نقش حروف ربط زمان‌دار در تعیین رابطهٔ زمانی بین رویدادهای فعلی در پیکرهٔ متون زبان فارسی معاصر

مرضیه مرتضوی‌نیا، محمود بی‌جن‌خان و پروین‌سادات فیض‌آبادی
گروه زبانشناسی، دانشکدهٔ ادبیات و علوم انسانی، دانشگاه تهران

چکیده:

این مقاله به پیش‌بینی روابط زمانی بین رویدادهای فعلی بر اساس حروف ربط زمان‌دار در متون می‌پردازد. برای انجام این کار، داده‌های مربوط به رویدادهای فعلی از پیکرهٔ متنی زبان فارسی استخراج و بررسی شدند و روابط زمانی بین رویدادها مشخص شد. با تحلیل داده‌ها براساس روابط زمانی در چهارچوب طبقه بندی بِرد و آلن، مشخص شد که حروف ربط زمان‌دار را می‌توان در چهار دستهٔ تعینی، اریب‌دار، شبه‌تصادفی و تصادفی جای داد. علائم تعینی به‌طور دقیق روابط زمانی بین رویدادهای مربوطه را پیش‌بینی کرده و علائم اریب‌دار با تقریب خوبی یک دسته رابطه را پیش‌بینی می‌کنند. علائم شبه‌تصادفی گرایش به یک نوع رابطه‌ی زمانی خاص ندارند و علائم تصادفی، به‌طور تقریبی هر گونه رابطهٔ زمانی را بین رویدادهای مربوطه امکان‌پذیر می‌کنند. این پژوهش نشان داد که پیش‌بینی روابط زمانی بِرد به‌وسیلهٔ علائم تا حدودی مطلوب‌تر از پیش‌بینی روابط آلن می‌باشد.

واژگان کلیدی: روابط زمانی رویدادها، حرف ربط، رویداد فعلی

۱- مقدمه

سیستم‌های پردازش زبان طبیعی در حوزه‌های مختلف علمی، آموزشی، سیاسی، تجاری و غیره و با اهداف مختلف به‌کار گرفته می‌شوند. ازجمله مهم‌ترین کاربردهای سیستم‌های پردازش زبان طبیعی می‌توان به موارد زیر اشاره کرد *(Bolshakov and Gelbuk, 2004: 53)*:

بازیابی اطّلاعات، ترجمهٔ خودکار، استخراج داده‌های حقیقی، فهم زبان طبیعی، تولید متن از تصویر، خلاصه‌سازی متن و غیره.

با گسترش روزفزون اطّلاعات متنی، به‌کار بردن ابزارهایی که به کاربران امکان خلاصه‌کردن متون را بدهند، بسیار سودمند می‌باشد. در خلاصه‌سازی متن، تلاش می‌شود تا محتوای اطلاعاتی رویدادها و ترتیب زمانی آن‌ها از درون متن استخراج و مهم‌ترین اطلاعات آن با توجه به نیاز کاربران به‌صورت خلاصهٔ متنی، به آنها عرضه شود *(Mani and Bloedorn, 1998: 35)*.

مسئلهٔ دیگری که در همین اواخر مورد توجّه زبان‌شناسی رایانه‌ای و هوش مصنوعی قرار گرفته، بازشناسی خودکار رویدادها و روابط زمانی در زبان طبیعی است. استخراج روابط زمانی بین رویدادها برای شرح وقایع به‌ترتیب زمانی در بسیاری از حوزه‌های پردازش زبان طبیعی اهمیت بسیار دارد *(Dorr and Gassterland, 2007: 1681)*.

طراحی سیستم‌هایی که توانایی درک روابط زمانی بین رویدادها را داشته باشند، مستلزم تولید پیکره‌هایی‌ست که در آن‌ها برچسب رویدادها و زمان‌ها موجود باشد. در سال‌های اخیر، چندین پیکره بر اساس استاندارد زبان نشانه‌گذاری زمان تایم‌ام‌ال[1] *(Sauri et.al, 2006: 1)*، برچسب‌دهی[2] شده‌اند. این پیکره‌ها شامل پیکرهٔ تایم‌بانک ۱-۱، تایم‌بانک ۱-۲ *(Pustejovsky et.al, 2003: 647)*، اکوینت[3] یا آپینین، اُتی‌سی[1] و تمپ ایول[2] می‌باشند. پیکره

[1] Time Markup Language
[2] Annotation
[3] Aquaint

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-02]

تایم‌بانک ۱-۱ بر اساس زبان نشانه‌گذاری زمان تایم‌ام‌ال ۱-۱ و پیکره‌های تایم‌بانک ۱-۲ و اکوینت بر اساس نسخهٔ ۱-۲-۱ تایم‌ام‌ال برچسب‌دهی شده‌اند *(Boguraev et.al, 2007: 1)*. پیکرهٔ اُتی‌سی تلفیقی از پیکرهٔ تایم‌بانک ۱-۲ و پیکرهٔ اکوینت می‌باشد. پیکرهٔ تمپ‌ایول برای کارگاه ارزیابی معنایی[3] ۲۰۰۷ در پراک طراحی شده و حاوی همان سندهای پیکرهٔ تایم‌بانک ۱-۲ می‌باشد *(Verhagen et.al., 2007: 75)*.

هدف از این مقاله، شناسایی و توصیف زبانی حروف ربط زمان‌دار به‌عنوان بخشی از علائم زبانی در تعیین رابطهٔ زمانی بین رویدادهای فعلی است. این پژوهش با تحلیل داده‌ها بر اساس روابط زمانی در چهارچوب طبقه‌بندی بِرد و آلن، نشان داد که حروف ربط زمان‌دار را می‌توان در چهار دستهٔ تعینی، اریب‌دار، شبه‌تصادفی و تصادفی جای داد. علائم تعینی، به‌طور دقیق روابط زمانی بین رویدادهای مربوطه را پیش‌بینی می‌کنند، علائم اریب‌دار با تقریب خوبی یک دسته رابطه را پیش‌بینی می‌کنند، علائم شبه‌تصادفی گرایش به یک نوع رابطه‌ی زمانی خاص ندارند و علائم تصادفی، به‌طور تقریبی هر گونه رابطهٔ زمانی را بین رویدادهای مربوطه امکان‌پذیر می‌کنند. در ادامهٔ این مقاله، ابتدا در (بخش ۲) مفهوم رویداد و تعاریف بِرد و آلن برای رویداد معرّفی می‌شوند. در (بخش ۳) پیکرهٔ متنی زبان فارسی معاصر معرّفی می‌شود. در (بخش ۴)، به معرّفی برچسب‌های استاندارد ایگلز پرداخته می‌شود. (بخش ۵)، داده‌های مورد بررسی را معرفی می‌کند. (بخش ۶) به تجزیه و تحلیل روابط زمانی میان رویدادها بر اساس روابط هفتگانه آلن می‌پردازد. در (بخش ۷) تحلیلی از این داده‌ها ارائه می‌شود. (بخش ۸) به تجزیه و تحلیل روابط زمانی میان رویدادها بر اساس روابط سه‌گانه بِرد می‌پردازد و در (بخش ۹) بحث و نتیجه‌گیری مقاله ارائه می‌شود.

۲- مفهوم رویداد

مفهوم رویداد مورد توجّه حوزه‌های مختلف علوم ازجمله منطق، فلسفه، روانشناسی، زبانشناسی نظری و ریاضیات می‌باشد. هر یک از این علوم از منظر خاصی به رویداد نگاه می‌کنند که در ادامه به تجزیه و تحلیل ریاضی روابط زبانی بین رویدادها می‌پردازیم.

بِرد در مقالهٔ خود *(Bird and Ewan, 1990: 33)*، هر رویداد را متشکل از یک بازه و یک خصیصه می‌داند و اصولی برای روابط زمانی میان رویدادها ارائه می‌دهد. وی ساخت رویداد را متشکل از رویدادها و روابط زمانی میان آنها می‌داند. بِرد روابط زمانی هم‌زمانی، تقدم و هم‌پوشی را به‌عنوان روابط زمانی حاکم بر واحدهای زمانی برمی‌شمرد و تعاریفی صوری از آنها ارائه می‌دهد. وی در تعاریف خود، روابط را مبتنی بر برخورداری یا عدم برخورداری از سه ویژگی انعکاسی، متقارن و متعدی می‌داند. بر این اساس:

هم‌پوشی زمانی، رابطه‌ای انعکاسی، متقارن و نامتعدی است.

تقدم زمانی، رابطه‌ای غیرانعکاسی، نامتقارن و متعدی است.

هم‌زمانی، رابطه‌ای متقارن، انعکاسی و متعدی است.

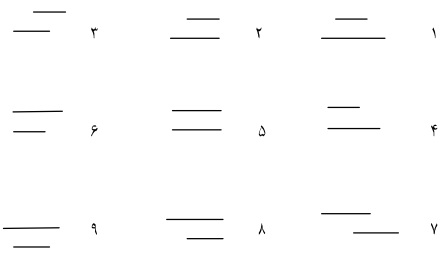
بِرد ادعا می‌کند که بازه‌ها می‌بایست بر اساس نقاط آغازین و پایانی‌شان و نه نقاط تشکیل دهندهٔشان تعریف شوند. او تعاریف سیجی را که بر اساس نقاط هستند، با تعاریف خود بر مبنای نقاط آغازین و پایانی اصلاح می‌کند:

الف) x بر y تقدم دارد، اگر و فقط اگر $\max(x)<\min(y)$

ب) x با y هم‌پوشی دارد، اگر و فقط اگر $\max(y)> \min(x)$ و $\max(x) > \min(y)$.

در روابط فوق، $\max(x)$ و $\max(y)$ به‌ترتیب، نشان‌دهندهٔ آخرین عنصر بازه x و y بوده و $\min(x)$ و $\min(y)$ نشان‌دهندهٔ اوّلین عنصر این دو بازه می‌باشند.

وی در تبیین رابطهٔ هم‌پوشی دو رویداد $e_1 o e_2$، ۹حالت را برمی‌شمرد که در (شکل ۱) آمده‌اند. در این نمودارها، هر خط نشان‌دهندهٔ بازهٔ یک رویداد است. خطّ بالا رویداد اوّل و خطّ زیرین، رویداد دوم را نشان می‌دهد.



(شکل۱): حالت‌های مختلف رابطهٔ هم‌پوشی

آلن *(Allen, 1984: 123)* طبقه‌بندی دیگری از روابط میان بازه‌های زمانی ارائه می‌دهد که به‌صورت زیر است:

الف) DURING(t1 , t2): بازهٔ زمانی t1 به‌طور کامل در بازهٔ t2 است.

[1] Opinion TimeBank Corpus (OTC)
[2] TempEval
[3] SemEval (Semantic Evaluation) Workshop

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-02]

ب) STARTS(t1 , t2): بازه‌های زمانی t1 و t2 دارای شروع یک‌سانی هستند؛ امّا بازهٔ t1، قبل از بازهٔ t2 تمام می‌شود.

ج) FINISHES(t1, t2): بازه‌های زمانی t1 و t2 با هم تمام می‌شوند؛ امّا بازهٔ t1 بعد از بازهٔ t2 شروع می‌شود.

د) BEFORE(t1 , t2): بازهٔ زمانی t1 قبل از بازهٔ t2 است و دو بازهٔ هیچ هم‌پوشی با هم ندارند.

ه) OVERLAP(t1 , t2): بازهٔ زمانی t1 قبل از بازهٔ t2 شروع می‌شود و دو بازهٔ با هم، هم‌پوشی دارند.

و) MEETS(t1 , t2): بازهٔ زمانی t1 قبل از بازهٔ t2 است، امّا هیچ بازه‌ای بین آن دو نیست؛ بدین معنی که درست در زمانی که t1 تمام می‌شود، t2 شروع می‌شود.

با در نظر گرفتن روابط فوق و روابط معکوس آنها می‌توان سیزده نوع رابطه زمانی در نظر گرفت. رابطهٔ زمانی که بین دو بازهٔ (A;B) برقرار است، می‌تواند یکی از این سیزده نوع رابطهٔ زیر را داشته باشد (A و B، زمان یا رویداد هستند.):

B برابر با A، A برابر با B، B قبل از A، A قبل از B، A بلافاصله قبل از B، B بلافاصله بعد از A، A هم‌پوشی دارد با، B به‌وسیلهٔ A هم‌پوشی شده‌است، A با B آغاز می‌شود، B با A آغاز می‌شود، A با B پایان می‌یابد، B با A پایان می‌یابد، A مشمول B است، B شامل A است.

بر اساس چمبرز و دیگران (*Chambers, et.al, 2997: 173*)، شش رابطه از مجموعه سیزده رابطه عنوان شده، معکوس شش رابطهٔ دیگر هستند و رابطه "برابر با"، همان رابطهٔ هم‌زمانی است. بنابراین با صرف‌نظر کردن از شش رابطهٔ معکوس که حشو محسوب می‌شوند، تعداد روابط به هفت رابطه کاهش می‌یابد. این هفت رابطه عبارتند از:
قبل، بلافاصله قبل، شامل، شروع می‌شود، پایان می‌یابد، هم‌زمان و هم‌پوشی.

با در نظر گرفتن اصول صوری که بِرد در مقالهٔ خود برای هم‌پوشی و تقدم و هم‌چنین برای روابط شمول، هم‌زمانی و تقدم بلافصل ارائه می‌دهد و با توجّه به روابط زمانی ارائه شده توسط آلن، می‌توان اصول صوری هفت‌گانهٔ زیر را برای هفت نوع رابطهٔ زمانی ذکر شده ارائه داد:

الف) قبل (BEFORE):
X قبل از y است اگر و فقط اگر $\max(x) \min(y)$

ب) بلافاصله قبل (IBEFORE):
به‌ازای هر $x, y \in E$ ، $x \prec^{\circ} y$ اگر و فقط اگر $x \prec y$ و هیچ $z \in E$ وجود نداشته باشد که $x \prec z \prec y$.

ج) شمول (INCLUSION):
به‌ازای هر $x, y \in E$ ، $x \subseteq y$ اگر و فقط اگر هر $z \in E$ که با x هم‌پوشی دارد، با y نیز هم‌پوشی داشته باشد و $\min(x) \neq \min(y)$ و $\max(x) \neq \max(y)$

د) شروع می‌شود (BEGINS):
به‌ازای هر $x, y \in E$ ، $x \lhd y$ اگر و فقط اگر $\min(x) = \min(y)$ و $\max(x) = \max(y)$

ه) پایان می‌یابد (ENDS):
به‌ازای هر $x, y \in E$ ، $x \rhd y$ اگر و فقط اگر $\min(x) \neq \min(y)$ و $\max(x) = \max(y)$

و) همزمان (SIMULTANEOUS):
به‌ازای هر $x, y \in E$ ، $x \leftrightarrow y$ اگر و فقط اگر $\min(x) = \min(y)$ و $\max(x) = \max(y)$.

ز) هم‌پوشی (OVERLAP):
XoY، اگر و فقط اگر $\max(x) > \min(y)$ و $\max(y) = \min(x)$.

۳- پیکرهٔ متنی زبان فارسی معاصر

برای شناسایی و توصیف روابط زمانی بین رویدادهای فعلی از پیکرهٔ متنی زبان فارسی معاصر استفاده شده‌است *(بی‌جن‌خان ۱۳۸۳، ۴۸)*.

پیکرهٔ متنی زبان فارسی معاصر، شامل ۳۵۰۵۸ پروندهٔ متنی است. هر پرونده شامل یک متن کامل یا نمونهٔ تصادفی منتخب از یک متن کامل است که براساس معیارهای غیرزبانی و هم‌چنین تعلقشان به گونه‌های فارسی معاصر انتخاب شده‌اند. حجم هر متن رشته‌ای از حدّاقل هزار کلمه است؛ اگرچه حجم بعضی از متون روزنامه‌ای که شامل اخبار کوتاه یا پیام‌های بازرگانی است، کمتر است. این پیکرهٔ بزرگ متنی شامل یکصد میلیون کلمه می‌باشد که از منابع مختلف و سیاق‌های متنوع جمع‌آوری شده‌اند. ده میلیون کلمه از آن با استفاده از راهنمای ایگلز به‌صورت سلسله‌مراتبی نشانه‌گذاری صرفی، نحوی و معنایی شده‌اند *(بی‌جن‌خان ۱۳۸۳، ۴۸)*.

محتوای پیکره، تحت ۲۴ مقولهٔ متنی سازمان‌دهی شده است*(بی‌جن‌خان ۱۳۸۳، ۴۸)*: داستانی، نمایشنامه رسمی، علمی، روزنامه، مجلات عمومی، مجلات تخصصی، آموزشی، کتابچه راهنما، قانون اساسی، طرح‌های دولتی، آیین‌نامه‌ها و مقررات، گزارش سفر، سالنامه، مکاتبات اداری، صورت‌جلسات، یادداشت‌های روزانه، نامه‌های شخصی، بروشور، سخنرانی فی‌البداهه، سخنرانی از قبل تهیه شده، نمایشنامه غیررسمی، اخبار، مکالمات تلفنی و مصاحبه.

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-02]

۴- برچسب‌های استاندارد ایگلز

استاندارد ایگلز ابتدا برای تفسیر متون زبان‌های اروپایی پیشنهاد شد. امّا به‌تدریج برای تفسیر زبانی متون سایر زبان‌ها از جمله زبان فارسی نیز به‌کار گرفته شد. راهنمای ایگلز، قائل به سیزده مقولهٔ اصلی دستوری است که همان اجزای کلام در دستور سنتی هستند. علاوه‌بر آن، به‌ازای هر مقولهٔ اصلی، حداکثر سه نوع زیرمقولهٔ فرعی در نظر گرفته شده‌است به‌طوری که هر کدام ویژگی‌های تفصیلی مقولهٔ اصلی مربوطه‌اش را توصیف می‌کند:

۱) مقوله‌های فرعی توصیه شده: شامل ارزش‌های تصریفی ساخت واژه؛

۲) مقوله‌های ذاتی: شامل ویژگی‌های معنایی که به‌صورت اختیاری به برچسب اضافه می‌شوند؛

۳) مقوله‌های خاص زبان: شامل ویژگی‌های خاص نحوی و معنایی زبان، که به‌صورت اختیاری به برچسب اضافه می‌شوند.

مقولهٔ زمان در پیکره برای مقوله‌های اسم و قید ذیل مقوله‌های ذاتی آورده شده‌اند

کاربرد استاندارد ایگلز در نشانه‌گذاری متون فرعی، منجر به تعریف ۱۴ برچسب برای مقوله‌های اصلی، ۵۳ برچسب برای مقوله‌های توصیه شده، ۲۵ برچسب برای مقوله‌های ذاتی و ۱۸ برچسب برای مقوله‌های خاص زبان فارسی شد که در مجموع ۱۱۰ برچسب می‌شوند *(روح‌پرور و بی‌جن‌خان ۱۳۸۶)*.

۵- داده‌ها

به‌منظور تحلیل روابط زمانی میان رویدادها با استفاده از علائمی که آنها را حروف ربط زمان‌دار می‌نامیم، ابتدا تمام حروف ربط پیکرهٔ متنی زبان فارسی استخراج شدند. برای استخراج این حروف از نرم‌افزار "جستجوگر پیکرهٔ متنی زبان فارسی"، تهیه شده در پژوهشکده پردازش هوشمند علائم، استفاده شد.

نتیجهٔ جستجو نشان داد که پیکرهٔ متنی دارای ۲۹۵ حرف ربط می‌باشد. از آنجا که در این تحقیق، حروف ربط زمان‌دار به‌عنوان علائم زمانی در نظر گرفته شدند و در پیکرهٔ متنی زبان فارسی، حروف ربط زمان‌دار به‌طور جداگانه برچسب نخورده‌اند، در این مرحله، این علائم از کل حروف ربط مجزا شدند.

پس از استخراج تمام علائم، علائمی که بیش از یک صورت نوشتاری دارند، مانند "هنگامیکه" و "هنگامی که" که در پیکره به‌صورت دو علامت متفاوت برچسب خورده‌اند و علائمی که قابلیت به کار رفتن به جای علامت دیگری را دارند، مثل "تا این که" که می‌توان آن را به‌جای "تا آن که" به کار برد، با هم در یک دسته قرار گرفتند.

پس از استخراج حروف ربط، جداکردن علائم (حروف ربط زمان‌دار) و مرتب کردن آن‌ها، جدول (۱) بر اساس فراوانی علائم در پیکرهٔ متنی به‌دست آمد.

(جدول ۱): علائم پیکره بر حسب فراوانی

ردیف	علامت	فراوانی
۱	تا	۹۸۷۳
۲	چون	۳۷۲۲
۳	وقتی	۲۹۷۷
۴	سپس	۲۰۲۳
۵	در حالی که/ در حالیکه	۱۷۸۹
۶	پس	۱۵۶۰
۷	بعد	۱۴۴۱
۸	در عین حال	۸۰۰
۹	آنگاه/ آن گاه	۶۱۰
۱۰	همان طور که/ همان طوری که/ همانوریکه/ همانطوری که	۵۷۹
۱۱	هنگامی که/ هنگامیکه/ هنگامی	۴۳۶
۱۲	ضمناً/ ضمنا/ در ضمن/ درضمن	۵۸۴
۱۳	در این حال	۳۰۶
۱۴	همچنان که/ همچنانکه	۲۲۳
۱۵	وقتی که/ وقتیکه	۱۵۹
۱۶	چو	۱۳۴
۱۷	در این میان	۱۲۹
۱۸	تااینکه/ تا این که/ تاآنکه/ تا آن که	۱۲۲
۱۹	ضمن اینکه/ ضمن این که	۱۱۴
۲۰	در همین حال	۱۰۲
۲۱	مادامی که/ مادامیکه/ مادامکه/ مادامی/مادام	۸۴
۲۲	همین که/ همینکه	۵۵
۲۳	به محض اینکه/ به محض این که/ به محض آنکه/ به محض آن که	۳۴
۲۴	زمانی که/ زمانیکه	۳۱
۲۵	پیش از آنکه/ پیش از آن که/ پیش از اینکه/ پیش از این که	۱۹
۲۶	به مجرد اینکه/ به مجرد این که/ به مجرد آن که/ به مجرد آنکه	۷
۲۷	در این بین	۶
۲۸	حالیه	۱
۲۹	زین پس	۱
۳۰	پس از آنکه/ پس از آن که	۱

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-02]

پس از تهیهٔ فهرستی از علائم، جملات حاوی آنها با استفاده از نرم افزار جستجوگر پیکره، استخراج شدند و سپس از بین جملات هر علامت، ده جمله به‌صورت تصادفی انتخاب شده و رویدادهای فعلی آنها که فعل اصلی جملات بودند به‌منظور تحلیل روابط زمانی مشخص شدند.

در این تحقیق به بررسی روابط زمانی میان رویدادهایی از نوع افعال زمان‌دار پرداخته شده‌است. بدین منظور برای هر علامت، جدولی مطابق (جدول۲) تهیه شده و افعال زمان‌داری که در هر سند با آن علامت به هم مربوط شده‌اند از سند مربوطه استخراج و در این جدول قرار داده شدند. سپس زمان افعال نیز در جدول قرار گرفتند و درنهایت رابطه یا روابط زمانی ممکن میان رویدادها بررسی شدند. (جدول ۲) نمونه‌ای از این جداول را نشان می‌دهد.

(جدول ۲): رویدادهای فعلی و رابطه‌ی زمانی آن‌ها برای هر علامت

ردیف	شماره سند	فعل (E1)	زمان فعل (E1.T)	حرف ربط	فعل (E2)	زمان فعل (E2.T)	نوع رابطه زمانی (relType)
۱	۱	برسیدم	ماضی ساده	چون	کرد	ماضی ساده	IBEFORE
۲	۲۹۸۵	دید	ماضی ساده	چون	شد	ماضی ساده	BEGINS
۳	۱۱	رفتیم	ماضی ساده	چون	در پیش امد	ماضی ساده	IBEFORE

۶- تجزیه و تحلیل روابط زمانی میان رویدادها بر اساس روابط هفت‌گانهٔ آلن

در تحلیل روابط زمانی با استفاده از روابط هفت‌گانه آلن، رابطهٔ بین دو رویداد فعلی در هفت دستهٔ Before، IBefore، Begins، Ends، Simultaneous، Inclusion، Overlap قرار می‌گیرد. (جدول۲) تحلیل زمانی حرف ربط زمان‌دار "چون"، بر اساس روابط هفتگانه آلن را نشان می‌دهد.

پس از تهیهٔ جدولی مانند (جدول ۲) برای تمام حروف ربط زمان‌دار، درصد پیش‌بینی روابط مشخص شده برای هر علامت، اندازه‌گیری شد. به‌عنوان مثال، هنگامی‌که در اسناد بررسی شده، علامت "تا" در بیست درصد موارد، رابطهٔ BEFORE را نشان داده‌است، پیش‌بینی می‌شود که در بیست درصد سندهای مربوط به این علامت در پیکره، رابطهٔ بین دو رویداد، از نوع رابطهٔ BEFORE باشد. این درصدگیری برای تمام علائم انجام شد.

۷- تحلیل داده‌ها

با توجّه به داده‌های به‌دست آمده، علائم زمانی بر حسب پیش‌بینی‌پذیری رابطه بین علامت و نوع رابطه زمانی در چهار طبقه قرار گرفتند:

۱) علائم تعینی: این علائم به‌طور دقیق روابط زمانی بین رویدادهای مربوطه را پیش‌بینی می‌کنند و به‌صورت صد درصد، تنها یک ستون از روابط را برای رویدادهای مربوط به خود پیش‌بینی می‌کنند. این علائم عبارتند از: مادامیکه/مادام که/ مادامی/ مادام، پیش از آنکه/ پیش از اینکه، به مجرد این که/ به مجرد آنکه/ به مجرد اینکه، حالیه، پس از آنکه، زین پس، بعد، ضمنا/ در ضمن.

۲) علائم اریب‌دار: این علائم به‌دلیل گرایش زیاد به یک نوع خاص از روابط زمانی، اریب‌دار نامیده می‌شوند؛ با تقریب خوبی یک نوع از روابط را پیش‌بینی می‌کنند و درصد پیش‌بینی آن‌ها برای یک نوع از روابط، تفاوت آشکاری با سایر انواع رابطه دارد. مانند علامت "در این میان" که در نود درصد موارد، رابطهٔ INCLUSION و در ده درصد موارد، رابطه BEFORE را پیش‌بینی می‌کند. علائم این دسته عبارتند از: هم‌چنان‌که، در این میان، تا اینکه/ تا آن که، همین که، زمانی که، در این بین، تا، در این حال، در عین حال، سپس، آن گاه، چون.

۳) علائم شبه‌تصادفی: که گرایش به یک نوع خاص از روابط زمانی ندارند؛ بلکه رفتار تاحدودی تصادفی دارند. مانند علائم "به محض اینکه/ به محض آن که" که در پنجاه درصد موارد رابطهٔ Begins و در پنجاه درصد موارد، رابطهٔ IBEFORE را پیش‌بینی

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-02]

کرده‌اند. علائم این دسته عبارتند از: به محض اینکه/ به محض آنکه، همان طور که، سپس.

۴) علائم تصادفی: این علائم مانند علائم شبه‌تصادفی، گرایش به یک نوع خاص از روابط زمانی ندارند و نمی‌توانند به‌خوبی رابطهٔ زمانی خاصی را بین رویدادها مشخص کنند و به‌طور تقریبی هر گونه رابطه‌ای را بین رویدادهای مربوط به خود امکان‌پذیر می‌دانند. مانند علامت "وقتی که" که در سی درصد موارد رابطه IBEFORE، در ده درصد موارد رابطهٔ INCLUSION، در سی درصد موارد رابطه‌ی SIMULTANEOUS، در بیست درصد موارد رابطهٔ IBEFORE و BEFORE و در ده درصد موارد روابط BEFORE، IBEFORE و INCLUSION را امکان‌پذیر می‌داند. علائم این دسته عبارتند از: چو، در همین حال، وقتی که، در حالی که، وقتی، هنگامی که، ضمن این که.

۸- تجزیه و تحلیل روابط زمانی میان رویدادها بر اساس روابط سه گانه بِرد

در تحلیل روابط زمانی با استفاده از روابط سه‌گانه بِرد آن، رابطهٔ بین دو رویداد فعلی در سه دستهٔ هم‌پوشی، تقدم و هم‌زمانی قرار می‌گیرد. در (جدول ۳) تناظر بین روابط آلن و بِرد آمده‌است.

(جدول۳): مقایسه تناظر روابط زمانی در چهارچوب آرای بِرد و آلن

ردیف	روابط بِرد	روابط آلن
۱	هم‌پوشی	INCLUSION
		BEGINS
		ENDS
		OVERLAP
۲	تقدم	BEFORE
		IBEFORE
۳	هم‌زمانی	SIMULTANEOUS

برای تحلیل روابط زمانی میان رویدادها بر اساس روابط سه‌گانهٔ بِرد، برای تمام حروف ربط زمان‌دار، جدولی مشابه جدول (۲) تهیه شد که در آن در ستون "نوع رابطهٔ زمانی"، از روابط بِرد استفاده شد. سپس درصد پیش‌بینی هر علامت بر حسب روابط بِرد محاسبه شد.

در تحلیل داده‌های این جدول نیز به یک دسته‌بندی مشابه دسته‌بندی روابط آلن دست می‌یابیم:

۱) علائم تعینی. این علائم عبارتند از: هم چنان که، تااینکه/ تا آنکه، حالیه، زین پس، پس از آنکه، تا، سپس، بعد، ضمنا/ در ضمن، مادامیکه/ مادام که/مادام، پیش از اینکه/ پیش از آنکه، به مجرد اینکه، به مجرد آنکه.

۲) علائم اریب‌دار: علائم این دسته عبارتند از: در این میان، همین که، زمانی که، در این بین، چون، آن گاه، در این حال، در عین حال.

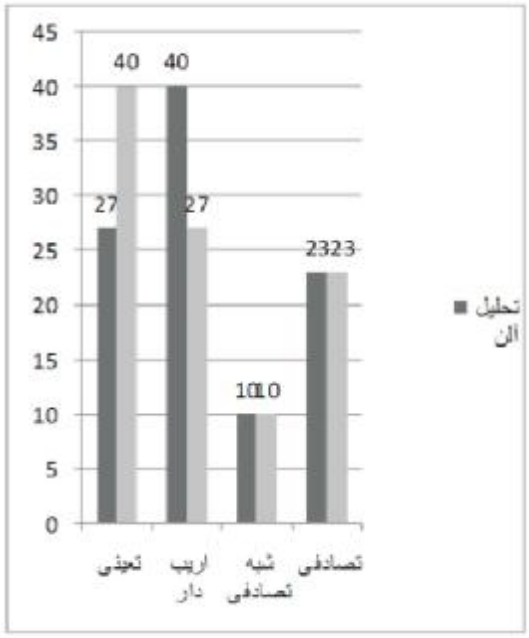
۳) علائم شبه‌تصادفی. این علائم عبارتند از: به محض اینکه/ به محض آنکه، پس، همان‌طور که.

۴) علائم تصادفی. این علائم عبارتند از: هنگامی‌که، چو، ضمن اینکه، در همین حال، وقتی که، درحالی که، وقتی.

۹- بحث و نتیجه گیری

در بخش‌های قبلی، روابط زمانی در چهارچوب طبقه‌بندی سه‌گانهٔ بِرد و طبقه‌بندی هفت‌گانهٔ آلن، تحلیل شدند. میزان پیش‌بینی روابط زمانی بِرد به وسیلهٔ علائم تا حدودی مطلوب‌تر از میزان پیش‌بینی روابط آلن به نظر می‌رسند و علّت آن این است که روابط بِرد، به سبب کلیت بیشتری که نسبت به روابط آلن دارند، پیش‌بینی‌های بهتری را برای علائم ارائه می‌کنند. مقایسهٔ پیش‌بینی روابط زمانی بر اساس روابط آلن و بِرد را می‌توان به‌صورت هیستوگرام در (شکل ۲) مشاهده کرد. طبق این مقایسه، تحلیل نتایج بر اساس روابط بِرد، تعداد بیشتری از علائم را در دستهٔ تعینی قرار می‌دهد. عدم رابطه یک‌به‌یک بین علائم اریب‌دار، شبه‌تصادفی و تصادفی با نوع رابطهٔ زمانی نشان می‌دهد که برای ابهام‌زدایی در این سطح، نیاز به یک روش احتمالاتی نحوی-معنایی از ساختار دستوری جملاتی داریم که در همسایگی علائم قرار دارند.

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-02]



(شکل ۲): مقایسهٔ هیستوگرامی طبقه‌بندی حروف ربط زمان‌دار بر حسب روابط بِرد و آلن

برای پیش‌بینی روابط زمانی بین رویدادها می‌توان هم‌چنین از علائم از نوع حروف اضافهٔ زمان نیز استفاده کرد. هم‌چنین می‌توان به بررسی روابط زمانی میان سایر انواع رویداد از قبیل افعال بدون زمان (مصدرها در زبان فارسی)، اسم‌گردانی‌ها، صفت‌ها، بندهای اسنادی و گروه‌های حرف اضافه‌ای و نیز به بررسی تأثیر ویژگی‌های معنایی رویدادها بر روابط زمانی، از جمله تداومی و لحظه‌ای بودن، پویا و ایستا بودن و هدفمند و بی‌هدف بودن افعال، پرداخت.

۱۰- مراجع

Bolshakov, Igor A; Gelbukh, Alexander; *Computational Linguistics: Models, Resources, Applications*, First Edition, Mexico, National Polytechnic Institute, 2004.

Mani,Inderjeet. Bloedorn, Eric., 1999. “Summerizing Similarities and Differences Among Related Documents.” *Information Retrieval*. 1, pp.35-67.

Dorr, Bonnie J. Gassterland, Terry. 2007. “Exploiting aspectual features and connecting words for summarization-inspired temporal-relation extraction.” *Information Processing and Management.* 43, pp.1681-1704.

Sauri, Roser. Littman, Jessica. Knippen, Bob Gaizauskas, Robert. Setzer, Andrea. Pustejovsky, James., 2006. “TimeML Annotation Guidelines, Version 1.2.”, *Event London*, 1, pp 1-30.

Pustejovsky, James. Hanks, Patrick. Sauri, Roser. See, Andrew. Day, David. Ferro, Lisa. Gaizauskas, Robert. Lazo, Marcia. Setzer, Andrea. Sundheim, Beth., 2003. “The timebank corpus.” *Corpus Linguistics.* pp. 647-656.

Boguraev, Branimir. Pustejovsky, James. Ando, Rie. Verhagen. Marc., 2007. “TimeBank evolution as a community resource for TimeML parsing”, *Language Resources And Evaluation*, 41, pp 91-115.

Verhagen, Marc. Gaizauskas, Robert. Schilder, Frank. Hepple, Mark. Katz, Graha,. Pustejovsky, James., 2007. “TempEval Temporal Relation Identification.” Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007), Prauge, pp. 75-80.

Bird, Steven. Klein, Ewan., 1990. “ Phonological Events.” *Journal of Linguistics*. 26, pp.33-56.

Allen, James F., 1984. “Towards a General Theory of Action and Time. “ *Artificial Intelligence*. 23, pp.123-154.

Chambers, Nathanael. Wang, Shan. Jarafsky, Dan., 2007. “Classifying Temporal Relations Between Events.” *Proceedings of the ACL Demo and Poster Sessions*, Prague, pp. 173-176.

بی‌جن‌خان، محمود. " نقش پیکره‌های زبانی در نوشتن دستور زبان: معرفی یک نرم‌افزار رایانه ای." مجله زبان‌شناسی، ۱۳۸۳، سال نوزدهم، شماره دوم: ۴۸-۶۷، تهران: مرکز نشر دانشگاهی.

روح‌پرور، رحیمه. بی‌جن‌خان، محمود. "به کارگیری یک نظام برچسب‌دهی برای تعبیر و تفسیر یک پیکره متنی زبان فارسی". مجموعه مقالات هفتمین همایش زبانشناسی ایران؛ ۱۳۸۶، جلد دوم.

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-02]



پـــروین ســـادات فـــیض‌آبـــادی
فارغ التحصیل رشتهٔ مهندسی کـامپیوتر از دانشگاه علـم و صـنعت ایـران و دارای مـــدرک کارشناســـی ارشـــد در رشـــتهٔ زبان‌شناسی همگـانی از دانشـگاه تهـران می‌باشد. وی هم اکنون دانشـجوی مقطـع دکتـرای دانشـگاه هایدلبرگ آلمان می باشد.
نشانی رایانامک ایشان عبارتست از:
p.feyzabadi@ut.ac.ir



محمود بی‌جن‌خان مدرک کارشناسی خود را در رشتهٔ ریاضی کاربردی در سال ۱۹۸۰ از دانشگاه ایالتی تگزاس -آرلینگتن- کسب کرد. وی مدارک کارشناسی ارشد و دکترای خود را به ترتیب در سال‌های ۱۳۶۹و ۱۳۷۴ در رشتهٔ زبان‌شناسی از دانشگاه تهران اخذ نمود. حوزهٔ تخصصی ایشان آواشناسی، واج‌شناسی و تولید پیکره‌های زبانی است. از وی تاکنون بیش از ۷۰ مقاله در نشریات و کنفرانس‌های داخلی و خارجی به چاپ رسیده است. ایشان هم اکنون عضو هیأت علمی گروه زبان‌شناسی و سرپرست آزمایشگاه زبان‌شناسی دانکشدهٔ ادبیات و علوم انسانی دانشگاه تهران می‌باشند.
نشانی رایانامک ایشان عبارتست از:
mbjkhan@ut.ac.ir



مرضیه مرتضوی نیـا ۲۰۱۱- تـاکنون؛ دانشـــجوی مقطـــع دکتـــری در رشـــته زبانشناسی دانشگاه مـک گیـل (McGill)، کانـادا، ۲۰۱۰-۲۰۱۱؛ دانشـجوی مقطـع دکتـری در رشـته علـوم شـناختی (Cognitive Science) دانشگاه کارلتون (Carleton)، کانـادا، ۲۰۰۸- ۲۰۱۰؛ مقطـع کارشناسـی ارشـد در رشـته زبـان‌شناسـی دانشـگاه تهـران، ۲۰۰۸-۲۰۰۴؛ مقطـع کارشناسـی رشـته زبـان و ادبیـات انگلیسی در دانشگاه شیراز.

نشانی رایانامک ایشان عبارتست از:
marzieh.mortazavinia@mail.mcgill.ca

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-02]