

## واژه‌شکن فارسی\*

سید مهدی سمائی (استادیار پژوهشگاه اطلاعات و مدارک علمی ایران)

### ۱ مقدمه

مهم‌ترین بخش پردازش زبان شامل درک متن<sup>۱</sup> و تشخیص گفتار<sup>۲</sup> است. درک متن گاهی در سطح نحو است. رایانه، در این حالت، متن را در سطح جمله و بند و کلّ متن تحلیل می‌کند. از درک مفهوم متن در ترجمه ماشینی نیز استفاده می‌شود. تحقیق در این بخش از پردازش زبان از اواخر دهه هشتاد میلادی آغاز شد.

متن معنی گسترده‌ای دارد. در مقیاسی عام، حتّی علایم راهنمایی و رانندگی را می‌توان متن خواند (Crystal 1992, p. 387). به این اعتبار هر لفظ را می‌توان متن شمرد و درک مفهوم متن را به حوزه کلمه نیز تعمیم داد. واژه‌ها یا بسیط‌اند یا ترکیبی<sup>۳</sup> یعنی یا ساده‌اند و از چند جزء معنی‌دار (تکواژ)<sup>۴</sup> ساخته نشده‌اند یا پیچیده‌اند و از چند تکواژ

\* «واژه‌شکن فارسی» طرحی بوده که به مدت چهار سال و نیم در پژوهشگاه اطلاعات و مدارک علمی ایران اجرا شده است. نرم‌افزار واژه‌شکن نخستین نرم‌افزار هوشمند صرف زبان فارسی و تجربه‌ای موفق در کار پردازش این زبان است. علاءالدین طباطبائی مشاور این طرح و مهندس مهرداد نوروزی برنامه‌نویس آن بوده است. داده‌های اولیه را مرضیه صناعتی و فریبا قطره، دانشجویان دوره دکتری زبان‌شناسی، و الهام قربانی استخراج کرده‌اند. نسترین یزدانی نیز کار قاعده‌نویسی نهایی و رمزگذاری آنها را به عهده داشته است.

1) text understanding

2) speech recognition

۳) ترکیبی در اینجا معادل compound نیست بلکه فقط در تقابل با بسیط در نظر گرفته شده است لذا شامل compound (مرکب) و complex (مشتق) می‌شود.

4) morphem

ساخته شده‌اند. تحلیل متن در سطح واژه مربوط به حوزه صرف است. واژه‌های ترکیبی را برحسب آنکه با اشتقاق یا با ترکیب ساخته شده باشند می‌توان به دو دسته اصلی تقسیم کرد: مشتق و مرکب. واژه‌های مشتق از یک پایه به اضافه وند ساخته می‌شوند؛ مانند دانشمند، بی‌باک. واژه‌های مرکب از اجتماع چند تکواژ مستقل از مقوله‌های اسم و صفت و قید و مبن فعل ساخته می‌شوند؛ مانند کتابخانه، عقب‌گرد، خانه‌نشین. واژه‌شکن فارسی برای تحلیل واژه‌های ترکیبی این زبان ساخته شده است. این نرم‌افزار واژه‌های مشتق و مرکب را تجزیه و مقوله آنها را مشخص می‌سازد و آن نخستین برنامه رایانه‌ای است که با این درجه از هوشمندی در حوزه صرف زبان فارسی عمل می‌کند. بخش‌های سه‌گانه این مقاله به ترتیب درباره «ساختمان واژه‌شکن»، «شیوه استخراج قواعد»، و «ویژگی‌های واژه‌شکن» است.

## ۲ ساختمان واژه‌شکن

واژه‌شکن دو بخش عمده دارد: بخش الفاظ و بخش قواعد.

### ۲-۱ بخش الفاظ

این بخش شامل انواع مقولات صرفی (اسم، صفت، عدد، قید، ضمیر، بن فعل) و وندها (پیشوندها و پسوندها) است.

#### ۲-۱-۱ اسم

در این بخش، ۳۴ نوع اسم با ملاک‌های صرفی و قابلیت پیوستن به وندها و شیوه ترکیب با دیگر مقولات دستوری و بن افعال تمیز داده شده‌اند (صفت و وند نیز با ملاک‌های صرفی از هم تمیز داده شده‌اند): اسم، اسم جاندار، اسم معنی، اسم عمل، اسم زمان، اسم خویش، اسم مصدر، اسم مرکب و جز آن.

#### ۲-۱-۲ صفت

در این قسمت ۱۲ نوع صفت به شرح زیر تشخیص داده شده است: صفت، صفت مفعولی، صفت مفعولی منفی، صفت مبالغه، صفت مبالغه منفی، صفت مرکب، صفت

بن ساختنی، صفت ترتیبی، صفت اشتقاقی، صفت مرکب اشتقاقی، صفت بن ساختنی  
اشتقاقی، صفت قید اشتقاقی...

### ۲-۱-۳ بن فعل

بن فعل شامل بن مضارع و بن ماضی افعال است. آن دسته از بن‌های مضارع و ماضی که  
تبدیل به اسم شده‌اند یا ابتدا اسم بوده‌اند و به صورت مصدر جعلی درآمده‌اند در هر دو  
فهرست «اسم» و «بن فعل» درج شده‌اند.

### ۲-۱-۴ قید و ضمیر

در این فهرست دو نوع قید (قید، قید جهت) و یک نوع ضمیر وجود دارد.

### ۲-۱-۵ عدد

در این فهرست، اعداد یک تا بیست همچنین سی تا صد (به صورت متناوب ده‌تایی)  
آمده و دیگر اعداد، چون ساخت قیاسی و ترکیبی دارند و وارد حوزه نحو می‌شوند،  
درج نشده‌اند.

### ۲-۱-۶ وند

شامل ۶ پیشوند و ۳۸ پسوند به شرح جدول‌های زیر است.

#### پیشوند

|            |    |
|------------|----|
| منفی ساز ۱ | ن  |
| منفی ساز ۲ | نا |
| اشتراک     | هم |

|          |    |
|----------|----|
| مبالغه   | ب  |
| دارندگی  | با |
| منفی ساز | بی |

پسوند

|           |     |
|-----------|-----|
| دارندگی   | گین |
| ترتیب     | م   |
| دارندگی ۱ | مند |
| ترتیب ۱   | مین |
| دارندگی ۲ | ناک |
| فاعلی ۱   | نده |
| شبهات ۴   | واز |
| دارندگی ۳ | ور  |
| شبهات ۵   | وش  |
| شبهات ۶   | ه   |
| مفعولی    | ه   |
| نسبت ۱    | ی   |
| جنس       | ین  |

|                      |      |
|----------------------|------|
| شبهات                | آسا  |
| آغستگی               | آگین |
| اسم معنی اشتقاقی ساز | ا    |
| فاعلی                | ا    |
| عمل                  | ار   |
| نسبت                 | انه  |
| مبالغه               | باز  |
| تقلید                | بازی |
| محافظت               | بان  |
| عمل ۱                | بندی |
| تصغیر                | چه   |
| ظرف                  | دان  |
| مکان                 | زار  |
| شبهات ۱              | سان  |
| مکان ۱               | ستان |
| مصدر ساز             | ش    |
| شبهات ۲              | فام  |
| تصغیر ۱              | ک    |
| کنندگی               | کار  |
| مکان ۲               | کده  |
| مکان ۳               | گاه  |
| مجموعه               | گان  |
| کنندگی ۱             | گر   |
| عمل ۲                | گری  |
| شبهات ۳              | گون  |

### ۳ قواعد

در فهرست قواعد انواع امکانات هم‌نشینی تکرارها و وندها با یکدیگر به صورت قاعده آمده است. در اینجا فهرست برخی قواعد به اختصار ذکر می‌شود.

#### واژه‌های مرکب

|                                                    |
|----------------------------------------------------|
| اسم + صفت = اسم جاندار مرکب                        |
| ..                                                 |
| اسم اشتقاقی + صفت = اسم جاندار                     |
| ..                                                 |
| اسم + اسم معنی = اسم مرکب                          |
| ..                                                 |
| اسم اشتقاقی + اسم معنی = اسم مرکب اشتقاقی          |
| ..                                                 |
| اسم معنی + اسم = اسم مرکب                          |
| ..                                                 |
| اسم + اسم زمان = اسم مرکب                          |
| ..                                                 |
| اسم اشتقاقی + اسم زمان = اسم مرکب اشتقاقی          |
| ..                                                 |
| اسم زمان + اسم = اسم مرکب                          |
| ..                                                 |
| اسم + اسم جاندار = اسم جاندار مرکب                 |
| ..                                                 |
| اسم اشتقاقی + اسم جاندار = اسم جاندار مرکب اشتقاقی |
| ..                                                 |
| اسم جاندار + اسم = اسم جاندار مرکب                 |
| ..                                                 |

|                                                           |
|-----------------------------------------------------------|
| اسم اشتقاقی + اسم = اسم جاندار مرکب اشتقاقی               |
| اسم عمل + اسم = اسم مرکب                                  |
| اسم عمل اشتقاقی + اسم = اسم مرکب اشتقاقی                  |
| اسم + اسم = اسم مرکب                                      |
| اسم اشتقاقی + اسم = اسم مرکب اشتقاقی                      |
| اسم جاندار + اسم جاندار = اسم جاندار مرکب                 |
| اسم جاندار اشتقاقی + اسم جاندار = اسم جاندار مرکب اشتقاقی |
| اسم جاندار + اسم عمل = اسم جاندار مرکب                    |
| اسم جاندار اشتقاقی + اسم عمل = اسم جاندار مرکب اشتقاقی    |
| اسم جاندار + اسم معنی = اسم جاندار مرکب                   |
| اسم جاندار اشتقاقی + اسم معنی = اسم جاندار مرکب اشتقاقی   |

### واژه‌های بن ساختی

|                                               |
|-----------------------------------------------|
| اسم + بن مضارع = اسم جاندار بن ساختی          |
| اسم اشتقاقی + بن مضارع = اسم بن ساختی اشتقاقی |

|                                                             |
|-------------------------------------------------------------|
| اسم مرکب + بن مضارع = اسم مرکب بن ساختی                     |
|                                                             |
| اسم جاندار + بن مضارع = اسم جاندار بن ساختی                 |
|                                                             |
| اسم جاندار اشتقاقی + بن مضارع = اسم جاندار بن ساختی اشتقاقی |
|                                                             |
| اسم جاندار مرکب + بن مضارع = اسم جاندار مرکب بن ساختی       |
|                                                             |
| اسم عمل + بن مضارع = اسم جاندار بن ساختی                    |
|                                                             |
| اسم عمل اشتقاقی + بن مضارع = اسم جاندار بن ساختی اشتقاقی    |
|                                                             |
| صفت + بن مضارع = اسم جاندار بن ساختی                        |
|                                                             |
| صفت اشتقاقی + بن مضارع = اسم جاندار بن ساختی اشتقاقی        |
|                                                             |
| صفت مرکب + بن مضارع = اسم جاندار مرکب بن ساختی              |
|                                                             |
| اسم معنی + بن مضارع = اسم جاندار بن ساختی                   |
|                                                             |
| اسم معنی اشتقاقی + بن مضارع = اسم جاندار بن ساختی اشتقاقی   |
|                                                             |
| اسم خویش + بن مضارع = اسم جاندار بن ساختی                   |
|                                                             |
| اسم خویش مرکب + بن مضارع = اسم خویش مرکب بن ساختی           |
|                                                             |

## واژه‌های اشتقاقی

|                                                                |
|----------------------------------------------------------------|
| اسم + وند شباهت = صفت اشتقاقی                                  |
| اسم مرکب + وند شباهت = صفت مرکب اشتقاقی                        |
| اسم جاندار + وند شباهت = صفت اشتقاقی                           |
| اسم جاندار مرکب + وند شباهت = صفت مرکب اشتقاقی                 |
| اسم جاندار بن‌ساختی + وند شباهت = صفت بن‌ساختی اشتقاقی         |
|                                                                |
| اسم + وند آغشتگی = صفت اشتقاقی                                 |
| اسم معنی + وند آغشتگی = صفت اشتقاقی                            |
|                                                                |
| صفت + وند اسم معنی اشتقاقی‌ساز                                 |
|                                                                |
| اسم جاندار + وند نسبت = صفت اشتقاقی                            |
| اسم جاندار مرکب + وند نسبت = صفت مرکب اشتقاقی                  |
| اسم جاندار بن‌ساختی + وند نسبت = صفت بن‌ساختی اشتقاقی          |
| اسم جاندار اشتقاقی + وند نسبت = صفت فید اشتقاقی                |
| اسم خویش + وند نسبت = صفت اشتقاقی                              |
| اسم زمان + وند نسبت = صفت اشتقاقی                              |
|                                                                |
| اسم + وند مبالغه = اسم جاندار اشتقاقی                          |
| اسم مرکب + وند مبالغه = اسم جاندار مرکب اشتقاقی                |
| اسم جاندار + وند مبالغه = اسم جاندار اشتقاقی                   |
| اسم جاندار مرکب + وند مبالغه = اسم جاندار مرکب اشتقاقی         |
| اسم جاندار بن‌ساختی + وند مبالغه = اسم جاندار بن‌ساختی اشتقاقی |
| اسم عمل + وند مبالغه = اسم جاندار اشتقاقی                      |
|                                                                |
| اسم خویش + وند مبالغه = اسم جاندار اشتقاقی                     |
|                                                                |

|                                                            |
|------------------------------------------------------------|
| اسم جاندار + وند تقلید = اسم عمل اشتقاقی                   |
| اسم جاندار مرکب + وند تقلید = اسم عمل مرکب اشتقاقی         |
| اسم جاندار بن ساختی + وند تقلید = اسم عمل بن ساختی اشتقاقی |
|                                                            |
| اسم + وند محافظت = اسم جاندار اشتقاقی                      |
| اسم مرکب + وند محافظت = اسم جاندار مرکب اشتقاقی            |
| اسم زمان + وند محافظت = اسم جاندار اشتقاقی                 |
| اسم معنی + وند محافظت = اسم جاندار اشتقاقی                 |
| اسم جاندار + وند محافظت = اسم جاندار اشتقاقی               |

#### ۴ شیوه استخراج قواعد

واژه‌شکن همانند مغز انسان عمل می‌کند یعنی، به جای تکیه کردن بر محفوظات، قواعدی حتی‌المقدور حشودایی شده و تعمیم‌پذیر را به کار می‌برد. هر تکواژ (خواه مستقل خواه مقید) برای قرار گرفتن در کنار تکواژهای دیگر محدودیت‌هایی دارد به این معنی که قبل و بعد از بعضی تکواژها می‌آید و در کنار برخی دیگر نمی‌آید. واژه‌های ترکیبی نیز، به نوبه خود، تابع محدودیت‌هایی هستند. برای دست یافتن به امکانات و محدودیت‌های هم‌نشینی تکواژها و واژه‌های ترکیبی، ابتدا این واژه‌ها (مرکب، بن ساختی، مشتق) از فرهنگ معین استخراج شدند. دلیل انتخاب فرهنگ معین این بود که ساخت‌های قدیمی و مهجور نیز در این فرهنگ وجود دارد. در مرحله بعد، فهرستی از اجزای تشکیل‌دهنده ترکیبات استخراج شده تهیه شد یعنی مقوله‌های متعارف صرفی نظیر اسم و صفت و قید و ضمیر همچنین وندها (پیشوندها و پسوندها) که در ساختمان ترکیبات به کار رفته‌اند متمایز و طبقه‌بندی شدند. داده‌های استخراج‌شده اولیه، در جریان کار، جرح و تعدیل‌های فراوانی شدند.

اجزای استخراج‌شده از واژه‌های مرکب یک به یک در بافتی که به کار رفته بودند و بافت‌های بالقوه‌ای که احتمال به کار رفتن آنها وجود دارد بررسی شدند و به هریک برجستگی داده شد. در این مرحله، معلوم گردید که مقوله‌های متعارفی که تاکنون مشخص و نام‌گذاری شده‌اند کفایت نمی‌کند و باید از مؤلفه‌های مکمل برای نامیدن

انجرای آنها استفاده کرد. در تحقیقاتی که تاکنون شده، عموماً از مقوله‌های صرفی اسم و صفت و ضمیر و قید و حرف اضافه و حرف ربط و اصوات مشخص گشته‌اند. محمودی از جمله محققانی است که به پانزده مقوله صرفی در زبان فارسی قایل شده است (Mahmoudi 1994). اما آنچه در کار واژه‌شکن اهمیت داشت شمار مقوله‌ها نبود، بلکه تقسیمات فرعی هر مقوله بود. مثلاً برای مقوله صرفی اسم بسته به نقشی که در هر واژه مرکب داشت می‌بایست مؤلفه‌ای جداگانه با برجسب جدید اختیار می‌شد. محمودی بر آن است که مقوله‌های صرفی باید بر اساس توزیع تعیین شوند (Ibid). استخراج مؤلفه‌های هر مقوله بر اساس همین نظر محمودی انجام شد. بدین ترتیب، برای هر مقوله تقسیمات معنایی قایل شدیم و بر اساس آنها مقوله‌ها به چند مقوله جدید تقسیم شدند. ۳۴ نوع اسم و ۱۲ نوع صفت حاصل همین کار بودند.

برای قاعده‌نویسی، دو کار انجام شد:

الف- قاعده هم‌نشینی مقوله‌ها به زبان طبیعی نوشته شد.

ب- قاعده‌ها با حروف لاتینی و اندیس‌های عددی به صورت رمز درآمدند تا مناسب

برنامه‌نویسی رایانه‌ای شوند.

### فهرست قراردادی قواعد

| اسم              | N01 |
|------------------|-----|
| اسم جاندار       | N02 |
| اسم معنی         | N03 |
| اسم عمل          | N04 |
| اسم زمان         | N05 |
| اسم خویش         | N06 |
| اسم مصدر         | N07 |
| اسم مرکب         | N08 |
| اسم جاندار مرکب  | N09 |
| اسم معنی اشتقاقی | N10 |

| صفت اشتقاقی          | A09 |
|----------------------|-----|
| صفت مرکب اشتقاقی     | A51 |
| صفت بن‌ساختی اشتقاقی | A52 |
| صفت قید اشتقاقی      | A54 |
| قید                  | B01 |
| قید جهت              | B02 |
| ضمیر                 | C01 |
| بن مضارع             | I01 |
| بن ماضی              | I02 |
| عدد                  | M01 |

| Formula Descriptions |                 |
|----------------------|-----------------|
| Code                 | Description     |
| A01                  | صفت             |
| A02                  | صفت مفعولی      |
| A03                  | صفت مفعولی منفی |
| A04                  | صفت مبالغه      |
| A05                  | صفت مبالغه منفی |
| A06                  | صفت مرکب        |
| A07                  | صفت بن‌ساختی    |
| A08                  | صفت ترتیبی      |

|     |                 |
|-----|-----------------|
| S25 | پسوند ترتیب ۱   |
| S26 | پسوند دارندگی ۲ |
| S27 | پسوند شباهت ۴   |
| S28 | پسوند دارندگی ۳ |
| S29 | پسوند شباهت ۵   |
| S30 | پسوند شباهت ۶   |
| S31 | پسوند نسبت ۱    |
| S32 | پسوند جنس       |
| S33 | پسوند فاعلی     |
| S34 | پسوند عمل ۲     |
| S35 | پسوند مصدر ساز  |
| S36 | پسوند فاعلی ۱   |
| S37 | پسوند مفعولی    |
| S38 | پسوند کنندگی ۱  |

|     |                            |
|-----|----------------------------|
| P04 | پیشوند منفی ساز ۱          |
| P05 | پیشوند منفی ساز ۲          |
| P06 | پیشوند اشتراک              |
| S01 | پسوند شباهت                |
| S02 | پسوند آغستگی               |
| S03 | پسوند اسم معنی اشتقاقی ساز |
| S04 | پسوند نسبت                 |
| S05 | پسوند میالنه               |
| S06 | پسوند تقلید                |
| S07 | پسوند محافظت               |
| S08 | پسوند عمل                  |
| S09 | پسوند تصغیر                |
| S10 | پسوند ظرف                  |
| S11 | پسوند مکان                 |
| S12 | پسوند شباهت ۱              |
| S13 | پسوند مکان ۱               |
| S14 | پسوند شباهت ۲              |
| S15 | پسوند تصغیر ۱              |
| S16 | پسوند مکان ۲               |
| S17 | پسوند مکان ۳               |
| S18 | پسوند مجموعه               |
| S19 | پسوند کنندگی               |
| S20 | پسوند عمل ۱                |
| S21 | پسوند شباهت ۳              |
| S22 | پسوند دارندگی              |
| S23 | پسوند ترتیب                |
| S24 | پسوند دارندگی ۱            |

|     |                             |
|-----|-----------------------------|
| N11 | اسم بن ساختی                |
| N12 | اسم جاندار بن ساختی         |
| N13 | اسم خویش مرکب               |
| N14 | اسم جاندار اشتقاقی          |
| N15 | اسم عمل اشتقاقی             |
| N16 | اسم اشتقاقی                 |
| N17 | اسم خویش اشتقاقی            |
| N18 | اسم ظرف اشتقاقی             |
| N19 | اسم مکان اشتقاقی            |
| N20 | اسم مجموعه                  |
| N21 | اسم زمان اشتقاقی            |
| N51 | اسم مرکب اشتقاقی            |
| N52 | اسم بن ساختی اشتقاقی        |
| N53 | اسم جاندار مرکب اشتقاقی     |
| N54 | اسم جاندار بن ساختی اشتقاقی |
| N55 | اسم عمل مرکب اشتقاقی        |
| N56 | اسم عمل بن ساختی اشتقاقی    |
| N59 | اسم مکان مرکب اشتقاقی       |
| N60 | اسم مکان بن ساختی اشتقاقی   |
| N64 | اسم مرکب بن ساختی           |
| N65 | اسم جاندار مرکب بن ساختی    |
| N66 | اسم خویش مرکب بن ساختی      |
| N67 | اسم معنی مرکب اشتقاقی       |
| N68 | اسم معنی بن ساختی اشتقاقی   |
| P01 | پیشوند میالنه               |
| P02 | پیشوند دارندگی              |
| P03 | پیشوند منفی ساز             |

| Formula |               |
|---------|---------------|
| Formula | Small Formula |
| N01S08  | N15           |
| N01S09  | N16           |
| N01S10  | N18           |
| N01S11  | N19           |
| N01S12  | A09           |
| N01S13  | N19           |
| N01S14  | A09           |
| N01S15  | N16           |
| N01S16  | N19           |
| N01S17  | N19           |
| N01S18  | N20           |
| N01S19  | N14           |
| N01S21  | A09           |
| N01S26  | A09           |
| N01S27  | A09           |
| N01S28  | N14           |
| N01S29  | A09           |
| N01S30  | N16           |
| N01S31  | A09           |
| N01S32  | A09           |
| N01S38  | N14           |
| N02I01  | N12           |
| N02N01  | N09           |
| N02N02  | N09           |
| N02N03  | N09           |

| Formula |               |
|---------|---------------|
| Formula | Small Formula |
| B02I01  | N11           |
| B02I02  | N11           |
| B02N01  | N08           |
| B02N01  | N08           |
| B02N04  | N08           |
| B02S17  | N19           |
| C01A02  | N09           |
| C01I01  | N12           |
| I01S35  | N07           |
| I01S36  | N14           |
| I02S37  | A02           |
| M01S23  | A08           |
| M01S25  | A08           |
| N01A01  | N09           |
| N01I01  | N12           |
| N01I02  | N11           |
| N01N01  | N08           |
| N01N02  | N09           |
| N01N03  | N08           |
| N01N04  | N08           |
| N01N05  | N08           |
| N01S01  | A09           |
| N01S02  | A09           |
| N01S05  | N14           |
| N01S07  | N14           |

| Formula |               |
|---------|---------------|
| Formula | Small Formula |
| A01I01  | N12           |
| A01I02  | A07           |
| A01N01  | N09           |
| A01N02  | N09           |
| A01N03  | N09           |
| A01N04  | N09           |
| A01N05  | N09           |
| A01N07  | N09           |
| A01S03  | N10           |
| A01S14  | A09           |
| A01S17  | N19           |
| A01S26  | A09           |
| A01S31  | N10           |
| A01S38  | N14           |
| A06I01  | N65           |
| A06I02  | N53           |
| A06S17  | N59           |
| A06S17  | N60           |
| A06S31  | N67           |
| A07S31  | N68           |
| A09I01  | N54           |
| A09I02  | A52           |
| B01I01  | N12           |
| B01N01  | N08           |
| B01S17  | N19           |

| Formula |               |
|---------|---------------|
| Formula | Small Formula |
| N05S14  | A09           |
| N05S17  | N2i           |
| N05S21  | A09           |
| N05S29  | A09           |
| N06I01  | N12           |
| N06N06  | N13           |
| N06S04  | A09           |
| N06S05  | N14           |
| N06S09  | N17           |
| N06S15  | N17           |
| N06S27  | A09           |
| N06S31  | N10           |
| N07I01  | N12           |
| N07S17  | N19           |
| N07S19  | N14           |
| N08I01  | N64           |
| N08I02  | N64           |
| N08S01  | A51           |
| N08S02  | A51           |
| N08S05  | N53           |
| N08S07  | N53           |
| N08S08  | N55           |
| N08S11  | N59           |
| N08S12  | A51           |
| N08S13  | N59           |

| Formula |               |
|---------|---------------|
| Formula | Small Formula |
| N03S07  | N14           |
| N03S10  | N18           |
| N03S13  | N19           |
| N03S16  | N19           |
| N03S17  | N19           |
| N03S18  | N20           |
| N03S19  | N14           |
| N03S22  | N14           |
| N03S24  | N14           |
| N03S26  | A09           |
| N03S28  | N14           |
| N03S31  | A09           |
| N04I01  | N12           |
| N04N01  | N08           |
| N04S05  | N14           |
| N04S12  | A09           |
| N04S13  | N19           |
| N04S17  | N19           |
| N04S19  | N14           |
| N04S20  | N15           |
| N05I01  | N12           |
| N05N01  | N08           |
| N05N02  | N09           |
| N05S04  | A09           |
| N05S07  | N14           |

| Formula |               |
|---------|---------------|
| Formula | Small Formula |
| N02N04  | N09           |
| N02N09  | N14           |
| N02S01  | A09           |
| N02S04  | A09           |
| N02S05  | N14           |
| N02S06  | N15           |
| N02S07  | N14           |
| N02S09  | N14           |
| N02S10  | N18           |
| N02S12  | A09           |
| N02S13  | N19           |
| N02S14  | A09           |
| N02S15  | N14           |
| N02S16  | N19           |
| N02S17  | N19           |
| N02S18  | N20           |
| N02S20  | N15           |
| N02S21  | A09           |
| N02S27  | A09           |
| N02S29  | A09           |
| N02S31  | N10           |
| N03I01  | N12           |
| N03I02  | N12           |
| N03N01  | N08           |
| N03S02  | A09           |

| Formula |               |
|---------|---------------|
| Formula | Small Formula |
| N14N04  | N53           |
| N14S04  | A54           |
| N15I01  | N54           |
| N15N01  | N51           |
| N16A01  | N53           |
| N16I01  | N52           |
| N16I02  | N52           |
| N16N01  | N53           |
| N16N01  | N51           |
| N16N02  | N53           |
| N16N03  | N51           |
| N16N04  | N51           |
| N16N05  | N51           |
| P01I01  | A04           |
| P02N03  | N14           |
| P03N03  | N14           |
| P04A02  | A03           |
| P04I01  | A05           |
| P05A02  | A03           |
| P05N02  | N14           |
| P06N01  | N14           |

| Formula |               |
|---------|---------------|
| Formula | Small Formula |
| N10I01  | N54           |
| N11S16  | N60           |
| N11S17  | N60           |
| N11S26  | A52           |
| N11S27  | A52           |
| N11S28  | N54           |
| N11S29  | A52           |
| N11S31  | A52           |
| N12S01  | A51           |
| N12S04  | A52           |
| N12S05  | N54           |
| N12S06  | N56           |
| N12S12  | A52           |
| N12S13  | N60           |
| N12S16  | N60           |
| N12S17  | N60           |
| N12S21  | A52           |
| N12S27  | A52           |
| N12S29  | A52           |
| N12S31  | N68           |
| N13I01  | N66           |
| N14I01  | N54           |
| N14I02  | N54           |
| N14N02  | N53           |
| N14N03  | N53           |

| Formula |               |
|---------|---------------|
| Formula | Small Formula |
| N08S16  | N59           |
| N08S17  | N59           |
| N08S19  | N53           |
| N08S21  | A51           |
| N08S26  | A01           |
| N08S27  | A51           |
| N08S28  | N53           |
| N08S29  | A51           |
| N08S31  | A01           |
| N08S38  | N53           |
| N09I01  | N65           |
| N09I02  | N65           |
| N09S01  | A51           |
| N09S04  | A51           |
| N09S05  | N53           |
| N09S06  | N55           |
| N09S12  | A51           |
| N09S13  | N59           |
| N09S16  | N59           |
| N09S17  | N59           |
| N09S20  | N55           |
| N09S21  | A51           |
| N09S27  | A51           |
| N09S29  | A51           |
| N09S31  | N67           |

## ۵ ویژگی‌های واژه‌شکن

یکی از مشکلات نظری در تهیه واژه‌شکن مسئله دستوری و غیردستوری بودن ساخت‌ها بود. غرض از ساختن واژه‌شکن آن بود که نرم‌افزار مفروض بتواند ساخت‌های دستوری را از ساخت‌های غیردستوری تمیز دهد و سپس الگوی ساخت آن را عرضه کند.

سؤال این بود که، اگر قرار باشد ساختی را غیردستوری بدانیم، معیار ما چه خواهد بود. چامسکی، در ساخت‌های نحوی، آنجا که درباره استقلال دستور زبان بحث می‌کند به این نکته می‌پردازد که چگونه جمله دستوری را از جمله غیردستوری تمیز دهیم (چامسکی، ص ۱۴-۲۰). وی در پاسخ این پرسش که رشته‌های دستوری و غیردستوری را در عمل بر چه پایه‌هایی از یکدیگر بازمی‌شناسیم؟ می‌نویسد:

پاسخ‌های چندی که بلافاصله به ذهن خطور می‌کنند نمی‌توانند درشت باشند ... دستور زبان پرتو رفتار گوینده است که بر پایه تجزیه محدود و تصادفی خود از زبان می‌تواند تعداد نامحدودی جمله تازه تولید کند ... در حقیقت هر توضیحی درباره مفهوم دستوری بودن در زبان را می‌توان به دست دادن توضیح این جنبه اساسی از رفتار زبانی شمرد. (همان، ص ۱۵)

به نظر چامسکی «مفهوم دستوری بودن را با مفهوم بهره‌ور بودن یا معنی‌دار بودن ... نمی‌توان همسان دانست» (همان‌جا). او جستجو «برای یافتن تعریف دستوری بودن بر پایه معنی‌شناسی» را کاری عبث می‌داند (همان‌جا). از طرفی، دستوری بودن را تابع بسامد و «از لحاظ تخمین آماری در حد بالا بودن» نیز نمی‌داند. (همان، ص ۱۶)<sup>۵</sup>

سه نکته مهم در نقل قول‌های مذکور وجود دارد:

الف- رفتار زبانی یکی از ملاک‌های تمیز جمله دستوری از غیردستوری است؛

ب- دستوری بودن را می‌توان فارغ از معنی بررسی کرد؛

ج- بسامد وقوع تأثیری در دستوری یا غیردستوری بودن ندارد.

این بحث هرچند در حوزه نحو صورت گرفته، می‌توان آن را به حوزه صرف تعمیم داد و در معنی واژه‌های مرکب تا حدی با تسامح برخورد کرد و غیرمتعارف بودن این

(۵) نگارنده، در رساله دکتری خود، در پی طراحی ماشین مفروضی بوده که جمله‌های دستوری را بدون توسل به معنی از جمله‌های غیردستوری تمیز دهد یعنی بر مبنای نظر چامسکی در ساخت‌های نحوی.

واژه‌ها را از حیث معنی ملاک، غیر دستوری بودن آنها ندانست.

واژه‌شکن با این دیدگاه ساخته شده است. یکی از تبعات برگزیدن چنین دیدگاهی آن است که بسیاری از ساخت‌های بالقوه را که خوش‌ساخت‌اند<sup>۶</sup> اما چه‌بسا تاکنون کسی به کار نبرده باشد می‌توان دستوری دانست. دستوری دانستن ترکیباتی نظیر جوب‌نشین و گرفتارانه که به قیاس با خانه‌نشین و دانشمندانه ساخته شده‌اند بر همین اساس است.

بی‌نظمی موجود در ساخت‌های ترکیبی به خصوص اشتقاقی نیز دلیل موجهی برای دستوری شمردن ترکیبات بالقوه‌ای است که معنی غیرمتعارف دارند مثلاً پسوند -ا را می‌توان به صفت‌های پهن و دراز اضافه کرد و پهن و دراز را ساخت. چنانچه این قاعده را تعمیم ندهیم و ساخت‌هایی مانند نازکا و سیختا و نرم را دستوری ندانیم ناچار خواهیم شد به استثنای زیادی قایل شویم که خلاف نظریه‌های یادگیری زبان است. در واقع، مجبور خواهیم شد همه قاعده‌ها را تعمیم ناپذیر فرض کنیم و ترکیبات موجودی را که با این قاعده ساخته شده‌اند مستقیماً در حافظه رایانه قرار دهیم، هرچند در مواردی چاره‌ای جز قرار دادن مستقیم ترکیبات در حافظه نداشتیم. مثلاً ترکیباتی که با این ماضی و پسوند -ار ساخته می‌شوند چنان از لحاظ معنایی ناهمگون‌اند که نمی‌توان برای آنها قاعده عام داد و باید آنها را مستقیم در حافظه گذاشت. یکی دیگر از مشکلات نظری این بود که آیا قواعد باید دوری<sup>۷</sup> باشد یا خیر. قواعدی را که می‌توانند تکرار شوند و جمله بسازند قواعد دوری یا مکرر<sup>۸</sup> می‌نامند (Crystal 1992, p. 328). مثلاً قاعده قرار گرفتن صفت بعد از اسم در زبان فارسی قاعده‌ای دوری است زیرا منطقاً می‌توان گفت: این کتاب قطور خواندنی، جذاب سبز ... این مسئله در کار واژه‌شکن که در حوزه صرف عمل می‌کند نیز مطرح می‌شود. مثلاً در حوزه ترکیب اشتقاقی اسم + وند تصغیر (کتاب + ک) ← اسم (کتابک)، که کتابک خود اسم است و منطقاً بار دیگر با پسوند -ک ترکیب می‌شود و کتابکک را. می‌سازد و الی آخر. یا اسم جاندار (کبوتر) + وند مبالغه (باز) ← اسم جاندار، که کبوتر باز خود اسم جاندار است و منطقاً بار دیگر با پسوند -باز می‌تواند ترکیب شود و کبوتر باز باز را بسازد و پس علی‌هذا، در زبان عادی، کاربرد مانع از دوری شدن ساخت‌ها می‌شود. در کار واژه‌شکن، مشکل به این صورت حل شد که برای کتابک و کبوتر باز نامی به جز اسم و اسم

6) well-formed

7) recursive

8) iterative

جاندار یعنی اسم تصغیر اشتقاقی و اسم جاندار مرکب اختیار شد و قاعده دیگری برای ترکیب، این ساخت‌های جدید با پسوند تک و -باز داده نشد تا مانع از دَوّری شدن قواعد شود. این مشکل در خوزه واژه‌های مرکب نیز وجود داشت؛ مثلاً اسم + اسم ← اسم مانند آب طلا، که در این صورت آب طلا نیز به نوبه خود می‌تواند دوباره در ترکیب شرکت کند و آب طلا را بسازد و به همین ترتیب ادامه پیدا کند. اما در این مورد هم قاعده به صورت اسم + اسم ← اسم مرکب در آمد.

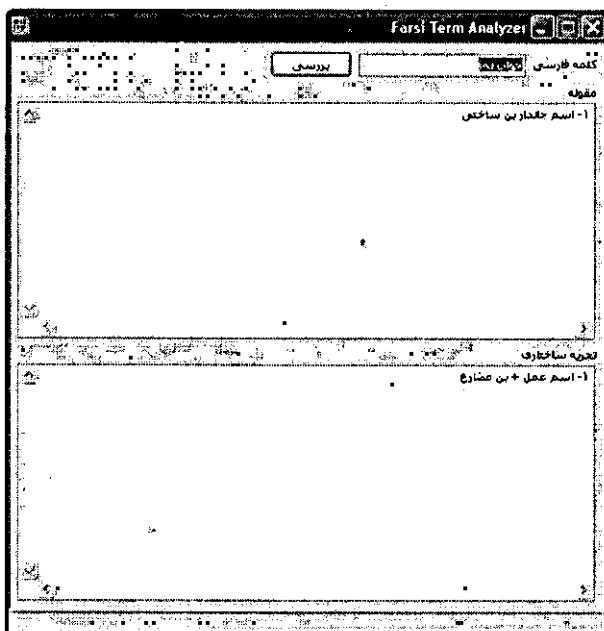
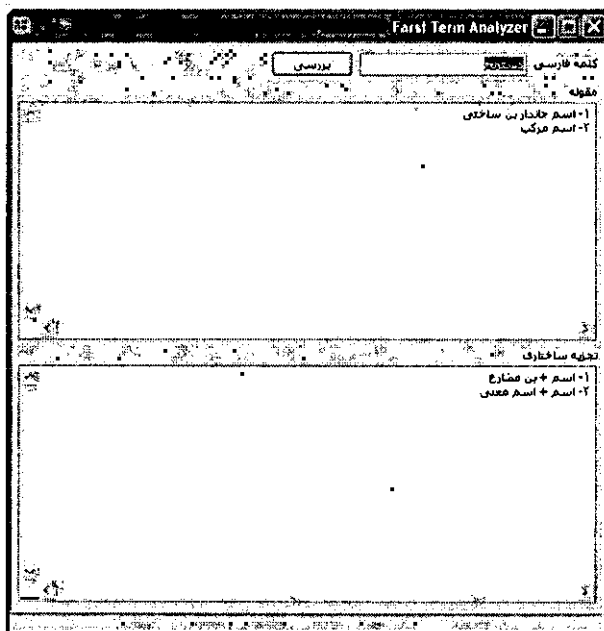
قابل شدن به دَوّری بودن قواعد در کار واژه‌شکن باعث ساخته شدن ترکیباتی دستوری می‌شود. اما، از آنجایی که این ترکیبات انتهایی ندارند و کارایی واژه‌شکن را مختل می‌سازند، این ویژگی مهار شده است از این راه که قواعد تکرارپذیر نباشند و در مرحله‌ای متوقف شوند. یکی از ویژگی‌های این نرم‌افزار آن است که پایگاه داده (شامل قواعد و فهرست کلمات) ویرایش‌پذیر است یعنی می‌توان اطلاعات جدیدی به آن وارد ساخت یا داده‌ای را حذف یا اصلاح کرد.

#### ۵-۱ شیوه تجزیه ترکیبات و تعیین مقوله دستوری

واژه‌شکن ابتدا مقوله واژه پیشنهادی را مشخص سپس آن را به اجزای سازنده‌اش تجزیه می‌کند. در صورتی که قاعده ساخت واژه پیشنهادی در حافظه رایانه موجود نباشد، فقط به تجزیه آن اکتفا می‌شود. این جمله مشروط به آن است که عناصر سازنده واژه مفروض در حافظه موجود باشد و، اگر موجود نباشد، می‌توان آن را به فهرست افزود سپس آن را برای تحلیل به واژه‌شکن سپرد. اینک دو مثال (دسترنج، خواب‌نما) برای نشان دادن کار واژه‌شکن عرضه می‌شود. (→ تصویر صفحه بعد).

#### توضیح

واژه‌شکن هر واژه ترکیبی را در دو خانه تحلیل می‌کند: خانه بالا و خانه پایین. در ردیف اول خانه بالا، مقوله دستوری و، در ردیف اول خانه پایین، شکل تجزیه‌شده آن را می‌دهد. مثلاً خانه‌نشین اسم جاندار بن ساختی است، یعنی از اسم و بن مضارع ساخته شده است که در بالا مقوله دستوری آن (اسم جاندار بن ساختی) و در پایین شکل



تجزیه شده (اسم + بن مضارع) می‌آید. واژه‌شکن البته در اغلب داده‌ها به این اکتفا نمی‌کند و تحلیل آن را نیز ارائه می‌دهد. در این حالت، فقط ردیف اول خانه بالا و ردیف اول خانه پایین معتبرند و باقی تحلیل‌ها مورد نظر نیستند. گاهی واژه‌ای دو مقوله دستوری دارد که هر دو معتبرند. در چنین حالتی، اطلاعات هر دو ردیف مربوط به واژه در خانه بالا و پایین معتبرند. البته تحلیل‌های اضافی واژه‌شکن نیز بی‌فایده نیستند و قابلیت نرم‌افزار را نشان می‌دهند. تحلیل‌های اضافی با علامت ستاره مشخص شده‌اند.

واژه‌شکن واژه دسترنج را اسم جاندار بن ساختی می‌داند؛ زیرا این ترکیب شامل اسم (دست) و بن مضارع (رنج) است یعنی، طبق قاعده و به قیاس با زورگیر و خانه‌نشین، می‌توان دسترنج را اسم جاندار دانست. واژه‌شکن در مرحله بعد دسترنج را مشمول قاعده اسم + اسم معنی می‌داند و آن را اسم مرکب می‌نامد.

صورت‌هایی که هم اسم و هم بن فعل‌اند با هر دو ویژگی در فهرست قواعد آمده‌اند. اما، چون قاعده بن مضارع + بن مضارع معتبر نیست، واژه‌شکن خواب‌نما را مشمول قاعده اسم عمل + بن مضارع می‌داند و آن را اسم جاندار می‌گیرد.

## منابع

- چامسکی، نعام، ساخت‌های نحوی، ترجمه احمد سمیعی، انتشارات خوارزمی، تهران ۱۳۶۲.
- سامعی، حسین (۱۳۷۵)، واژه‌سازی در زبان فارسی، رساله دکتری زبان‌شناسی، دانشگاه تهران، تهران.
- سمائی، سیدمهدی (۱۳۷۷)، واژگان در دستورسنج، رساله دکتری زبان‌شناسی، دانشگاه تهران، تهران.
- صادقی، علی‌اشرف (۱۳۷۰-۱۳۷۲)، «شیوه‌ها و امکانات واژه‌سازی در زبان فارسی معاصر»، نشر دانش، سال ۱۱، شماره ۴، ص ۱۸-۱۲ و شماره ۵، ص ۱۲-۶؛ سال ۱۲، شماره ۳، ص ۲۱-۲۵ و شماره ۴، ص ۳۹-۴۵ و شماره ۵، ص ۱۵-۱۹ و شماره ۶، ص ۱۹-۲۳؛ سال ۱۳، شماره ۳، ص ۹-۱۵ و شماره ۴، ص ۱۵-۲۳ و شماره ۵، ص ۲۱-۲۵.
- طباطبائی، علاءالدین (۱۳۷۶)، فعل بسیط فارسی و واژه‌سازی، مرکز نشر دانشگاهی، تهران.
- \_\_\_\_\_ (۱۳۸۲)، اسم و صفت مرکب در زبان فارسی، مرکز نشر دانشگاهی، تهران.
- فلاح رستگار، گیتی (۱۳۵۰)، «شیوه ترکیب در زبان فارسی»، دومین کنگره تحقیقات ایرانی، ج ۱، به کوشش حمید زرین‌کوب، انتشارات دانشگاه مشهد، ص ۲۵۰-۲۶۳.
- کانن، اس.پی.ان و دیگران، «تهیه بانک پر حجم واژگان برای کاربرد در نظام‌های متن‌سازی و سازه‌یابی و بازیابی اطلاعات»، ترجمه سیدمهدی سمائی، نشر اطلاع‌رسانی، دوره ۱۲، شماره ۲، ص ۴۷-۷۳.
- کشانی، خسرو (۱۳۶۷)، «پسوندهای بازی»، مجله زبان‌شناسی، سال ۵، شماره ۲، ص ۶۵-۶۹.
- \_\_\_\_\_ (۱۳۷۹)، اشتقاق پسوندی در زبان فارسی امروز، مرکز نشر دانشگاهی، تهران.

کلباسی، ایران (۱۳۶۹)، «نقش‌های 'مم' در زبان فارسی»، مجلهٔ زبان‌شناسی، سال ۷، شمارهٔ ۲، ص ۵۶-۵۸.  
\_\_\_\_\_ (۱۳۷۱)، ساخت اشتقاقی واژه در فارسی امروز، مؤسسهٔ مطالعات و تحقیقات فرهنگی، تهران.  
وحیدیان کامیار، تقی (۱۳۷۱)، «بررسی صفت مفعولی و اهمیت آن در زبان فارسی»، مجلهٔ زبان‌شناسی،  
سال ۹، شمارهٔ ۲، ص ۶۱.

ANDERSON, S.R. (1988), "Morphological Theory", in an MT framework, *ASLIB* processing, vol. 47, N.3, pp. 73-82.

CRYSTAL, D. (1992), *An Encyclopedic Dictionary of Language and Languages*, Blackwell, Oxford.

KATZ, J. and J. Fodor (1963), "The Structure of a Semantic Theory", *Language*, 39, pp. 170-210.

KATZ, J. and P. Postal (1964), *An Integrated Theory of Linguistic Descriptions*, MIT Press, Cambridge.

KING, M. (1992), "Machine Translation", *International Encyclopedia of Linguistics*, vol.II, W. Briqht. (editor in chief), pp. 365-367.

MAHMOUDI, S.M. (1994), *Contribution au traitement automatique de la langue persane*, thèse de doctorat, Lyon II.

SPENCER, A. (1991), "Morphological Theory: An Introduction to Morphology", *Generative Grammar*, Blackwell, Oxford.

