اطلاعات، مفهوم حشو به تکرار یک پیغام مربوط است. اگر N پیغام مختلف وجود داشته باشد و تنها یکی را، مثلاً ۳ بار، ارسال کنیم، با فرض اینکه کانال انتقال بدون گره<sup>۱۳</sup> باشد، تنها اولین پیغام با معنی

$$Re = \% \frac{3-1}{3} = \% \frac{2}{3} = \% 66/6$$

آنتروپی، واقع، - ماکزیمم آنتروپی، ماکزیمم آنتروپی

محدودیت مربوط به سحر زبان محیسی است.

حدود ۷ حرف وجود دارد که می تواند بعد از D واقع شود که حدود ۲/۸ بیت خواهد بود. یعنی برای یک کلمه ۲ حرفی ۲/۸ بیت / نماد خواهیم داشت. فرض کنیم حرف دوم I است. انتخاب زیاد دیگری برای حرف سوم وجود دارد. اگر حروف بعدی به ترتیب R, T, S باشند بنابراین به کلمه DISTR رسیده ایم، ولی باز هم برای انتخاب حرف ششم حالتی وجود دارد. اگر حروف بعدی U, B, I باشند، بنابراین می توانیم کلمه را DISTRIBUTION یا DISTRIBUTIVE حدس بزنیم. ملاحظه می کنیم که آخرین اطلاعات کسب شده بخش کوچکی از اطلاعات را تشکیل می دهند؛ زیرا دنباله ای که از

از آنجایی که سه بار پیغام ارسال شده، پس سه بیت اشغال شده که ماکزیمم آنتروپی است و آنتروپی واقعی برابر ۱ است زیرا یک کد (۰ یا ۱) تنها یک بیت اشغال می کند.

کلمات بخش وسیعی از احتمالات را تشکیل می دهند. به این ترتیب که یک کلمه ممکن است دامنه وسیعی از لغات را تشکیل دهد. یک مکالمه تلفنی را در نظر بگیرید، برای گفتن یک کلمه در زبان انگلیسی ۲۶ انتخاب می توان برای حرف اول کلمه در نظر گرفت، که حدود ۱/۴ بیت / نماد اشغال می کند. فرض کنید اولین حرف D باشد. انتخاب حرف بعدی محدودتر است، زیرا در انگلیسی کلمه ای که مثلاً با DD شروع شود، وجود ندارد. این

مشخص می‌کند؛ مثلاً، زوائد زبان انگلیسی حدود ۷۸ درصد است (Useher, 1984: 14). پس ۷۸ درصد علائم انگلیسی را ساختار زبان تعیین می‌کند و فقط ۲۲ درصد اختیاری است. از دیگر ویژگیهای زوائد می‌توان به ساختن جدولهای بزرگ اشاره کرد. با زبانی که زوائد آن ۳۳ درصد باشد، از لحاظ تئوری، می‌توان جدول کلمات متقاطع سه بعدی ساخت (Guiasu, 1977:130; Shannon, 1984:399). قابل ذکر است که دیدگاه نظریهٔ بهینگی، نیز ما را به نقطه‌ای مشابه در نظریهٔ اطلاعات می‌رساند.

### محاسبهٔ آنتروپی و حشو در زبان فارسی

اکنون قسمتی از تئوری اطلاعات را برای زبان فارسی پیاده می‌کنیم. از آنجا که زبان فارسی ۳۲ حرف دارد، پس برای اینکه کلمه‌ای را ادا کنیم ابتدا یک حرف از ۳۲ حرف را باید انتخاب نماییم که احتمال چنین انتخابی (با توجه به متساوی بودن همهٔ احتمالها برای تمام حروف)  $\frac{1}{32}$  است.

مقدار آنتروپی سیستم زبان فارسی در زیر آمده است:

$$H = -\sum_{i=1}^n P_i \log p_i$$

$$H = -(p(\bar{)} \log p(\bar{)}) + p(\bar{)} \log p(\bar{)} + \dots + p(\bar{)} \log p(\bar{)}$$

$$H = -\left(\frac{1}{32} \log_2 \frac{1}{32} + \frac{1}{32} \log_2 \frac{1}{32} + \dots + \frac{1}{32} \log_2 \frac{1}{32}\right)$$

← ۳۲ مرتبه →

$H = 5$  بیت / سمبل

مقایسه، درمی‌یابیم که حشو زبان فارسی در روی حرف اول حدود ۲ درصد بیشتر از وضعیت مشابه در زبان انگلیسی است. از آنجا که زوائد زبان فارسی ۸۰ درصد است، پس ۸۰ درصد علائم زبان فارسی را ساختار زبان تعیین می‌کند و فقط ۲۰ درصد اختیاری است. تا اینجا تنها منابع مجزای اطلاعاتی بررسی شد و فقط حرف اول کلمه مورد بحث قرار گرفت. چنانچه بحث بر سر حرف دوم کلمه را آغاز کنیم، صحبت از وابسته یا مستقل بودن حروف اول و دوم

### 13. Node

حروف DISTRIBU تشکیل می‌شود رویکرد مناسبی برای حدس زدن کلمه پیش روی ما قرار می‌دهد. این دنباله تقریباً همان چیزی است که در دستور زبان به آن بن کلمه گفته می‌شود، البته این دنباله گاهی از نظر تعداد حروف کمتر از تعداد حروف بن کلمه می‌باشد. شانن و ویور<sup>۱۴</sup> (1949) مطالعات و بررسیهای زیادی در مورد زوائد زبان انگلیسی انجام دادند. ولی محاسبهٔ زوائد یک متن که دارای تعدادی جمله است کار بسیار مشکلی است. زیرا، همانطور که می‌دانید، هر متن از چندین جمله و هر جمله از چندین کلمه و هر کلمه از چندین حرف تشکیل شده است و با توجه به اینکه همهٔ متنها یکی نیست و در واقع هر متنی با متن دیگر تفاوت دارد، پس محاسبهٔ زوائد یک متن با متن دیگر متفاوت خواهد بود.

یکی از بهترین کاربردهای حشو در زبان آن است که حدود اختیار گوینده و ساختار زبان را

پس آنتروپی زبان فارسی ۵ بیت / نماد است که در مقایسه با آنتروپی زبان انگلیسی که برابر ۱/۴ بیت / نماد است به تعداد حدود ۹/۰ بیشتر است. این اختلاف ناشی از تعداد بیشتر حروف در زبان فارسی است. برای محاسبهٔ زوائد زبان فارسی فقط روی حرف اول داریم:

$$Re = \% \frac{5-1}{5} = \% 80$$

بنابراین، زوائد زبان فارسی حدود ۸۰ درصد است. زوائد زبان انگلیسی حدود ۷۸ درصد است که، با

زایشی، سمت، تهران؛

فارسیان، محمدرضا (۱۳۷۸)، جنسیت و واژگان، پایان نامه

کارشناسی ارشد، دانشگاه تهران؛

**Archangeli, D.** (1997), *Optimality Theory: An introduction to linguistics in the 1990s*, Basil Blackwell;

**Atkinson, K.** (1985), *Elementary numerical analysis*, New York: Wiley;

**Beerbower, J. R. & Jordan, D.** (1960), "Application of Information Theory to Palaeontological Problems", *Journal of palaeontology*;

**Bromberger, S. & Halle, M.** (1991), *Why phonology is Different*, Basil Blackwell;

**Chomsky, N.** (1973), *Condition on transformation*, New York: Holt, Rinehart & Winston;

\_\_\_\_\_ (1981), *Lectures on Government and Binding*, Mass: MIT press, Cambridge;

**Chomsky, N. & Hall, M.** (1968), *The Sound Pattern of English*, Harper and Row, New York;

**Elliott, R. E.** (1970), "Simulation of a Productive Coal Measures Sequence", *the A mercian Geologist*;

**Ferguson, J.** (1980), "Application of Information Theory: Dataprocessing", *Quarterly Journal of the Geological society of London*;

\_\_\_\_\_ (1982), "The Application of Information Theory to Trend Surface Analysis", *Mathematical Journal*;

**Idem**, (1983);

**Glass, J. C.** (1980), *An introduction to mathematical methods in economics*, New York: Mc Graw-Hill;

**Gray, J. R.** (1967), *Probability*, Oliver and Boyd Edinburgh;

**Guiasu, S.** (1977), *Information Theory with Application*, New York: Mc Graw - Hill;

**Harbaugh, J. W. & Merriam, D. F.** (1968), *Computer Applications in Stratigraphic Analysis*, New York: Wiley;

**Kager, R.** (1999), *Optimality Theory*, Cambridge press;

**Pelto, C. R.** (1954), "Mapping of Multicomponent System", *Journal of Geology*;

**Ross, J.** (1967), Constraints on variables in syntax. oh.D. Dissertation MIT;

\_\_\_\_\_ (1986), *Infinite syntax*. Norwood, N.J., Ablex

به میان می‌آید. در این حالت، به احتمالات مشترک

14. W. Weaver

اشاره خواهد شد. ابتدا بحث وقوع یا عدم وقوع اتفاقات متوالی و سپس، احتمالات مربوط را خواهیم داشت. این بحث سنگین و طولی است و در این مقال جای نمی‌گیرد. همین مقدار بدانیم که این بحث را نباید خاتمه یافته تلقی کرد.

### نتیجه

این مقاله کاربرد مختصری از ریاضی را در بخش کوچکی از زبان نشان داد. تعریف و به کار بردن مفهوم جدید از نظریه اطلاعات، که مربوط به بخش انتقال اطلاعات از فرستنده به گیرنده است و با زبان ریاضی بیان می‌شود را در قالب زبان فارسی روی حرف اول یک کلمه قرار دادیم و نتیجه آن شد که آنتروپی سیستم زبان فارسی روی حرف اول کلمه ۵ بیت / نماد است و زوائد آن حدود ۸۰ درصد می‌باشد؛ بنابراین، اختیار ما برای سخن گفتن ۲۰ درصد و احاطه زبان فارسی بر گفتار ما حدود ۸۰ درصد است. در زبان انگلیسی آنتروپی سیستم حدود ۴/۱ بیت / نماد است و زوائد آن روی حرف اول کلمه ۷۸ درصد است. این مقاله می‌تواند راهگشای شاخه جدیدی از تحقیقات در زبان فارسی باشد که پایه آن ریاضی است. بستر سازی مناسب ریاضی برای زبان می‌تواند به درک، تسهیل در یادگیری و شناخت بیشتر زبان منجر شود. از این رو، توجه به این مسئله در خور اهمیت است.

### منابع

جان نژاد، محسن (۱۳۸۰)، زبان و جنسیت، تفاوت‌های زبانی میان گویشوران مرد و زن ایرانی در تعامل مکالمه‌ای، رساله دکتری زبان‌شناسی دانشگاه تهران؛

دیر مقدم، محمد (۱۳۸۳)، «نظریه بهینگی»، مجله زبان و ادب، دانشگاه علامه طباطبائی، شماره ۲۰؛

\_\_\_\_\_ (۱۳۸۳)، زبان‌شناسی نظری، پیدایش و تکوین دستور



2ndedn, New York: Wiley;

**Twaddell, W. F.** (1935), "On defining the phoneme", *language Monograph*, N.16, reprinted in M. Joos(ed), *Readings in linguistics I*, Fourth edition (1966), the university of chicago press;

**Usher, M. J.** (1984), *Information Theory of Information Technologists*, London: Mc Millan.■

publishing corporation;

**Shannon, C. E. & Weaver, W.** (1949), *A Mathematical Theory of Communication*, univ, of Illinois press, Champaign, Illinois;

**Shannon, C. E.** (1984), "A mathematical Theory of Communication", Technical Journal of the Bell system;

**Tasch, P.** (1980), *Palaeobiology of the invertebrates*,

