

زبان‌شناسی پیکره‌ای و مسایل نظری و کاربردی ترجمه^(۱)

پروانه خسروی زاده فروشانی

نگارنده در این مقاله به معرفی پیکره‌زبانی، انواع پیکره‌های زبانی و کاربرد آنان، نقش پیکره‌های زبانی در مطالعات زبان‌شناختی قرن گذشته، ارتباط پیکره‌های زبانی با ترجمه و نظریه‌های ترجمه و استفاده از زبان‌شناسی پیکره‌ای به عنوان روشی مکمل برای نظریه دستور زایشی می‌پردازد و ضمن ارائه شواهدی از کنش‌های واقعی زبان، بر ضرورت استفاده از روش‌های احتمالاتی و مدل‌های آماری در تحلیل داده‌های زبان تأکید دارد.

زبان‌شناسی پیکره‌ای شاخه‌ای از زبان‌شناسی رایانه‌ای و مبحثی در زبان‌شناسی کاربردی است، اما امروزه اهمیت استفاده از مدل‌های آماری در بررسی‌های علمی زبان به حدی است که علم زبان‌شناسی آتی دنیا را زبان‌شناسی داده‌بنیاد می‌نامند. زبان‌شناسی پیکره‌ای به آن جنبه از ساخت‌های زبانی می‌پردازد که در آنها عدم قطعیت وجود دارد و بر اساس منطق دو ارزشی تحلیل پذیر نیستند. چنین ساخت‌هایی درگیر مسایل درک و تولید در دو شاخه یادگیری و ترجمه‌اند. مقاله حاضر به ضرورت موضوع به بررسی مسایل ترجمه در حوزه ابهام‌های واژگانی، صرفی و نحوی می‌پردازد.

با بهره‌گیری از زبان‌شناسی پیکره‌ای و روش‌های تحقیق آماری در ترجمه می‌توان فرضیه‌های مرتبط با "انتخاب از میان چند ساختار مُحتمَل موازی" را در حوزه درک محک

۱- استاد گرانقدرم جناب آقای دکتر محمود بی‌جن خان مبحث زبان‌شناسی پیکره‌ای را برای نگارنده این سطور گشودند و کلاس‌های درس ایشان پایه‌ای شد برای ارائه این نوشته. ضمن آنکه با سخاوت تمام اجازه دادند تا از پژوهش در دست چاپ ایشان به نام «نقش پیکره‌های زبانی در نوشتن دستور زبان: معرفی یک نرم‌افزار رایانه‌ای»، بهره‌گیرم. مسلماً مسئولیت هرگونه اشتباهی در ارائه متن حاضر به عهده این شاگرد کمتربین خواهد بود.

زد و از ساخت‌های مبهم به کمک روش‌های آماری ابهام‌زدایی کرد. در پردازش آماری زبان لایه‌های زبانی در هم تنیده و تفکیک ناپذیرند. بنابراین قالب سازی احتمالاتی در دو سطح درک و تولید را می‌توان در بررسی فرایندهای آوایی، صرفی، نحوی، واژگانی و گفتمانی به کار گرفت. هریک از کاربردهای فوق می‌تواند در ترجمه شفاهی و مکتوب مدرسان باشد و در تأیید یا رد فرضیه‌های رایج در مباحث نظری ترجمه راهگشا باشد. پیکره زبانی مجموعه‌ای وسیع از داده‌های زبانی است که به صورت متون نوشتاری، گفتاری یا واژگانی تهیه می‌گردد. این مجموعه، نمونه آماری داده‌های یک زبان خاص است و از دو بعد نظری و کاربردی مورد بررسی قرار می‌گیرد.

پیکره‌های زبانی می‌توانند یک زبانه، دو زبانه یا چند زبانه باشند.

گرایش زبان‌شناسان، نظریه پردازان ترجمه و به طور کلی محققان علم زبان نسبت به بهره‌گیری از پیکره‌های زبانی در طول یک قرن گذشته فراز و نشیب بسیار داشته است. نودستوریان در اواخر قرن نوزدهم از پیکره‌های چند زبانه برای کشف روابط خویشاوندی میان زبان‌ها، استخراج خانواده‌های زبانی و بازسازی صورت‌های آوایی کلمات در زبان مادر بهره می‌گرفتند.

نیمه اول قرن بیستم شاهد گرایش قابل ملاحظه زبان‌شناسان ساختگرا به توصیف زبان بر اساس داده‌های زبانی بود. ساختگرایان امریکایی به منظور مطالعه زبان سرخپوستان بومی به تحلیل پیکره‌های گفتاری پرداختند. بلومفیلد در دهه ۳۰ توصیف ساختاری خود از زبان را به کمک تجزیه و تحلیل داده‌های زبانی در سطوح واج، صرف و نحو ارائه نمود.

زلیگ هریس نیز درست در سال‌های میانی قرن بیستم دستورالعمل صوری یا روال‌های کشف را در استخراج دستور زبان پیشنهاد کرد. وی معتقد بود با استفاده از

عملکرد تعداد محدودی دستورالعمل صوری بر حجم قابل ملاحظه‌ای از داده‌های زبانی، می‌توان به کشف ساخت‌های واجی، صرفی و نحوی هر زبان موردنظری پرداخت و در نهایت دستور آن زبان را به دست داد.

با مطرح شدن آرای چامسکی در اوایل نیمه دوم قرن بیستم، جایگاه پیکره‌های زبانی در تجزیه و تحلیل دستوری ضعیف شد اما در آغاز دهه ۹۰ توجه زبان‌شناسان مجدداً به جمع‌آوری داده‌های زبانی معطوف گردید.

امروزه استفاده از پیکره در مطالعات زبانی از جایگاهی بسیار والا برخوردار است. پیکره‌ها امکان پردازش زبان طبیعی را فراهم کرده‌اند و امیدهایی برای پیشبرد ترجمه ماشینی ایجاد کرده‌اند. اکنون بررسی‌های علمی ترجمه از دو بُعد انسانی و ماشینی آن با نظریه‌های نوین زبان‌شناسی عجین است. از یک سو انگاره دستوری زبان‌شناسی زایشی به عنوان ابزاری برای تحلیل ساخت‌های زبانی در حوزه ترجمه عمل می‌کند و از سوی دیگر زبان‌شناسی پیکره‌ای و روش‌های تحقیق آماری در خدمت نظریه پردازان ترجمه قرار گرفته است.

زبان‌شناسی داده بنیاد، گامی در جهت ارائه نظریه کنش زبانی

از دیدگاه زبان‌شناسان زایشی ساخت‌های بالقوه زبان از منطقی دو ارزشی برخوردارند. جملات زبان در منطق زبان‌شناسی زایشی دارای یکی از دو ارزش صفر و یک یعنی خوش ساخت یا بد ساخت بودن هستند. در حالی که بسیاری از قضاوت‌های زبانی در میان اهل زبان دو ارزشی نیستند.

چامسکی خود نخستین بار در کتاب جنبه‌هایی از نظریه نحو (۱۹۶۵) نظریه توانش دستوری را بخشی از نظریه کنش زبانی می‌داند. از آنجا که جنبه خلاق کاربرد زبان به

کنش زبانی نسبت داده شده است، هر نظریه‌ای در مورد کنش زبانی باید بتواند علاوه بر توصیف و تبیین انگاره دستوری زبان - به عنوان زیربنایی برای درک و تولید کنش‌های زبانی - به تبیین قضاوت‌های چند ارزشی سخنگویان زبان، و نیز به بررسی معناشناختی و کاربرد شناختی زبان بر اساس داده‌های واقعی بپردازد. در مقام نظر، ارائه نظریه کنش زبانی به دو طریق امکان‌پذیر است:

نخست از طریق بازسازی مدل ذهن انسان بر اساس:

الف) دستوری که بر پایه آن، آنچه یک سخنگوی ایده‌آل بیان می‌کند به عنوان جملات خوش ساخت آن زبان تلقی گردد.

ب) نظامی کاربرد شناختی که بر اساس قواعد آن آنچه یک سخنگوی ایده‌آل در بافتی خاص بیان می‌کند مناسب و بجا تلقی شود.

ج) جزء سومی که سخنگو را قادر می‌سازد تا وقایع گفتاری جدید و بدیعی را خلق کند بی آنکه نه از قواعد دستوری و کاربرد شناختی تخطی کرده باشد.

ساخت چنین مدلی علاوه بر ارائه توانش دستوری نیازمند ارائه نظریه‌ای از توانش کاربرد شناختی، شناخت کلیه مکانیسم‌های روانی و نظام‌های عصبی، و نیز بررسی بافت حاکم بر کنش زبانی است. متأسفانه دانش کنونی بشر برای ارائه مدل ذهنی فوق در مراحل اولیه خود به سر می‌برد.

طریق دوم برای ارائه نظریه کنش زبانی، استفاده از رویکرد احتمالاتی و پردازش آماری داده‌هاست. بر این اساس زیانشناسی پیکره‌ای به بررسی زبان برون‌ی پرداخته، هدف خود را شناخت الگوهای معمول در کنش‌های زبانی بر اساس محاسبات احتمالاتی و آماری قرار می‌دهد و در جهت نیل به این هدف به پردازش ساخت‌های بالفعلی از زبان می‌پردازد که از منطق دو ارزشی پیروی نمی‌کنند و با عدم قطعیت روبرو هستند. از این

منظر زبان‌شناسی پیکره‌ای نه در جهت نفی زبان‌شناسی زایشی، که در جهت تکمیل بررسی‌های زبان‌شناختی قبلی قدم برمی‌دارد و گامی است در زمینه ارائه نظریه کنش زبانی.

امروزه بهره‌گیری از راه‌حل دوم یعنی استفاده از رویکرد احتمالاتی و پردازش آماری داده‌ها در صدر توجه زبان‌شناسان عصر حاضر قرار دارد. نگاهی به نطق افتتاحیه دومین همایش بین‌المللی پردازش زبان‌های گفتاری در کانادا بیانگر اهمیت رویکرد داده‌بنیاد در مطالعات زبان‌شناختی امروز است.

پیتر لوفوگد در سخنرانی افتتاحیه خود چنین ابراز می‌دارد:

«استفاده از پیکره‌های زبانی احتمالاً به دگرگونی کامل آواشناسی، واج‌شناسی و حتی کل زبان‌شناسی در سال‌های آتی منجر خواهد شد. از زمان ظهور نوام چامسکی زبان‌شناسی بر توصیف توانش‌گوینده، یعنی ساخت ذهنی زبان و نه تظاهر زبان در گفتار، تأکید داشته است. امروزه عاقلانه نیست که به زبان‌شناسی آمریکایی در آغاز دهه ۱۹۵۰ برگردیم... اما به همان اندازه نیز نشستن بر روی صندلی راحتی و ارائه حکم قطعی در خصوص زبان‌گفتاری که در ذهن یک گوینده فرضی قرار دارد، غیرمنطقی است.»

بنش (۱۹۹۱) نیز معتقد است اگر فن‌آوری اطلاعات در دهه‌های ۵۰ و ۶۰ در حد پیشرفت‌های اواخر قرن بیستم بود، احتمالاً روش تحقیق در مطالعات زبان به نفع مکتب زایشی تغییر جهت نمی‌داد و در نتیجه تحقیق بر روی کنش زبان بر اساس پیکره‌های زبانی متمرکز می‌شد.

زبان‌شناسی پیکره‌ای به عنوان ابزاری عملی و نظری در حوزه ترجمه

مباحث ترجمه در برگیرنده دو حوزه کاربردی و نظری در شاخه‌های عمده‌ای چون

ترجمه ماشینی و انسانی، ترجمه شفاهی و مکتوب و ترجمه از زبان مادری به زبان غیرمادری و بالعکس می‌باشد.

نوشته حاضر سعی بر آن دارد تا با توجه به داده‌های زبانی، نگاهی هرچند گذرا به مباحث مطروحه فوق داشته باشد.

بررسی قالب‌های صرفی، نحوی و آوایی در حوزه کاربردی ترجمه

۱- بخش ضمیمه فرهنگ لغت آکسفورد (۱۹۹۶) شامل یادداشت زیر است:

«امروزه تجزیه و تحلیل متون گفتاری و نوشتاری توسط کامپیوتر امکان‌پذیر است و می‌توان مجموعه کلماتی را که در با هم آیی با کلمات دیگر قرار می‌گیرند از میان متون گردآوری شده استخراج کرد. بطور مثال اگر از رایانه بخواهیم پیکره درونداد را بررسی کند و لیستی از واژه‌هایی را که غالباً با واژه **hair** بکار می‌روند به دست دهد، بروندادی از این دست خواهیم داشت:

long / black / dark / brown / cut / grey / blonde / short / red / fair / curly / dry / white / thick /...

پیکره نشان می‌دهد که واژه **hair** با طیفی از صفاتی همراه می‌شود که طول، رنگ و نوع آن را توصیف می‌کند. به لحاظ معنایی و مقوله دستوری واژه **thick** را می‌توان مترادف واژه **dense** دانست. نمونه‌های زیر نشان می‌دهند که قرار گرفتن هریک از دو صفت **thick** و یا **dense** در کنار واژه‌هایی چون **fog**، **smoke** و **forest** موجب تمایز معنایی نمی‌شود:

thick fog	dense fog
thick smoke	dense smoke
thick forest	dense forest

به اعتقاد ویلیام کرافت بسامد به کارگیری عبارت اضافی **of** نسبت به اضافه ملکی تصریفی در داده‌های زبانی بیشتر است و چنین فراوانی و قوعی به مقوله‌بندی اسم تحت عنوان «جاندار» یا «بی‌جان» بستگی ندارد. از سوی دیگر نمونه‌های شماره (۳) و (۴) مؤید آن است که استفاده از عبارات اضافی **of** تنها طریق ممکن برای بیان مالکیت گروه‌های اسمی سنگین **heavy NPs** می‌باشد.

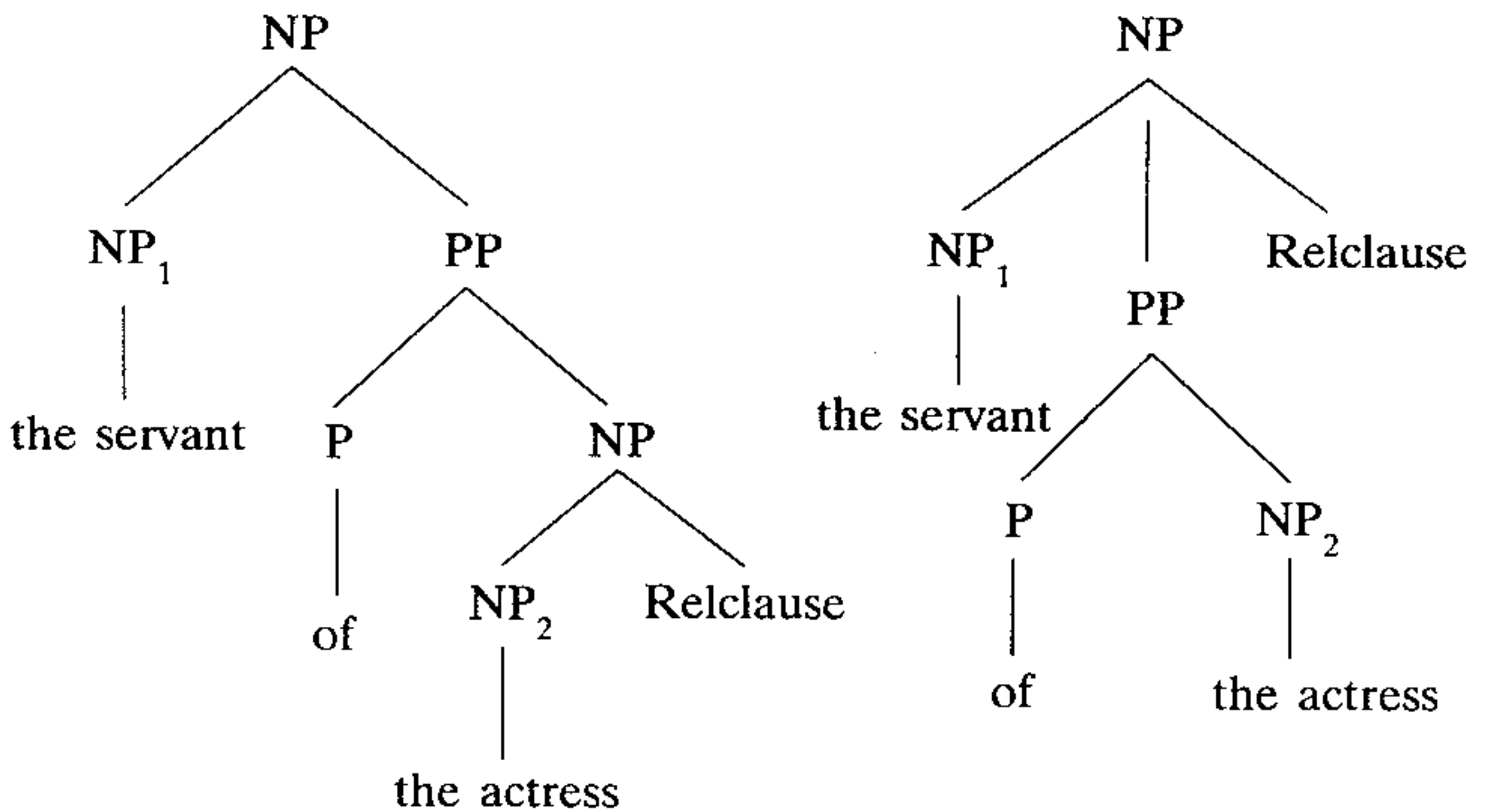
نگاهی به گروه‌های اسمی زیر در زبان انگلیسی شاهدهی دیگر بر لزوم استفاده از پیکره‌های زبانی است. کرافت بار دیگر بسامد وقوع ساخت‌هایی با عبارت اضافی **of** را فراتر از ساخت‌های "اسم + اسم" می‌داند.

- | | |
|-----------------------------|-------------------|
| 8. the library of Boston | Boston library |
| 9. the University of Tehran | Tehran University |
| 10. the back of the chair | the chair back |

چنین یافته‌هایی از بسامد وقوع ساخت‌های محتمل موازی می‌تواند از سردرگمی مترجمان فارسی زبان در ترجمه متون فارسی به انگلیسی بکاهد.

۲- دستور زایشی در تحلیل نحوی جملاتی نظیر (۱۱) به دو زیر ساخت متفاوت اشاره دارد.

11. Someone shot [_{NP₁} the servant] of [_{NP₂} the actress] who was on the balcony. (Cuetos et al).



تحلیل چنین ساختی از نقطه نظر توانش دستوری جدا از کنش واقعی زبان است. مترجمی که با جمله‌ای از این دست روبرو می‌شود در تحلیل نحوی جمله فوق مشکلی ندارد بلکه برای حفظ امانت در ترجمه نیازمند آن است که بداند عبارت موصولی **who was on the balcony** به کدام یک از گروه‌های اسمی موجود در جمله برمی‌گردد. به عبارت دیگر آیا "هنرپیشه" در بالکن بوده است یا "خدمتکار هنرپیشه". بررسی‌های آماری که بر روی درک و تولید جمله فوق در میان گویشوران انگلیسی زبان صورت گرفته است نشان می‌دهد گویشوران انگلیسی زبان به منسوب کردن عبارت موصولی به گروه اسمی دوم در جملاتی نظیر جمله (۱۱) گرایش دارد در حالی که گویشوران اسپانیولی زبان تمایل به نسبت دادن عبارت موصولی به گروه اسمی اول دارند. دستیابی به چنین نتایجی صرفاً با بهره‌گیری از روش‌های آماری در تحلیل پیکره‌های زبانی امکان‌پذیر است.

بررسی‌های آماری بر روی واژه **that** در **Brown Corpus** نشان می‌دهد که این

واژه می تواند در مقام وابسته اسم (**determiner**)، متمم نما (**complementizer**)، ضمیر (**pronoun**)، یا تشدیدکننده (**intensifier**) ظاهر شود. فراوانی خوانش **that** به عنوان ضمیر بسیار بالاتر از خوانش های دیگر این واژه است اما بسامد وقوع این مقوله های متفاوت به بافت نحوی بستگی دارد.

	Determiner	Complementizer
Start of sentence	35%	11%
After verb	6%	93%

12. The lawyer insisted that experienced diplomat would be very helpful.

13. The lawyer insisted that experienced diplomats would be very helpful.

14. That experienced diplomat would be very helpful to the lawyer.

15. That experienced diplomats would be very helpful made the lawyer confidence. (Jurafsky)

نمونه های فوق احتمال خوانش نادرست جملات (۱۲) و (۱۵) را در نگاه اول، بر پایه محاسبات آماری مطرح می نماید.

۳- در حوزه مطالعات واژ-واجی، همجواری یک همخوان انسدادی لثوی با یک نیم واکه کامی در مرز دو واژه منجر به کامی شدگی همخوان لثوی می گردد. این قاعده آوایی قادر به توجیه تلفظ کلماتی چون **would , told/did** در همجواری واژه **you** است:

did you	[did ₃ yu]
told you	[tauld ₃ yu]
would you	[wud ₃ yu]

اما همین قاعده آوایی قابل تعمیم به ساخت‌هایی از قبیل **at you** و **but you** و **good you** نخواهد بود.

بایبی و شیب‌مان (۱۹۹۹) عدم تعمیم قاعده آوایی فوق را در ساخت‌هایی از این دست وابسته به فراوانی وقوع واژه‌های همنشین دانسته، با محاسبه احتمال وقوع جفت کلماتی از این قبیل معتقدند در جفت کلمات با احتمال وقوع زیاد، انسدادی‌های لثوی بیشتر به سمت کامی شدگی پیش می‌روند در حالی که جفت کلماتی که از احتمال وقوع چندانی برخوردار نیستند غالباً به سمت کامی شدگی پیش نمی‌روند.

زبان‌شناسی پیکره‌ای و نظریه‌های ترجمه

هر چند تحقیقاتی از این دست، می‌تواند در ترجمه شفاهی و مکتوب به یاری مترجمان آید، ولی به نظر می‌رسد استفاده از تحقیقات زبان‌شناسی پیکره‌ای در حوزه نظری ترجمه کاربردی وسیع‌تر دارد. نگاهی به نمونه‌های زیر گویای مطلب است.

16. The old man the boat. (Milne 1982)
17. The complex houses married and single students and their families. (Jurafsky 1992)
18. The man who you saw saw me. (Gibson 1998)
19. The horse raced past the barn fell. (Bever 1970)

نمونه‌های فوق همگی در زبان انگلیسی خوش ساخت تلقی می‌گردند اما درک و ترجمه چنین جملاتی لااقل در نگاه اول ساده نیست.

جورافسکی معتقد است تأخیر در درک جملاتی نظیر (۱۶) و (۱۷) به دلیل ظاهر شدن واژه‌های **man** و **house** در حالت فعلی و نیز ظهور واژه‌های **old** و **complexe** در حالت اسمی است. از آنجا که تجلی واژه‌های فوق در حالت‌های یاد شده از فراوانی وقوع بالایی برخوردار نیست، درک چنین جملاتی با تأمل صورت می‌گیرد.

در مورد جمله (۱۸) گیسون و بسیاری از محققان دیگر معتقدند درک جملات با عبارت موصولی مفعولی به مراتب مشکل‌تر از درک جملات با عبارت موصولی فاعلی است. دلیل این امر فراوانی وقوع ساخت‌هایی از نوع دوم در مقایسه با ساخت‌های نوع اول ذکر شده است. به اعتقاد جورافسکی تأمل در پردازش و درک جمله شماره (۱۹) به علت ضعیف بودن احتمال ظهور فعل **face** به عنوان یک فعل لازم است. این فعل غالباً به صورت یک فعل متعدی همراه با گروه حرف اضافه‌ای ظاهر می‌شود.

نمونه‌های فوق مثنی از خروارند که می‌توانند مورد عنایت نظریه پردازان ترجمه ماشینی قرار گیرند و در تهیه نرم‌افزارهای رایانه‌ای ترجمه بکار روند.
پایان سخن:

در این مختصر سعی بر آن بود تا ضرورت استفاده از روش‌های احتمالاتی و مدل‌های آماری در تحلیل داده‌های زبان بازگو گردد.

صفوی (۱۳۸۰) معتقد است معنی واژه‌ها برحسب همنشینی با واژه‌های دیگر تغییر می‌کند و درک هریک از معانی چندگانه یک واژه منوط به درک معنای واحدهای هم‌نشین شده در کنار یکدیگر است. به اعتقاد وی می‌توان با قاطعیت گفت که در زبان خودکار کمتر واژه‌ای یافت می‌شود که در همنشینی با واژه‌های دیگر تغییر معنی ندهد. از سوی

دیگر وی حذف واژه‌های هم‌نشین را نیز سبب تغییر معنی واژه‌های محذوف می‌داند. چنین استدلالی می‌تواند در مورد تمامی نمونه‌هایی که تا کنون در این نوشته ذکر شده صادق باشد، اما نمونه (۲۰) وضعیتی متفاوت را بیان می‌کند:

20. Sometimes I feel like nuts and sometimes I feel like nuts.^(۱)

نمونه (۲۰) یک کنش واقعی زبان است که نگارنده خود با آن روبرو شده است. پردازش جمله (۲۰) به تأخیر بسیار صورت می‌گیرد و به نظر می‌رسد نمونه‌ای از چند معنایی در سطح جمله باشد که از طریق چند معنایی ترکیبی یا چند معنایی انتخابی بر پایه مقاله یاد شده توجیه نمی‌گردد. از یک سو هم‌نشینی واحدهای هم‌نشین شده در کنار یکدیگر به تعبیرهایی چندگانه منجر خواهد شد و از سوی دیگر انتخاب استعاری واژه‌ای به جای واژه دیگر در این نمونه صورت نگرفته است.

به اعتقاد صفوی هم‌آوایی - هم‌نگاشتی در نظام واژگان زبان دارای نقش است. به همین دلیل هم در واژگان زبان واحدهای مستقلی هستند و هر کدام، در فرهنگ لغت، مدخل جداگانه‌ای دارند. از این رو می‌توان دو واژه **like** به مفهوم دوست داشتن و **like** در حالت حرف اضافه‌ای و به مفهوم شباهت را دو واژه هم‌آوا - هم‌نگاشت تلقی کرد اما واژه **nuts** در فرهنگ لغت **oxford** دو زیر یک مدخل اسمی با معانی متفاوت "آجیل" و "دیوانه" ذکر شده است.

مقایسه نمونه (۲۰) با نمونه‌ای از جملات فارسی که دارای ابهام معنایی هستند مفید به نظر می‌رسد. نمونه (۲۱) بیانگر ابهام معنایی در سطح جمله در زبان فارسی است:

۲۱ آن یکی شیری است اندر بادیه و آن یکی شیری است اندر بادیه

۱- این حالت برگرفته از یک آگهی تلویزیونی است.

تفاوت دو نمونه فارسی و انگلیسی در این نکته نهفته است که در نمونه (۲۱) هر دو واژه "شیر" و "بادیه" دارای دو معنای جداگانه‌اند که در مدخل‌هایی جداگانه در فرهنگ لغت ذکر شده‌اند. به عبارت دیگر "شیر" در معنای مایعی خوراکی و "شیر" در معنای حیوانی درنده دو واژه هم‌آوا - هم‌نگاشت محسوب می‌شوند. همچنین "بادیه" در معنای صحرا و "بادیه" در معنای ظرف نیز بر همین اساس هم‌آوا - هم‌نگاشت هستند. در حالی که در نمونه (۲۰) لااقل یکی از واژه‌هایی که در ابهام معنایی جمله نقش دارد به عنوان واژه چند معنی تلقی شده است.

به زعم نگارنده این سطور، بر اساس شم زبانی و نه بر اساس یک تحقیق آماری، فراوانی وقوع واژه "شیر" در دو مفهوم جداگانه مورد نظر یکسان به نظر می‌رسد. از سوی دیگر واژه "بادیه" نیز دارای دو مفهوم جداگانه است که هیچ یک در واژگان امروزه فارسی زبانان معمول نیست و باز بر اساس شم زبانی نگارنده از فراوانی و نوعی تقریباً معادل صفر برخوردارند. مقایسه‌ای تقریباً مشابه از نظر فراوانی وقوع در مورد واژه‌های ابهام زای نمونه (۲۰) نیز بیانگر آن است که واژه‌های **like** در حالت فعلی و **like** در حالت حرف اضافه‌ای احتمالاً از بسامد وقوع تقریباً یکسانی برخوردارند. همانگونه که واژه **nuts** نیز به عنوان یک واژه چند معنایی می‌تواند به ترتیب احتمال در داده‌های زبانی بطور تقریباً یکسانی ظاهر گردد.

به نظر می‌رسد تحلیل داده‌هایی از این دست با مددگیری از پیکره‌های زبانی و محاسبات احتمالاتی میسر نباشد. هرچند پردازش جمله‌ای همچون نمونه (۲۱) به احتمال قوی با توجه به بافت گفتمان صورت می‌گیرد، اما نگارنده این سطور خود در معرض داده شماره (۲۰) با تصویرپردازی تلویزیونی یعنی بافت غیر کلامی قرار گرفته است.

تردیدی نیست که نقش بافت کلامی و غیر کلامی در پردازش زبان بسیار اساسی است،

اما آنگاه که گویشوران فارسی زبان در مواجهه با جمله‌ای نظیر نمونه (۲۱) و گویشوران انگلیسی زبان در رویارویی با نمونه (۲۰) دچار تردید و تأمل می‌گردند و درک معنای جمله توسط آنان با تأخیر صورت می‌گیرد، آیا می‌توان انتظار داشت که مترجم برای برگرداندن این مفهوم به زبانی دیگر دچار حیرت نگردد و به طریق اولی چگونه می‌توان در نرم‌افزار رایانه‌ای ترجمه چنین ساخت‌هایی را پیش‌بینی کرد و انتظار پردازش آنها را از ماشینی ذیشعور داشت. به نظر می‌رسد علم با تمامی پیشرفت‌های روز افزونش هنوز در ابتدای راه قرار دارد و استفاده از زبان‌شناسی پیکره‌ای و داده بنیاد تنها بخشی کوچک از مشکلات علم زبان را حل می‌کند. اما آنچه مسلم است این است که زمینه‌های تحقیقاتی بسیاری در رابطه با مسایل یاد شده وجود دارد.

References:

- Croft, W.** 1995. *Typology and Universals*. Cambridge: Cambridge University Press.
- Crystal, D.** 1994. *An Encyclopedic Dictionary of Language and Languages*. Oxford: Oxford University Press.
- Crystal, D.** 1992. *A Dictionary of Linguistics and Phonetics*. London: Blackwell.
- Jurafsky, D.** 2002. *Probabilistic Modeling in Psycho-Linguistics: Linguistic Comprehension and Production*. Michigan: MIT Press.
- Manning, C. and Schutze Heinrich.** 1999. *Foundation of Statistical Natural Language Processing*. Michigan: MIT Press.

بی جن خان، محمود. نقش پیکره‌های زبانی در نوشتن دستور زبان: معرفی یک نرم‌افزار رایانه‌ای. در دست چاپ در مجله زیانشناسی (۱۳۸۲).

صفوی، کورش. مقام هم معنایی و چند معنایی در ترجمه، هفت گفتار درباره ترجمه، کتاب ماه (۱۳۸۰).

صفوی، کورش. نگاهی تازه به مسأله چند معنایی واژگانی، نامه فرهنگستان، شماره ۲ (۱۳۸۰).

۹- منبع اینترنتی:

<http://linguistlist.org/issues/14/14-2103.htm>.