

تاریخ ارسال: ۱۴۰۱/۶/۱۹

تاریخ پذیرش: ۱۴۰۲/۷/۱۰

doi 10.22034/nf.2023.361515.1212

## فرهنگ بسامدی فارسی؛ اثری نیازمند ویرایش یا تألیف دوباره؟

غلامرضا مستعلی پارسا\* (دانشیار گروه زبان و ادبیات فارسی، دانشکده ادبیات فارسی و زبان‌های خارجی، دانشگاه علامه طباطبائی، تهران، ایران)

سیدحمید حسینی (دانشجوی دکتری گروه زبان و ادبیات فارسی، دانشکده ادبیات فارسی و زبان‌های خارجی، دانشگاه علامه طباطبائی، تهران، ایران)

**چکیده:** مجموعه «فرهنگ‌های بسامدی راتلج»، تا به امروز، دست‌کم در پانزده عنوان منتشر شده و جمعاً مشتمل بر چهارده زبان است. این مجموعه، که از سوی انتشارات راتلج (گروه تیلور و فرانسس) منتشر شده، به‌ویژه در میان کسانی که در زمینه تدوین «فرهنگ‌های بسامدی» و/یا «زبان‌شناسی پیکره‌ای» پژوهش می‌کنند، معروفیت جهانی یافته است. هر یک از کتاب‌های این مجموعه در قالب فرهنگی بسامدی و بر پایه پیکره‌های زبانی غالباً مستقل تدوین شده است. بیشتر فرهنگ‌ها هرکدام ۵۰۰۰ مدخل اصلی/سرمدخل/بن‌واژه دارند و بخش اعظم متن هر کتاب را فهرست این واژه‌ها — به ترتیب بسامد و از پرسامدترین به کم‌بسامدترین واژه — تشکیل می‌دهد. مبنای چینش واژه‌ها اساساً بسامد خام یا بسامد مطلق نیست، بلکه شاخص پراکنش هر یک از واژه‌ها نیز مد نظر بوده است. بلافاصله پس از فهرست بسامدی، فهرست الفبایی واژه‌ها — به ترتیب الفبای زبان هر فرهنگ — آمده است. فهرست‌های واژگانی موضوعی (در جعبه‌های مستقل و در لابه‌لای صفحات بخش اصلی اثر، یعنی فهرست بسامدی) و نمایه هویت‌های دستوری/اجزای سخن/پاره‌های گفتار نیز بعد از آن قرار گرفته‌اند. پیش از بخش اصلی، یعنی فهرست بسامدی مدخل‌ها، عموماً پیش‌گفتاری به قلم مارک دیویس (دبیر مجموعه)، سپس مقدمه‌ای به قلم مؤلفان فرهنگ (به‌علاوه منابع مقدمه)، و چند فقره دیگر آمده است.

\* mastali.parsa@gmail.com

فرهنگ بسامدی فارسی در سال ۲۰۱۷ م منتشر شد و چاپ دوم آن در ۲۰۱۸ م انتشار یافت. نگارندگان، در مقاله حاضر، پس از معرفی اجمالی مجموعه، به معرفی و نقد این اثر پرداخته‌اند و نشان داده‌اند که اشکالات و معایب آن، هم از نگاه روش شناختی و هم از نظر موردی، چندین برابر محاسن آن است.

کلیدواژه‌ها: فرهنگ بسامدی، زبان‌شناسی پیکره‌ای، بسامد واژگانی، فرهنگ‌نویسی، فرهنگ لغت، فرهنگ‌های بسامدی راتلج.

### ۱- درآمد: معرفی مجموعه «فرهنگ‌های بسامدی راتلج»

هدف اصلی از تولید و انتشار مجموعه «فرهنگ‌های بسامدی راتلج»<sup>۱</sup>، به‌دست دادن منابعی است برای کسانی که به آموزش زبان می‌پردازند (نک: Davies, p. vii: پیش‌گفتار مجموعه به‌قلم تونی مک‌انری و پل ریزن<sup>۲</sup> بر کتاب؛ نیز نک: Davies & Gardner, p. vii: پیش‌گفتار مجموعه به‌قلم هم‌ایشان بر کتاب). اغلب فرهنگ‌های این مجموعه در بردارنده ۵۰۰۰ واژه پرکاربردند. این واژه‌ها عموماً بر پایه پیکره‌هایی متشکل از متن‌های امروزی یا بعضاً معاصر— از منابع نوشتاری و گفتاری— و با حجم‌های گوناگون انتخاب شده‌اند.

تا به امروز برای این زبان‌ها فرهنگ بسامدی تدوین شده، و انتشارات راتلج، در فاصله سال‌های ۲۰۰۶ تا ۲۰۲۱ م، آن‌ها را چاپ و منتشر کرده است: آلمانی، اسپانیایی، انگلیسی (آمریکایی معاصر)، پرتغالی، ترکی [استانبولی]، چکی، چینی ماندارین، روسی، ژاپنی، عربی (عمومی)، عربی (ادبیات داستانی)، فارسی، فرانسوی، گره‌ای، هلندی. فرهنگ بسامدی انگلیسی بریتانیایی نیز در دست انتشار است.

تدوین این فرهنگ‌ها عموماً بر پایه زبان دوره معاصر و/ یا روز صورت گرفته است.

### ۱-۲ ویژگی‌های مجموعه «فرهنگ‌های بسامدی راتلج»

چنان‌که اشاره کردیم، هر یک از فرهنگ‌های این مجموعه شامل ۵۰۰۰ واژه یا بن‌واژه است. هر فرهنگ یا، در اصطلاح کتاب‌داری و اطلاع‌رسانی، «تک‌نگاره»<sup>۳</sup> شامل فهرست‌های الفبایی و بسامدی (بر اساس بیشترین تا کمترین تکرار، عموماً از میان ۵۰۰۰ واژه انتخاب‌شده مؤلفان) است. در

1. Routledge Frequency Dictionaries (USA, Routledge)

بر پایه تلفظ نام این انتشارات معروف، املاي فارسی آن را به یکی از سه صورت «راتلج/ روتلج/ راوتلج» می‌توان نوشت. تلفظ اصلی رایج آن به «الفبای آوانگاری بین‌المللی» (International Phonetic Alphabet: IPA) چنین است: [ˈraʊtlɪdʒ].

2. Tony McEnery & Pay Rayson: "Series preface" (2005)

3. monograph

فهرست‌ها، در هریک از مدخل‌ها، به هویت دستوری واژه‌ها اشاره شده است. همچنین، در همه تک‌نگاره‌ها، به‌جز در فرهنگ بسامدی انگلیسی آمریکایی معاصر (نک: Davies)، مدخل‌ها با معادل یا معادل‌های رایج‌تر انگلیسی و یک شاهدمثال/شاهد مثال<sup>(۱)</sup> از کاربرد واژه در بافت زبان بر اساس خود پیکره، همراه با ترجمه انگلیسی شاهدمثال، آورده شده‌اند.

۱-۲-۱. منظور از واژه (word) عنصر یا واحدی واژگانی (lexical element/ lexical unit: LU) است که در یک فرهنگ لغت استاندارد، به‌عنوان مدخل (entry) به‌طور مستقل ثبت می‌شود. و اما مراد از بن‌واژه (lemma) اصطلاحی است که بیشتر در «ساخت‌واژه/ صرف» (morphology)، «فرهنگ‌نگاری» (lexicography)، و «زبان‌شناسی پیکره‌ای» (corpus linguistics) مطرح است، و آن صورت صرف‌نشده هر واژه است. برای مثال، «درخت» بن‌واژه است و به همه صورت‌هایی مانند «درختان»، «درخت‌ها»، «درخت‌هایی»، «درختی» (با «ی» نکره)، «درختشان»، و گاه «درخت» (خود بن‌واژه + کسره اضافه یا وصف)، و حتی خود «درخت»، اصطلاح صورت‌واژه یا صورت‌کلمه (word-form/ word form) اطلاق می‌شود. همچنین، فعل «داشتن» (صورت صرف‌نشده یا مصدر) بن‌واژه است و همه صورت‌های صرفی «دارم»، «داشتیم»، «داشته‌اند»، «داشته است»، «داشته‌اند»<sup>(۱)</sup> (ماضی نقلی، سوم شخص مفرد)، «داشته‌ام»<sup>(۲)</sup> (فعل وصفی)، «بدارند»، «ندارد»، و مانند این‌ها صورت‌واژه یا صورت‌کلمه‌های «داشتن» اند.

نکته: در ادامه مقاله، هر جا اصطلاح واژه را به‌کار برده‌ایم، عموماً منظورمان «بن‌واژه» یا «مدخل» است.

۱-۲-۲. در تدوین مجموعه «فرهنگ‌های بسامدی راتلج» عموماً از روشی کارآمد و جذاب برای بهبود توانش و کنش زبانی<sup>۴</sup> زبان‌آموزان در سطح‌های مختلف بهره گرفته شده است. انتظار می‌رود که چنین پیکره‌هایی بهترین و به‌روزترین منبع برای مطالعه در حوزه زبان‌شناسی پیکره‌ای به‌شمار آیند. مطالب داده‌های اغلب فرهنگ‌های این مجموعه، افزون بر کتاب، بر روی لوح فشرده نیز در دسترس است. بدین‌گونه، می‌توان بیشترین بهره را در پژوهش‌های مربوط به «واژگان پایه»<sup>۵</sup> از این مجموعه‌ها گرفت.

ظاهراً همه مجموعه فرهنگ‌های بسامدی راتلج، با جلد سخت (گالینگور) و جلد نرم (شمیز)، به همت انتشارات راتلج (گروه تیلور و فرانسیس)<sup>۶</sup> چاپ و منتشر شده‌اند.

4. language competence and performance

5. basic vocabulary

6. Routledge (Taylor & Francis Group)

## ۲- فرهنگ بسامدی فارسی

در این مقاله به بررسی و نقد این کتاب پرداخته‌ایم:

*A Frequency Dictionary of Persian: Core Vocabulary for Learners.* Corey Miller, Karineh Aghajanian-Stewart, USA/UK: Routledge, 2017, 2<sup>nd</sup> inprint: 2018.

[عنوان فارسی: فرهنگ بسامدی فارسی؛ واژگان پایه‌ی زبان و ویژه‌ی زبان‌آموزان. کوری میلر، کارینه آقاجانیان-استوارت. ایالات متحده آمریکا/ بریتانیا: راتلج، ۲۰۱۷، چاپ دوم: ۲۰۱۸ م.]  
این کتاب نخست بار در سال ۲۰۱۷ م به همت انتشارات راتلج/ راتلج/ روتلج منتشر شد. چاپ دومی از این فرهنگ، با اصلاح یک خطای عددی بر روی جلد، در ۲۰۱۸ م منتشر شد. این اثر جزو مجموعه «فرهنگ‌های بسامدی راتلج» چاپ و منتشر شده است.

### ۱-۲. مؤلفان کتاب

کوری [آندرو] میلر مدیر پژوهش‌های «بازشناسی خودکار گفتار»<sup>۷</sup> در شرکت «رو»<sup>۸</sup> (تبدیل خودکار گفتار- به- متن<sup>۹</sup>)، در ایالت تگزاس، ایالات متحده است.  
کارینه آقاجانیان-استوارت عضو هیئت علمی پژوهشی در «مرکز مطالعات پیشرفته زبان»<sup>۱۰</sup> در دانشگاه مریلند، ایالات متحده است.  
بر روی جلد و در صفحه‌ی عنوان تنها نام این دو را می‌بینیم؛ اما، در صفحه‌ی عنوان دوم، نام چهارده تن دیگر را در زیر نام دو مؤلف اصلی آورده‌اند. یک نام ایرانی است و سایر نام‌ها غیر ایرانی و غالباً فرنگی‌اند.

### ۲-۲. پیکره مبنا

مبنای استخراج ۵۰۰۰ واژه پربسامد در فرهنگ بسامدی فارسی، بنا بر آنچه مؤلفان نوشته‌اند، پیکره‌ای متنی با ۱۵۰ میلیون رکورد<sup>۱۱</sup> [و جمعاً با اندکی بیش از ۱,۳۹۵,۰۰۰ گونه‌ی واژگانی یا صورت‌واژه<sup>۱۲</sup>] از متن‌های نوشتاری و گفتاری از دنیای فارسی‌زبان بوده است (نک: Miller & Aghajanian-Stewart, p. 1-2) (۲)؛ اما

7. automatic speech recognition/ ASR      8. Rev Company (rev.com)  
9. speech-to-text (STT)      10. the Center for Advanced Study of Language  
11. token      12. token [or word-form]

قرآینی روشن نشان می‌دهد که حجم پیکره مبنا کمتر از یک‌دهم مقداری است که اظهار داشته‌اند. مهم‌ترین قرینه، بسامد ذکرشده برای واژه‌های پرسامد است. برای مثال، در پیکره‌های فارسی، بسامد کاربرد حرف ربط «و» – که به «واو عطف» نیز معروف است – حدود ۴,۵ تا ۵,۵ درصد کل واژه‌هاست؛ بر این اساس، بسامد «و» به نسبت ۱۵۰ میلیون، در این پیکره باید حدود ۶,۸ تا ۷,۵ میلیون بار می‌بود؛ اما چنین نیست و بسامد آن تنها ۵۳۳,۷۱۲ بار است. برپایه این برآورد، حجم پیکره مبنا مؤلفان در تدوین این فرهنگ احتمالاً حدود ۱۰ تا ۱۲ میلیون واژه است.<sup>(۳)</sup>

2 A Frequency Dictionary of Persian

Table 1.1 Persian Corpora

Corpus	Tokens	Types	Modality	Comments
Hamshahri	148,438,042	1,066,204	Text	subject-labeled
Bijankhan	2,409,535	76,540	Text	POS-tagged, subject-labeled
Fiction	1,001,754	61,692	Text	novels, plays and short stories
Dari	1,099,752	65,705	Text	news, government reports, academic articles
Blogs	561,482	63,142	CMC	blogs
LDC CALLFRIEND Farsi	198,098	11,182	Speech	telephone conversations
Raytheon BBN Broadcast Monitoring System (IRINN) Top 10k	2,783,424	50,712	Speech	speech recognition of television broadcasts

جدول (۱) فهرست پیکره‌هایی که مؤلفان آن‌ها را مبنا کار خود معرفی کرده‌اند

## ۲-۳. فهرست مطالب کتاب

۱. مطالب کتاب عبارت‌اند از:

- [فهرست] «فهرست‌های موضوعی واژگان»؛
- پیشگفتار مجموعه (به قلم پل ریزن، از دانشگاه لَنکَسْتِر، بریتانیا، متخصص پردازش زبان طبیعی<sup>۱۳</sup>، و مارک دیویس، زبان‌شناس آمریکایی و متخصص زبان‌شناسی پیکره‌ای، دانشگاه بریگم یانگ، شهر پرُو، ایالت یوتا، ایالات متحده<sup>۱۴</sup>)؛
- سپاس‌گزاری؛
- فهرست کوتاه‌نوشت‌ها/ نشانه‌های اختصاری؛
- مقدمه؛

13. Paul Rayson, Lancaster University, UK; 'natural language processing/ NLP' expert

14. Mark Davies, Brigham Young University (BYU), Provo, Utah, USA; 'corpus linguist'

- منابع [مقدمه]؛
- فهرست بسامدی ۵۰۰۰ واژه (بر اساس بسامد یا کاربرد واژه‌ها، از بیشترین تا کمترین تکرار)؛
- فهرست الفبایی ۵۰۰۰ واژه؛
- فهرست‌های موضوعی واژگان، شامل ۳۱ جعبه، در لابه‌لای صفحات بخش اصلی کتاب، یعنی نمایه فهرست بسامدی<sup>(۴)</sup>؛
- نمایه هویت‌های دستوری واژه‌ها، از جمله تفاوت‌های میان کاربرد نوشتاری و محاوره‌ای فعل‌ها، ساختار همکردها (فعل‌های سبک)، فرایندهای ساخت واژی و واژه‌های عربی دخیل و نیز واژه‌های بومی (ایرانی)۔ در برخی موضوعات.

## ۲. محاسن کتاب

مهم‌ترین محاسن فرهنگ بسامدی فارسی عبارت‌اند از:

۱-۲ به دست دادن بیشتر واژه‌های پایه فارسی امروز، گرچه نه با دقتی خاص. بررسی‌هایی که نگارندگان در این باره صورت داده‌اند، مقایسه ۵۰۰۰ مدخل این اثر با ۵۰۰۰ بن‌واژه پرسامدتر «پیکره فارسی امروز» است.<sup>۱۵</sup> با اندکی اغماض و گذشته از اشکالات فراوان اثر، نزدیک به ۷۵ درصد مدخل‌های این فرهنگ را می‌توان جزو ۵۰۰۰ واژه یا بن‌واژه پرسامدتر فارسی امروز به‌شمار آورد.

توضیح: «پیکره فارسی امروز»، گردآورده سید حمید حسنی، جمعاً شامل ۱۵,۰۰۰ متن کوچک و بزرگ از منابع نوشتاری و گفتاری نیمه دوم سده چهاردهم هجری خورشیدی (۱۳۵۱-۱۴۰۰) و مجموعاً مشتمل بر بیش از ۹۰,۰۰۰,۰۰۰ صورت‌واژه است. حسنی این پیکره را در طول بیش از ۲۵ سال گذشته گرد آورده است. برچسب‌گذاری/ برچسب‌دهی دستوری و واژگانی<sup>۱۶</sup> بیش از نود درصد متن‌های پانزده‌هزارگانه به پایان رسیده است. این فرایند برچسب‌گذاری به صورت دستی - نه به صورت خودکار و با کمک گرفتن از برنامه‌های برچسب‌گذاری خودکار<sup>۱۷</sup> - صورت گرفته است.

۲-۲ دادن «فهرست‌های موضوعی واژگان»<sup>۱۸</sup> در قالب جعبه‌های موضوعی، شامل ۳۱ جعبه مستقل، که بیشتر آن‌ها حاوی ده‌ها بن‌واژه استخراج‌شده از پیکره مبنای کار مؤلفان بوده‌اند. برخی جعبه‌ها شامل واژه‌هایی فراتر از ۵۰۰۰ مدخل اصلی موجود در کتاب‌اند. این جعبه‌ها در خلال صفحات بخش اصلی، یعنی بخش فهرست بسامدی، آورده شده‌اند.

۱۵. در بخش مربوط به «معایب و اشکالات فنی» به این موضوع پرداخته‌ایم.

16. grammatical & lexical tagging

17. automatic tagging /labelling programs

18. Thematic vocabulary lists

۲-۳ دادن فهرست واژه‌ها برپایه هویت‌های دستوری/ اجزای سخن/ پاره‌های گفتار (ص 317-365؛ شامل نه فقره: صفت‌ها، قیده‌ها، حروف ربط، اسم‌ها، حروف اضافه، ضمائر، اسم‌های خاص، کتبت‌نماها، فعل‌ها)<sup>(۵)</sup>.

۲-۴ فعل‌های مرکب را نیز در شمار آورده‌اند و یکی از جعبه‌های موضوعی را هم به این دسته اختصاص داده‌اند. البته، اشکالاتی در شیوه کارشان هست. در بخش «معایب و اشکالات فنی» به اشکال‌های مرتبط اشاره کرده‌ایم.

### ۳. معایب و اشکالات فنی کتاب

اشکالات و معایب و کاستی‌های کتاب از دو گونه روش‌شناختی و موردی یا موضوعی است و شمار آن‌ها فراوان است. در ادامه به معایب اصلی کتاب و اشکالات فنی آن از هر دو گونه و به‌ویژه از نگاه فن‌پیکره‌نگاری<sup>۱۹</sup> می‌پردازیم و از اشکالات جزئی درمی‌گذریم.

#### ۳-۱. اشکالات فراوان در ترتیب الفبایی، در فهرست‌های الفبایی

کتاب در ترتیب الفبایی سه حرف «ک» و «و» و «ه» اشکال‌های اساسی دارد؛ به طوری که ترتیب شش حرف پایانی چنین است: م، ن، ه، و، ک، ی. همچنین، آنجا که میان عناصر بن‌واژه (۶) فاصله قرار دارد، نظم و ترتیب الفبایی همچنان به صورت ماشینی و اصلاح نشده باقی مانده است. برای مثال، «در حال» و «در زدن» بعد از «در آمد» و «در آمدن» آمده‌اند، و «گرامیداشت» پس از «گرامی داشتن» آمده است. نمونه‌های اینچنینی در کتاب بسیار زیاد است.

#### ۳-۲. نبود واژه‌ها/ بن‌واژه‌های بسیار پربسامد و پربسامد در اثر

در این فرهنگ از وجود این واژه‌های پربسامد خبری نیست. چند نمونه از بن‌واژه‌هایی که به دلیل فقدان روشی درست و تصویری صحیح از بن‌واژه‌سازی<sup>۲۰</sup> از قلم افتاده‌اند (به ترتیب الفبا): آخری، آخرین، آلا، ...، دیگران (ضمیر)، دیگری (ضمیر)، راه، صبر، صبر کردن، لزوماً، یادآوری کردن، یاد دادن، یادشده، ... همچنین: اولی، اولین، نخستین، دومی، دومین، سومی، سومین، چهارمین، پنجمین، ششمین، هفتمین، ... و فعل‌های مرکبی مانند احتمال دادن، اختصاص دادن (به)، اخراج کردن (از)، ادا کردن (به)، ارائه دادن (به)، ارائه کردن (به)، ارتقا دادن (از، به)، ارتقا یافتن (از، به)، ارجاع دادن (به)، استخراج کردن (از)، اعتماد کردن (به)، اعدام کردن، اعزام کردن (به)، اعطا کردن،

افتتاح کردن، افتخار کردن (به)، افراط کردن، افزایش دادن (به)، افشا کردن، اقدام کردن (به)، ... درضمن، واژه‌هایی نظیر «ارتباطات» و «اطلاعات» در فارسی مفردند، و جای این‌ها هم در فرهنگ خالی است. برعکس این، برای نمونه، «مذاکرات» را مدخل کرده‌اند؛ درحالی‌که جمع «مذاکره» است.

مجموع بسامد «باید» و «نباید»، «بایستی» و «نبایستی»، «می‌باید» و «نمی‌باید»، «می‌بایست» و «نمی‌بایست»، و نیز «می‌بایستی» و «نمی‌بایستی» را در مدخل «بایستن» (ش ۱۹) آورده‌اند. اما، نقد وارد بر این اقدام این است که کلمه‌های یادشده در فارسی امروز «فعل» نیستند و صرف هم نمی‌شوند، بلکه «قید»ند (در فارسی کهن، برای مثال، «باید» و «بایستی» فعل مضارع سوم شخص مفرد بوده‌اند). بنابراین، هر کدام از فقره‌های یادشده، به‌ویژه «باید» و «نباید» و «بایستی» و «نبایستی»، که پربسامدترند، در فهرست‌های پیکره‌ها و فرهنگ‌های بسامدی فارسی امروز، با هویت دستوری قید، به مدخلی جداگانه نیاز دارند.

«آیت» را، که در فارسی امروز کم‌بسامد است، مدخل کرده‌اند. روشن است که این واژه حاصل بی‌توجهی به فاصله‌گذاری است و، در غالب موارد، جزء آغازین «آیت‌الله» بوده است.

۳-۳. جابه‌جایی اعداد بیش از یک رقمی در متن برخی شاهد مثال‌ها و بی‌دقتی در برگردان به انگلیسی

اعداد در برخی شاهد مثال‌ها وارونه‌یا، به اصطلاح خودمانی، به صورت «چپه» ثبت شده‌اند. این وارونگی یا چپگی در تاریخ‌ها به شکلی عجیب، نامتناسب با بافت، و حتی مضحک درآمده است. برای نمونه، در مدخل «لیسانس» (ش ۴۰۱۴)، شاهد مثال این چنین ثبت شده است: «لیسانس خود را در سال ۱۱۳۱ در این مکان گرفتند.» [به‌جای ۱۳۱۱]. در ترجمه نیز، نه تنها همین عدد ۱۱۳۱ آمده، بلکه شاهد مثال درست ترجمه نشده است:

He received his bachelor's in this location in the year 1131.

به نظر می‌رسد که فعل جمع در جمله اصلی (فارسی)، که به‌عنوان شاهد مثال آمده، محترمانه است و نه جمع حقیقی؛ یعنی، مثلاً فاعل آن «ایشان» یا چیزی شبیه این است. بسامد خام «لیسانس» ۲۴۸ بار است، و گردآورندگان به‌سادگی می‌توانستند شاهد مثالی بهتر به جای شاهد مثال کنونی بگذارند.

نمونه‌ای دیگر: در مدخل «مبلغ، مبلغ» (ش ۱۵۱۰)، شاهد را تنها برای «مبلغ» آورده‌اند و به جای «۷۵ سنت» نوشته‌اند «۵۷ سنت».

۳-۴. ثبت نادرست بن‌واژه



«کاهیدن» (۴۷۵۸) به جای «کاستن». مصدر «کاهیدن» در زبان فارسی سابقه کاربرد دارد، اما این سابقه تاریخی یا در زمانی یا تاریخی<sup>۲۱</sup> است (نک: Ullmann, p. 159; Trask (1); طباطبایی). وقتی نویسنده یا گوینده، در فعلی ماضی، از ستاک گذشته/ بن ماضی فعل استفاده کرده باشد، مصدر آن تابع ستاک گذشته خواهد بود. چند مثال برای کاربردهای تاریخی (قدیمی و/ یا منسوخ): آراییدم، آلایده‌اند، بازید، تازیدید، دوزیدیم، سازیدی، سوزیدند، کاهیدی، ... > آراییدن، آلایدن، بازیدن، تازیدن، دوزیدن، سازیدن، سوزیدن، کاهیدن، ...

۳-۵. ثبت بن‌واژه‌های دوتلفظی با تنها یک تلفظ، به دلیل عدم استفاده از فرهنگ‌های استاندارد/ معیار

«کاووش» (kavof: ۴۴۱۶)؛ تلفظ اصلی (و امروزه تلفظ ثانویه آن) kavef (اسم یا اسم مصدر از «کاویدن») است. در فرهنگ بزرگ سخن (انوری (۱)) دو تلفظ داده شده است: kāv(e)š<sup>۲۲</sup>.

۳-۶. اشکال در دادن هویت‌های دستوری/ اجزای سخن/ پاره‌های گفتار (PoS) به ذکر چند نمونه بسنده می‌کنیم:

«کرد» (ش ۱۸: kærd, kord) (هویت دستوری «کرد») را تنها فعل گزارش کرده‌اند؛ درحالی‌که شاهد مثال این مدخل «کرد» (صفت، اسم، اسم خاص؛ انگلیسی: Kurd) است. نتیجه عدم برچسب‌گذاری دستوری و واژگانی این است که صورت‌واژه‌ای از بن‌واژه «کردن» (کُند: سوم‌شخص مفرد گذشته) را با صفت/ قید «کُند» یک‌جا بیاورند. دقیقاً مدخل پیشین آن (ش ۱۷) مصدر «کردن» است.

«کند» (konæd, kond) (هویت دستوری «کند») را تنها صفت گزارش کرده‌اند؛ درحالی‌که ابتدا تلفظ «کُند» (فعل) را آورده‌اند و سپس «کُند» (صفت [و قید]) را. در اینجا نیز با نتیجه عدم برچسب‌گذاری دستوری و واژگانی روبه‌رویم. صورتی صرف‌شده از مصدر «کردن» (کُند) را با صفت/ قید «کُند» یک‌جا آورده‌اند. حال که چنین کرده‌اند، باید یادآور شویم که «کُند» گذشته سوم‌شخص مفرد از «کندن» را، که بسیار بعید است در پیکره مبنای مؤلفان کاربرد نداشته بوده باشد، از قلم انداخته‌اند.

«مُسلِم، مسلّم» (moslem, mosællæm): «مسلّم» با هویت دستوری قید هم در فارسی امروز کاربرد دارد، به معنای مسلماً: «مسلّم این کار از او بر نمی‌آید. ○ مفت و مسلّم به چنین موهبت عظمایی رسیده‌ای.» (نک. انوری (۱)): «مسلّم»؛

۲۱. diachronic ≠ synchronic; historical (در مقابل همزمانی؛ معاصر)

۲۲. به شیوه آوانویسی/ واج‌نویسی [مجموعه فرهنگ‌های «سخن»

«مفصل، مفصل» (۳۵۰۹: mæfsæl, mofæssæl): «مفصل»، افزون بر هویت دستوری صفت، کاربرد قیدی هم دارد (به معنی به تفصیل، مفصلاً): «مفصل راجع به موضوع با او صحبت کردم». (نیز نک: انوری (۱): «مفصل»، معنای ۳)

### ۳-۷. آوردن عدد و معدود به صورت نامنظم

برای نمونه، این‌ها را نزدیک (بی فاصله یا چسبیده) نوشته‌اند: «دوبار»، «یکبار»؛ «یکسال». این کار نادرست است؛ زیرا عدد و معدود ساده را باید فاصله داد؛ به این دلیل که هرکدام یک واحد واژگانی مستقل‌اند (نک: ۱-۲-۱). عدد و معدود را تنها در جاهایی که واحد واژگانی می‌سازند باید نزدیک — یعنی چسبیده یا بی فاصله یا با نیم فاصله — نوشت؛ مانند: یکسر (قید)، دورو (صفت/قید)، سهره (اسم)، چهارسو/چارسو (صفت/اسم).

### ۳-۸. بی‌نظمی در ثبت برخی واژه‌های مرکب و همایندها

برخی بن‌واژه‌ها را که در اصل ترکیب اضافی یا وصفی (و نیز از گونه ترکیب مقلوب) بوده‌اند یا پیشوندواره دارند و در فرایند ساخت واژی - نحوی<sup>۲۳</sup> به واژه‌ای مرکب تبدیل شده‌اند، به‌علاوه بعضی قیده‌ها، حروف اضافه، حروف ربط مرکب، ترکیب‌های عطفی، شبه جمله‌ها، ضمائر، و همچنین بعضی اسم‌های خاص مرکب، مدخل کرده‌اند. چند نمونه: «دردسر»، «محیط‌زیست»، «مدیرعامل»، «مدیرکل»، «تیرماه»، «مهرماه»، «بجای» (سرهم)، «بجز» (سرهم)، «بخاطر» (سرهم)، «درحال»، «درحالی‌که»، «زیرویمی»، «علیه السلام»، «بخصوص» (سرهم)، «بزودی» (سرهم)، «درواقع»، «هرگاه»، «هرکس»، «غیر دولتی»، «غیر قانونی»، «غیر نظامی»، و حتی «قانون اساسی»، «مواد مخدر» (بی فاصله: مواد مخدر)، «حقوق بشر» (سرهم: حقوق بشر!)، «نماینده [مجلس]»، «خط لوله»، «خاورمیانه»، و «خلیج فارس»... این شیوه در پیکره‌نگاری و فرهنگ‌نگاری، با وجود تفاوت‌هایی در شیوه، امری پسندیده است؛ اما، چون به امر یک‌دستی و برچسب‌گذاری نظام‌مند مبنی بر واژه‌شناسی توجه نداشته‌اند، ترکیب‌های پربسامد مشابه بسیاری را به‌سادگی از کف داده‌اند. برای نمونه، «خط‌مشی» را نیآورده‌اند و شاهد مثال آن، که خود یک کلمه است، در مدخل «مشی» آمده است.

### ۳-۹. عدم برچسب‌گذاری درست بن‌واژه‌ها

مراجعه‌کننده، با جست‌وجو در مدخل‌های گوناگون و مقایسه آن‌ها، با قدری تأمل ویژه، درمی‌یابد که همه عناصر واژگانی به‌دقت برچسب‌گذاری نشده‌اند. برای مثال، در مدخل «کام» (۳۵۷۹)، با

هویت دستوری اسم (n) <sup>۲۴</sup>، این معادل‌ها آمده‌اند: palate, goal, com. این سه واژه، از نظر ریشه‌شناختی، از سه ریشه واژگانی متفاوت‌اند:

(۱) «کام» فارسی (= palate) به معنای دهان و سقف دهان؛

(۲) «کام» (= goal) در معنای آرزو و خواسته دل؛

(۳) «کام» انگلیسی (< com) در «دات کام» (.com)، دامنه اینترنتی عام. در پیکره‌هایی که به درستی و با دقت برچسب‌گذاری می‌شوند، این سه را باید جدا کرد و بسامد هر کدام را جداگانه به دست داد. نمونه‌های دیگر:

الف) با تلفظ یکسان یا تلفظ‌های نزدیک، اما با ریشه‌های متفاوت: آرامی، بار، بر، تار، تیر، در، کف، کیست؛ گرمی، گرمی (gæirmi). تلفظ شهر گرمی (در استان اردبیل)، گرمی (Germi) است، و آن را هم ثبت نکرده‌اند.

ب) با تلفظ و ریشه یکسان، اما با هویت‌های دستوری متفاوت: کلان («کلان» را، هم در کاربرد صفتی/وصفی و هم به‌عنوان پیشوندواره، یک بن‌واژه گرفته‌اند. معادل‌های انگلیسی آن را macro- و big- آورده‌اند، البته با یک هویت دستوری: (adj = adjective). شاهد مثال را تنها برای «کلان» صفت آورده‌اند.؛ کمال (kæmal, Kæmal)؛ مشتری (moʃtæri)؛ هادی (hadi, Hadi)؛ ... البته موارد مشابه پرشماری را از قلم انداخته‌اند. برای نمونه، به ذکر مدخل‌هایی بسنده می‌کنیم که با حرف «آ» آغاز می‌شوند؛ مواردی که مؤلفان در آن‌ها فراموش کرده‌اند که این‌ها اسم خاص هم هستند (و در محدودی از این‌ها، به شیوه‌ای نامنظم، به اسم خاص بودنشان اشاره کرده‌اند: pn (کوته‌نوشت proper name/ proper noun)؛ اما یادشان رفته آن‌ها را با حرف آغازین بزرگ/ کپیتال هم بیاورند): آبادانی، آدم، آرام، آرمان، آزاد، آزاده، آزادی، آستانه، آفرین، آقایی، آن.

نکته ۱: حتی در جعبه نام‌های خانوادگی (جعبه موضوعی شماره ۱۲، ص ۱۰۶-۱۰۵) آبادانی و آزاده و آفرین را آورده‌اند.

نکته ۲: در جعبه نام‌های زنانه (جعبه موضوعی شماره ۹، ص ۸۱) آذر و آزاده و آفرین را آورده‌اند. آزاده و آفرین را، هم در جعبه نام‌های زنانه فهرست کرده‌اند و هم در جعبه نام‌های خانوادگی.

نکته ۳: در جعبه نام‌های مردانه (جعبه موضوعی شماره ۱۴، ص ۱۲۷-۱۲۶) آدم و آرام را آورده‌اند. در ضمن، در همین سه جعبه، دست به کاری نامعمول زده‌اند؛ یعنی ترجمه یا صورت انگلیسی بسیاری از نام‌های خاص را هم ذکر کرده‌اند. برای نمونه، در جعبه‌ها، نام‌های زنانه اکرم، پروانه،

شادی، و نرگس را<sup>۲۵</sup> dearer, butterfly, joy, narcissus؛ «آرام»، «پویا»، «صادق»، و «یوسف» را peaceful، dynamic, truthful, Joseph و نام‌های خانوادگی اسلامی، الهی، ایرانی، بزرگی، و بیرونی را Islamic، divine, Iranian, magnitude, external را، به گونه‌ای که ملاحظه شد، با برابرنهاد یا صورت انگلیسی آورده‌اند. این‌ها تنها نمونه‌هایی معدود از آن‌همه نام کوچک زنانه و مردانه و همچنین نام خانوادگی بی‌جهت ترجمه شده بود. اهل فن آگاه‌اند که این آفت خطرناک نیز باز حاصل کم‌دقتی در برچسب‌گذاری واژگانی و دستوری پیکره است؛ زیرا این کار، در مقیاس بزرگ، فوق‌العاده دشوار و بلکه طاقت‌فرساست. نمونه‌هایی که دیدیم، از نظر روش شناختی، به کلی غریب و نامعمول‌اند.

همه این‌ها در حالی است که، در اواخر کتاب، در فهرست اسم‌های خاص (ص 361-351)، که جزو بخش بزرگ‌تر «نمایه هویت‌های دستوری/ اجزای سخن/ پاره‌های گفتار» است، ۵۸۶ نام را به ترتیب بسامد آورده‌اند و به گونه‌ای متفاوت عمل کرده‌اند؛ آن‌چنان‌که گویی شیوه خود در سایر فهرست‌ها و در جعبه‌ها را از یاد برده‌اند. این فهرست با «آن» (Anne) آغاز می‌شود (بن‌واژه شماره ۱۳؛ همان‌که در فهرست بسامدی و در ص 8 با هویت‌های دستوری «اسم، ضمیر، اسم خاص» آمده‌بود).

پ) با تلفظ‌های متفاوت، اما با هویت‌های دستوری و ریشه یکسان: «رحم، رجم»، «مرتفع، مرتفع». شاهد مثال از آن «مرتفع» است: «ساختمان‌های مرتفع». در ضمن، میان «ساختمان» و «های» نیز، مانند بسیاری جاهای دیگر، فاصله‌ای نابه‌جا افتاده.

ت) با تلفظ‌ها و هویت‌های دستوری متفاوت، اما با ریشه یکسان: «مراجع، مراجع» (خود «مراجع» جمع «مرجع» و صورت‌واژه‌ای از این بن‌واژه است و نباید بن‌واژه‌ای جداگانه گرفته می‌شد.)؛ «مُسَلِّم، مسلم»؛ «مشترک، مشترک»؛ «مشرّف، مشرف»؛ «مفصل، مفصل».

ث) با تلفظ‌ها و هویت‌های دستوری و ریشه‌های متفاوت: «کرد، گرد»؛ «کره، گره<sup>[۱]</sup>، گره<sup>[۲]</sup>، کره» (گره [۱]: شکل هندسی (عربی، اسم، صفت)؛ گره [۲]: نام دو کشور شمالی و جنوبی (اسم خاص جغرافیایی). همین «گره»، در فارسی چند دهه اخیر، به معنای دیگری هم کاربرد داشته و دارد؛ اما، چون کم‌کاربردند، از اشاره به آن‌ها درمی‌گذریم و فرض می‌کنیم که در پیکره مبنای کار مؤلفان این فرهنگ به‌کار نرفته‌اند.)؛ «کُشته، کشته» (خود «کُشته» و «کِشته» هم، به نوبه خود، بسته به بافت سخن (context)، بالفعل، سه یا چهار کاربرد و هویت دستوری متفاوت دارند: اسم، صفت، فعل ماضی نقلی بدون «است»، فعل وصفی.)؛ «کُشتی/ کشتی»، «کُشتی» (خود «کُشتی» و «کِشتی» هم، باز، بسته به بافت سخن، بالفعل، به‌مانند فقره پیشین، سه یا چهار کاربرد و هویت دستوری متفاوت دارند.)؛ «کلیه، کلیه» (خود «کلیه» (اولی) دو معنی دارد از دو ریشه متفاوت: ۱) عضو دوتایی لوبیاشکل داخلی بدن...، از عربی؛ ۲) گردن‌بند، از فرانسوی (collier).)؛ «کُندی، کیدی»؛ «کی، کی»؛ «کیف،

۲۵. نام گل «نرگس» در انگلیسی Narcissus است و حرف نخست آن را با حرف بزرگ/ کپیتال باید نوشت.

کیف)؛ «کیفی، کیفی»؛ «گرد، گرد، گُرد» (در اینجا، درست بر پایه شیوه خود مؤلفان، «-گرد» (gærd-)، به عنوان جزء پسین کلمه‌های مرکب (در معنای «گردنده»)، مانند «جهان‌گرد»، «دوره‌گرد»، «شب‌گرد»، «ول‌گرد»، و چند کلمه دیگر، از قلم افتاده است. همچنین، «گرد» (gerde [یا: gerd-e]) حرف اضافه است، و آن را هم از قلم انداخته‌اند.)؛ «گردان، گردان» (در اینجا، باز بر پایه شیوه خود مؤلفان، «-گردان» (gærdan-) از قلم افتاده است: جزء پسین کلمه‌های مرکبی (در معنای «گرداننده» و «چرخاننده»)، مانند «آتش‌گردان»، «جوجه‌گردان»، «روگردان/روی‌گردان»، «سرگردان»، «سگ‌گردان»، و چند کلمه دیگر.)؛ «گل، گل، گُل<sup>[۱]</sup>، گُل<sup>[۲]</sup>» («گل» (gæl) نیز در فارسی امروز کاربرد دارد و آن را از قلم انداخته‌اند: گل میخ، گل چوب‌رختی/رخت‌آویز. همچنین، «گل» (gæle [یا: gæl-e]) حرف اضافه است به معنای «[در] بالای» (کَلّه او و امثالش گل دار خواهد رفت.) و «[در] کنار» (دور روز تعطیل گل هم افتاده بود.) و «[به] دُور» (طناب افتاد گل پایم.) (نک: انوری (۱): «گل»، ترکیب «گل» (حرف اضافه))، و آن را هم از قلم انداخته‌اند.)؛ «گله، گله» («گله» (gole) هم در فارسی امروز داریم: ۱) شکلی دایره‌مانند بر سطح چیزی (نک: انوری (۱): «گله» (۱)؛ ۲) اخگر، شاره، شعله (نک: همو، همان؛ نیز نک. گُل ۱، معنای ۷)؛ ۳) اسم خاص شخص: نام خانوادگی (دست‌کم فریدون گله.)؛ «گلی، گلی، گُل<sup>[۱]</sup>، گُل<sup>[۲]</sup>» («گلی»: ۱) به رنگ سرخ روشن، گُل‌رنگ (نک: انوری (۱): گلی، معنای ۱: کراوات خاکستری با گُل‌های گلی)؛ ۲) گُل فروش (نک: همو، همان: گلی، معنای ۲: آهای گلی! گُل‌ها را چند می‌دهی؟)؛ ۳) اسم خاص شخص: نام کوچک زنانه.)؛ «ماده، ماده» (در فارسی امروز، به ویژه در گفتار، «ماده»ی فارسی (مقابل «نر») را هم کمابیش با تشدید تلفظ می‌کنند. دیهیم، در مدخل «ماده (ضدِ نر)»، نتیجه پژوهش میدانی را ۱۴ بار با تلفظ madde و ۳ بار با تلفظ made داده است.)؛ «مبلغ، مبلغ»؛ «مُجاز، مُجاز» (هویت دستوری «مُجاز، مُجاز» را تنها صفت نوشته‌اند؛ درحالی‌که «مُجاز» صفت است و «مُجاز» هم اسم و هم صفت (نک: انوری (۱): مُجاز ۱).)؛ «مُحرم، مُحرم» («مُحرم» هم در فارسی امروز داریم: ویژگی آن‌که... احرام می‌بندد (نک: انوری (۱): «مُحرم»). فارغ از این‌که «مُحرم» را باید جدا می‌کردند (و اصولاً این شیوه را مراعات نکرده‌اند)، اگر در پیکره مبنای مؤلفان، حتی کم‌بسامد بوده باشد، ذکر آن لازم بود.)؛ «مُد/مَد» (mædd, mod)؛ «والا، والا» (شاهدمثالی که آورده‌اند، مانند همه شاهدمثال‌های مدخل‌های دارای بیش از یک تلفظ، ویژه دومی (والا: va?ella) است: «زود آب برسان به من [،] والا می‌میرم.»؛ و برای «والا» (vala) شاهدمثال نیاورده‌اند). اما، برعکس، برای مثال، در «کُشتن» از تلفظ «کشتن» اثری نمی‌بینیم؛ درحالی‌که «کشتن» هم – گرچه کمتر از «کاشتن» – در فارسی دهه‌های اخیر کاربرد دارد.

۳-۱۰. بی‌توجهی یا کم‌توجهی به به‌هنجارسازی/هنجارسازی پیکره مینا

۳-۱۰-۱. عناصر واژگانی و دستوری پیکره مینا

به‌هنجارسازی/هنجارسازی/نرمال‌سازی<sup>۲۶</sup> عموماً شامل تمیزکاری نظام‌مند و ویرایشی متن‌ها

متناسب با کار پیکره‌نگاری، یکدست‌سازی روشمند، تعدیل ناهنجاری‌ها، و رمزگذاری<sup>۲۷</sup> درست آن‌هاست. برای مثال، در بسیاری از مدخل‌های ختم‌شونده به پسوند «ی» نسبت، به‌قرینه جایگاه بسامدی، یعنی میزان کاربرد واژه، به نکره بودن «ی» در برخی موارد کاربرد می‌توان پی برد. اینک نمونه‌هایی روشن از دو حرف «آ» و «ا» را، که در آن‌ها به‌هنگارسازی مغفول مانده، به‌ترتیب الفبا می‌آوریم: آدمی، آرامی، آزادی، آزمایشگاهی، آزمایشی، آسانی، آشنایی، آقایی، ابری، اتفاقی، اجتماعی، احتمالی، اخلاقی، ادبی، ارتباطی، ارزی، اساسی، استاندارد، استانی، استثنایی، استواری، اسمی، اصلی (و اصولی)، اعتباری، اعتمادی، افتخاری، افراطی، اقتصادی، امامی، امپراتوری، امنیتی، امیدواری، امیدی، امینی، انتخاباتی، انتقادی، انسانی، انقلابی، ایالتی، ایمنی.

۳-۱۰-۲. عدم اصلاح املائی و رسم‌الخطی (به‌هنگارسازی/هنگارسازی املائی)  
در این بند تنها به ذکر دو مورد می‌پردازیم که به اصلاح جدی املائی نیاز داشته‌اند، و متعرض تغییرات رسم‌الخطی و چندوچون اصلاح آن‌ها نمی‌شویم؛ زیرا مبحث رسم‌الخط بایستی است مفصل و نیازمند فصلی مشیع.

در مدخل «آزمایش» (۱۱۲۶): «انجام آزمایش تیروئید قبل از بارداری ضروریست». سامانه، هنگام تجزیه خودکار صورت‌واژه‌ها، «ضروریست» [سرهم] را یک صورت‌واژه می‌گیرد؛ درحالی‌که «ضروری» و «ست» (= است) دو صورت‌واژه جداگانه‌اند و می‌بایستی آن‌ها را تفکیک می‌کردند.  
در مدخل «اکتشاف» (۳۶۳۲): «در طول تاریخ روشهای مختلفی برای اکتشاف نفت بکار برده شده است.» چون «بکار» سرهم است، مراجعه‌کننده حرفه‌ای حق دارد احتمال بدهد که در فهرست الفبایی و بسامدی کامل اثر هم با «بکار» روبه‌رو خواهد شد؛ به‌ویژه اینکه «بکار» اصطلاحی فرانسوی (be carre) در موسیقی نیز هست، در فارسی هم کاربرد دارد، و در پیکره‌های بزرگ، ولو با بسامد کم، پیدا می‌شود.

در امر رمزگذاری نامتناسب در کتاب، تنها به خودداری مؤلفان از یکدست‌سازی نویسه‌های فارسی و عربی اشاره می‌کنیم: آنان حتی به تفاوت دو نویسه «ک» و «ی» فارسی و دو نویسه «ك» و «ي» عربی و یکدست کردن آن‌ها توجه نکرده‌اند. با شمارش مکانیکی چهار نویسه یادشده، از طریق فایل پی‌دی‌اف اثر، به این نتیجه دست یافتیم که ۲۶,۷۷۶ «ی» فارسی، ۷۱۶ «ي» عربی، ۷,۷۲۴ «ك» فارسی، و ۱۴۴ «ك» عربی در سراسر متن وجود دارد.

۳-۱۰-۳. کم‌دقتی جدی در فاصله‌گذاری کلمات در مثال‌ها (بخشی از به‌هنجارسازی ویرایشی)  
از آنجاکه سامانه مورد استفاده در این کار، «تجزیه<sup>۲۸</sup>»ی صورت‌واژه‌ها را به شیوه «تجزیه خودکار/ ماشین/ مکانیکی»<sup>۲۹</sup> انجام داده است، این اشکال در مثال‌ها به مسئله‌ای مضاعف تبدیل می‌شود و اساس صورت‌واژه و بن‌واژه را بهم می‌ریزد. این اشکال تنها به ویرایش صوری و/یا ویرایش فنی مربوط نیست. چند نمونه موضوع را روشن‌تر می‌سازد:

مثال مدخل «اندیشه» (۳۰۰): «خداوند متعال فکر و اندیشه [ندیش] انسان را به این حقیقت بزرگ رهنمون می‌شود که اونی-تواند بیرون از سنتهای الهی زندگی کند.» آندراسکور [underscore] نشانه فاصله کاملی است که میان «می-» (پیشوند صرفی فعل) و «شود» افتاده است. همچنین، به جای «اونمی-تواند»، با دو واحد «اونمی-» و «تواند» روبه‌رویم؛ یعنی ضمیر «او» به پیشوند صرفی منفی فعل چسبیده و واحدی بی‌معنا و نادستوری ایجاد کرده است.

مثال مدخل «کلان» (۱۷۱۰): «بر برنامه‌ریزی کلان برای رفع آلودگی‌های ناشی از توسعه و صنعتی شدن کشور تاکید<sup>۳۰</sup> کرد.»

مثال مدخل «وارد» (۱۱۴): «مسعود بدون موجودیت یک حکومت دلخواه و مورد قبول مردم بار دیگر نمی‌خواست وارد کابل شود.» افزون بر فاصله کامل نابه‌جا میان «می-» و «خواست»، اتفافی بدتر افتاده: در واژه «دلخواه»، میان «دلخوا» و «ه» (حرف پایانی) هم فاصله کامل افتاده است. اما، به‌عکس، مثال مدخل «کلانتری» (۴۹۷۶): «۷۰ درصد کلانتری‌ها دوربین دارند.» میان «۷۰» و «درصد» باید یک فاصله کامل می‌دادند تا سامانه، در تجزیه خودکار صورت‌واژه‌ها، «۷۰ درصد» را یک واحد نگیرد.

با نگاهی گذرا در مدخل‌های کتاب، نمونه‌های بسیار دیگر از فاصله‌گذاری نامناسب در شاهد مثال‌ها را می‌توان یافت.

### ۳-۱۱. شاهد مثال نامناسب و/یا نادرست

در مدخل «مراقب» (۲۸۲۱): «ازش مراقبت کنه.» شاهد مثال برای «مراقبت کردن» است نه برای «مراقب». در مدخل «آخر» (۱۸۸): «آخرین کتلت رو از روی ماهیتابه برمی‌دارم.» خود واژه «آخرین» واژه‌ای مستقل و بن‌واژه‌ای جداست و صورت‌واژه‌ای از «آخر» نیست؛ درست مثل «اول» و «اولی» و «اولین».

28. parsing

29. automatic parsing/ machine parsing/ mechanical parsing

۳۰. «تاکید» را با الف و بدون همزه (به‌صورت «تاکید») نوشته‌اند.

«نخست» و «نخستین»، «دوم» و «دومی» و «دومین»، و مانند این‌ها. بنابراین، جای مدخل «آخرین» در این فرهنگ بسامدی خالی است. در پیکره‌ای استاندارد متعلق به فارسی امروز، بسامد کاربرد بن‌واژه «آخرین» از بسامد کاربرد بن‌واژه «آخر» کمتر نیست. افزون بر این، در فعل «برمی‌دارم»، میان «برمی‌[-]» (پیشوند فعل پیشوندی + پیشوند صرفی فعل) و «دارم» یک فاصله کامل افتاده؛ اما در قلم/فونت به‌کاررفته در کتاب موردبحث ما به چشم دیده نمی‌شود؛ ولی هنگامی که آن را از نسخه پی‌دی‌اف کتاب کپی کنیم و، مثلاً، به فایل ورد ببریم، فاصله کامل آن ظاهر می‌شود.

در مدخل «برای» (ش ۱۰): «دنیا اسباب‌بازی‌هایش را برای بچه‌ها گذاشته است.» آندراسکور [underscore]؛ معروف به آندرلاین: underline) نشانه فاصله نابه‌جا در شاهدمثال است.<sup>(۷)</sup> این شاهدمثال، جدا از اشکالات فنی‌اش، از نظر معنایی هم ابهام دارد. «دنیا» نام دختر است، و ترجمه انگلیسی شاهدمثال نیز همین را نشان می‌دهد. مثال در زیر مدخل «برای» آمده، و همه آگاهیم که «برای» از واژه‌های بسیار پرکاربرد فارسی است (رتبه آن در این فرهنگ، از نظر بسامدی، دهم است). بی‌تردید می‌توانستند ده‌ها هزار شاهدمثال بهتر برای «برای» بیاورند.

در مدخل «پیکر» (۲۵۹۳): «این سرزمین از یک پیکر می‌دانیم.» در این شاهدمثال، «این سرزمین» گروه مفعولی است، و کاربرد آن با فعل «دانستن» در این مثال به «را» نیاز دارد. بر این اساس، ظاهراً «را» از قلم افتاده.

در مدخل «تشریح» (۱۷۵۵): «وضعیت را برای رئیس تشریح کرده‌ای؟» در این شاهدمثال، املاي «رأیس» \* به‌شدت غریب و حتی مشکوک است. همچنین، جدا از اشکال فاصله‌گذاری در «کرده‌ای»، مثال از آن «تشریح کردن» است نه «تشریح».

در مدخل «شریعتی» (۴۰۲۴)، که اسم خاص شخص است، این شاهدمثال آمده: «شریعتی از دانشگاه علامه رفت.» این «شریعتی»، «سید صدرالدین شریعتی»، رئیس اسبق دانشگاه علامه طباطبائی، است. قطعاً در پیکره مبنا بیشتر شواهد به «علی شریعتی» اشاره داشته؛ پس نیکوتر آن بود که شاهدی مناسب‌تر می‌آوردند. هویت دستوری این مدخل را «اسم، اسم خاص» (n, pn) ذکر کرده‌اند، و این ناظر به کاربرد «شریعت [اسم] + ی [پسونده نکره]» است. اشتباه اخیر، یعنی خلط اسم و اسم خاص ده‌ها بار در کتاب تکرار شده است.

در مدخل «که» (ش ۵): «بنظر [سرهم] می‌آید که شلوغتر از جاهای دیگر است.» شاهدمثال برای حرف ربط «که» آمده؛ یعنی برای بن‌واژه‌ای بسیار پرکاربرد که در فرهنگ حاضر، از نظر جایگاه بسامدی، در رتبه پنجم است. قطعاً ده‌ها هزار شاهدمثال بهتر از این می‌توانستند برای «که» بیاورند.



بزرگ‌ترین عیب این مثال سرگردان بودن سیاق<sup>۳۱</sup> زبان آن میان زبان نوشتار و زبان گفتار است. تبصره: این نمونه‌ها تنها مثنوی از خروار بود. به دلیل ضیق فضا در مقاله، تنها به چند نمونه اشاره کردیم.

### ۳-۱۲. شاهد مثال نامفهوم، مبهم، و/یا اشکال‌دار

در مدخل «حوالی» (۳۵۳۷): «در حوالی میدان، بساطش را پهن کرده بود؛ فریب می‌فروخت.» جدا از اشکال فاصله‌گذاری (اشکال تکرارشونده بسیار رایج در کتاب)، شاهد مثال کاملاً مبهم و فراواقع‌گرایانه است.

در مدخل «طلبه» (۳۱۱۷): «من در این سنم تمام عمرم طلبه بودم.» قطعاً شاهد مثال‌های بهتر برای «طلبه» داشته‌اند.

در مدخل «هلندی» (۴۶۳۲): «هلندی‌ها پل فولادی چاپ می‌کنند.» ظاهراً شاهد مثال به چاپگر سه‌بعدی اشاره دارد. اما، در فرهنگ‌های این‌چنینی، مثال‌های روشن و رسا در اولویت‌اند.

### ۳-۱۳. شاهد مثال نادرست، نامناسب، یا ناقص برای مدخل

در مدخل «امیدوار» (۱۰۸۵): «در پایان انتقادی کلی را عنوان می‌کنیم که امیدواریم تاثیرگذار باش.» این شاهد مثال چند اشکال دارد. افزون بر فاصله نابه‌جا میان «می‌کنیم» و «کنیم» از سوی، و فاصله نابه‌جا میان «تاثیر» (چنین است، بدون همزه بر کرسی الف) و «[گذار]» از طرف دیگر، «باشد» را «باش» نوشته‌اند. آیا شاهد مثال در پیکره به همین صورت آمده بوده؟ اگر چنین بوده، چرا شاهد مثالی مناسب‌تر برای «امیدوار»، با آن‌همه پربسامدی‌اش، نیاورده‌اند؟

چند نمونه دیگر:

در مدخل «اخذ» (۱۸۷۱): «پس از اخذ مجوز از سازمان مدیریت و برنامه‌ریزی [کشور].» شاهد مثال ناقص و ابتر است. حتی اگر «پس از» را در سرآغاز مثال حذف می‌کردند، عبارت پذیرفتنی می‌شد.

در مدخل «اطلاع» (۲۰۹)، شاهد مثال از آن «اطلاعات» است. خود واژه «اطلاعات» در کاربرد فارسی امروز بن‌واژه است، از نظر کاربردی مفرد است، و مطلقاً صورت‌واژه‌ای از «اطلاع» نیست.

در مدخل «اقدام» (۲۲۵): «استان‌هایی که فاقد روزنامه بوده و یا یک روزنامه بیشتر ندارند باید اقدام به انتشار روزنامه کنند.» این شاهد مثال برای «اقدام کردن [به]» است.

در مدخل «سیمان» (۲۹۶۸): «بخش سیمان ما حدوداً سه هزار و پانصد تن در روز تولید داریم.» در سرآغاز شاهد مثال ظاهراً حرف اضافه «در» از قلم افتاده؛ «حدوداً» تنوین ندارد؛ «تن» نیز، چون

حرکت ندارد، بعد از تفکیک، معلوم نیست «تُن» است یا «تَن».

و امّا، در مدخل «فی» (۳۰۵۴): «مذاکره] ایران و اروپا تنها راه حل رفع سوء تفاهمات فی مابین است.» شاهد مثال به «فی مابین» مربوط است نه به «فی»؛ زیرا «فی مابین» در فارسی یک عنصر واژگانی/ واحد واژگانی است و نباید و نمی توان آن را به سه جزء «فی» و «ما» و «بین» تجزیه کرد.

از این‌ها گذشته، ذکر شاهد مثال‌های حاوی فعل مرکب یا فعل گروهی/ عبارت فعلی برای واژه بسیط (واژه‌ای که نقش فعل‌یار را در مثال دارد) در کتاب فراوان است. اینک تنها به ذکر چند نمونه از واژه‌های آغاز شده با دو حرف «آ» و «ا» اشاره می‌کنیم: در مدخل‌های آزار، اثبات، اجازه، احترام، احتیاط، احساس، احضار، اختصاص، اذعان، اذیت، اراده، ارزش، ارزیابی، اسارت، استخدام، استراحت، اشاره، اشغال، اصابت، اصرار، اضافه، اطاعت، اطمینان، اعتراض، اعتقاد، اعزام، اعطا، افتخار، افزایش، اقدام، اکتفا، امانت، انجام، انس، انصاف، اوج، اهدا، ایجاب، ایجاد، ایراد، و ایفا، همه مثال‌ها فعل مرکب یا فعل گروهی/ عبارت فعلی است؛ یعنی این‌ها: آزار دادن، به اثبات رساندن، اجازه دادن، احترام گذاشتن، احتیاط کردن، احساس کردن، احضار کردن، اختصاص یافتن، اذعان کردن، اذیت کردن، اراده کردن، ارزش داشتن، ارزیابی کردن، به اسارت درآمدن، استخدام کردن، استراحت کردن، اشاره کردن، اشغال کردن، اصابت کردن، اصرار کردن، اضافه کردن، اطاعت کردن، اطمینان داشتن، اعتراض کردن، اعتقاد داشتن، اعزام داشتن، اعطا کردن، افتخار کردن، افزایش دادن، اقدام کردن، اکتفا کردن، امانت دادن، انجام دادن، انس داشتن، انصاف داشتن، اوج گرفتن، اهدا کردن، ایجاب کردن، ایجاد کردن، ایراد داشتن، و ایفا کردن. افزون بر فعل‌های یاد شده، فعل‌هایی را مرکب گرفته‌اند که تنها کاربرد مجهول فعل‌های متعدی‌اند. برای مثال، فعل‌های آشکار شدن و استفاده شدن و تولید شدن تنها کاربردهای مجهول فعل‌های متعدی آشکار کردن و استفاده کردن و تولید کردن‌اند. اینک به نمونه‌هایی از دو حرف «آ» و «ا» به علاوه فعل‌هایی دیگر از داخل جعبه (ص ۱۱۸-۱۱۴)، به ترتیب الفبا، اشاره می‌کنیم: آشکار شدن، اختراع شدن، استخراج شدن، استفاده شدن، اصلاح شدن، اطلاق شدن، افتتاح شدن، اعدام شدن، افزون شدن، افشا شدن، اقامه شدن، امضا شدن، انتصاب گشتن<sup>(۸)</sup>، ترجمه شدن، تولید شدن، صادر شدن، کامل شدن، گفته شدن، معرفی شدن، ملی شدن.

افزون بر این، گروه کلمه‌هایی را فعل مرکب گرفته و مدخل قرار داده‌اند که مطلقاً و بر پایه هیچ نظریه معقولی فعل مرکب نیستند. کنار هم قرار گرفتن صفت و فعل ربطی «بودن» فعل مرکب نمی‌سازد. مؤلفان این‌ها را فعل مرکب فرض کرده‌اند: آگاه بودن، امیدوار بودن، حاکی بودن، درست بودن، شاهد بودن، قاتل بودن، مربوط بودن، مشغول بودن، مواجه بودن.

مؤلفان کتاب حتی «اینجا بودن» (قید + فعل ربطی) را فعل مرکب گرفته‌اند. غریب‌تر از آن

این هاست: سال بودن، سال داشتن، سالگی، ساله، ساله بودن، و ماهه. برای مثال، جمله «۱۸ ساله [با فاصله] بودم که حس مادری را تجربه کردم.» شاهد مثالِ مدخلِ سال بودن است. این‌ها مصداق‌هایی برای نهایتِ کارِ مکانیکی اند.

### ۳-۱۴. سایر اشتباهات یا افتادگی‌ها

در مدخلِ «طی» (۲۵۸) «در طی این سالها چون چرای در اصول نکرده.» در این مثال، در «چون چرای»، «و» افتاده و باید «چون و چرای» بوده باشد. «چون و چرا» واحد واژگانی مستقل (ترکیب عطفی واژگانی) و بن‌واژه است.

در مثالِ مدخلِ «ثروت» (۱۲۳۸)، که جمله‌ای کامل است، نقطه پایان را نگذاشته‌اند.

نمونه‌های مشابه در کتاب فراوان است.

اما، در مدخلِ «رجب» (۴۸۳۹)، برعکس عمل کرده‌اند؛ یعنی نقطه گذاشته‌اند: «روز سوم رجب سال دویست و پنجاه و چهار.» این مثال، که عبارتی غیر جمله‌ای است، در پایان به نقطه نیاز ندارد. نمونه‌های اشکالات نقطه‌گذاری در شاهد مثال‌های این اثر کم نیست است؛ چند مصداق بارز دیگر را در این مدخل‌ها می‌توان دید: امداد، انتقالی، انعکاس، پایه، یارانه.

### ۳-۱۵. تشدید پایانی در تلفظ بن‌واژه‌ها و مشکل ناهماهنگی

ملاک‌ها نامشخص و تعریف نشده‌اند. برای مثال، تلفظِ حق (hæv)، عام (am)، کل (kol)، محل (mæhæl)، مستبد (mostæbed)، منجر (monʒær)، و مهم (mohem) ... را بدون تشدید پایانی داده‌اند؛ اما بن‌واژه‌هایی مانند سن (senn)، متضاد (motæzadd)، مختص (moxtæss)، مختل (moxtæll)، مستمر (mostæmærr)، مسن (mosenn)، منحل (monhæll)، ... را با حرف پایانی مشدد آوانویسی کرده‌اند.

### ۳-۱۶. تلفظ‌های نامعمول یا نادر

در آوانگاری برخی بن‌واژه‌ها تلفظ‌های نامعمول، نادر، یا احیاناً گویشی هم می‌بینیم. برای مثال، تلفظِ «گریبان» (geriban: ۳۷۸۱) معمول نیست، و تلفظِ نخستِ فارسی معیار «گریبان» (gæriban) است.

### ۳-۱۷. ناهماهنگی در دادنِ تلفظِ عناصر واژگانی مشابه و هم‌ارز

تلفظِ صفت فاعلی واژه‌ها از باب «تفعل» با کسره در پایانه هجای دوم («مِت»):

متأسفانه (moteʔæssefane)<sup>۳۲</sup>، متخصص (motexæsses)، متوجه (motevæʒʒeh)، متوجه شدن (motevæʒʒeh fɔdæn)، متوقف (motevæʒʒef)، متوقف شدن (motevæʒʒef fɔdæn)، متولی (motevælli).

۳۲. افزودگیِ خطِ زیرِ /e/ (آندرلاین: e) در آوانگاریِ واژه‌ها از نگارندگان است.

این در حالی است که سایر واژه‌های مشابه (اسم فاعل از دو باب «تفعل» یا «تفاعل») را با فتحه در پایانه هجای دوم، یعنی «مُتَّ»، ثبت کرده‌اند؛ در این واژه‌ها:

متأثر (motæ?æsser)<sup>۳۳</sup>، متجاوز (motæ?ævez)، متحمل (motæ?æhæmmel)، متخلف (motæ?ællef)، متداول (motæ?ædavel)، متشکر (motæ?ækker)، متشکل (motæ?ækkel)، [متضاد/ متضاد] (motæ?æzadd)، متعادل (motæ?ædel)، متعارف [ / متعارف] (motæ?æref/ motæ?æraf)، متعال [ > متعالی] (motæ?æal)، متعدد (motæ?ædded)، متعلق (motæ?ælleʋ)، متعهد (motæ?æhhed)، متغیر (motæ?æjjjer)، متفاوت (motæ?æfævet)، متفکر (motæ?æfækker)، متقابل (motæ?ævabel)، متقاضی (motæ?ævazi)، متقاعد (motæ?æva?ed)، متمادی (motæ?æmadi)، متمایز (motæ?æmajez)، متمرکز (motæ?æmærkez)، متناسب (motæ?ænaseb)، متنوع (motæ?ænævveʋ)، متوالی (motæ?ævvali)، متوسط (motæ?ævæsset)، متوسطه (motæ?ævæssete)، متوسل (motæ?ævæssel)، متوقف کردن (motæ?ævævvef kærðæn)، متولد (motæ?ævællled)، متولد شدن (motæ?ævællled ʃodæn).

تلفظ معیار کدام است؟ این همه تفاوت برای چیست؟ در تلفظ فارسی صفت‌های فاعلی (و مفعولی) از دو باب «تفعل» و «تفاعل»، دست‌کم در فارسی معاصر و به‌ویژه در فارسی امروز، هم فتحه (/æ/) باب است و هم کسره (/e/). بنابراین، هیچ‌کدام از این دو گونه تلفظ غیر معمول نیستند و قطعاً کسی که به تلفظ طبیعی اعتقاد دارد، نمی‌تواند تلفظ با کسره را، که در میان عامه مردم — و حتی گاه در میان خواص — رواج دارد، نادرست بدانند. در ضمن، ذکر این نکته هم ضروری است که، بر پایه قاعده «همگونی»<sup>۳۴</sup>، که گونه‌ای فرایند واجی یا قانون واجی<sup>۳۵</sup> معمول درون‌کلمه‌ای در زبان است، تلفظ هجای دوم با ضمه (/o/) نیز، هم در فارسی تهرانی و هم در بسیاری دیگر از گویش‌های مناطق گوناگون ایران، کمابیش، رواج دارد؛ برای مثال، «متوقع» را بعضاً به صورت /motovævveʋ/ تلفظ می‌کنند.

### ۱۸-۳. ناهماهنگی در عناصر واژگانی «فهرست‌های موضوعی واژگان»<sup>۳۶</sup>

جعبه‌های موضوعی در سراسر بخش اصلی فرهنگ بسامدی فارسی، یعنی در بخش فهرست واژه‌ها به ترتیب بسامد و به فاصله هر چند صفحه یک جعبه، چیده شده است. برخی جعبه‌ها نیم صفحه یا یک صفحه، و بعضی دیگر بیش از یک صفحه را اشغال کرده‌اند. هر جعبه شامل گروهی

۳۳. ایضاً افزودگی خط زیر /æ/ (آندرلاین: æ) در آوانگاری واژه‌ها از نگارندگان است.

34. assimilation

35. phonological process, phonological rule

36. Thematic vocabulary lists

از واژه‌هاست که، از نظر معنایی و کاربردی، به یک، دو، یا چند موضوع نزدیک به هم مرتبط‌اند. مؤلفان مراجعه‌کننده را به این امر توجه داده‌اند که معادل‌های انگلیسی<sup>(۹)</sup> درون جعبه‌ها لزوماً با معادل‌هایی که در فهرست اصلی آمده‌اند تطابق کامل ندارند. باین حال، چون خود مؤلفان به این نکته اذعان کرده‌اند، و در بخشی از مقدمه (ص 3-4) نیز، همراه با ذکر چند مثال، به این امر اشاره کرده‌اند، تنها به بررسی جعبه نخست و مهم‌ترین واژه‌های ناساز، ناهمخوان، و ناهماهنگ آن اشاره می‌کنیم و نقد باقی جعبه‌ها را به فرصتی دیگر و مقاله‌ای جداگانه وامی‌گذاریم.

Frequency index 15

1 Animals							
Rank	Headword	Pronunciation	Gloss	Rank	Headword	Pronunciation	Gloss
351	ماده	made	female	6993	مورچه	murtʃe	ant
1434	ماهی	mahi	fish	6999	نر	nær	male
1593	شیر	ʃir	lion	7124	پشه	pæʃe	mosquito
1675	کره	korre	foal	7770	نهنگ	næhæng	whale
1927	حیوان	hejvan	animal	8091	خزنده	xæzænde	reptile
1952	پرنده	pærænde	bird	8120	روبه	rubah	fox
2761	اسب	æsb	horse	8209	کبوتر	kæbutær	pigeon
3050	پروانه	pærvane	butterfly	8421	بید	bid	moth
3405	سگ	sæg	dog	9254	خرگوش	xærquʃ	rabbit

جدول ۲) جعبه شماره ۱ از «فهرست موضوعی واژگان»: حیوانات (تصویر بخشی از بالای جعبه)

در «جعبه شماره ۲: حیوانات»، از هشت واژه نخست، پنج واژه را بررسی می‌کنیم:

۱) چنان‌که در تصویر می‌بینیم، «ماده» (با تلفظ بدون تشدید *made* و تنها در معنای مؤنث)، از نظر بسامد کاربرد، در پیکره مبنا در رتبه ۳۵۱ است و در میان واژه‌های این جعبه در جایگاه نخست قرار دارد. اشکالی که وارد است این است که در خود مدخل، «ماده» و «ماده» را با هم (*made, madde*) داده‌اند. بنابراین، اولاً «ماده» دو بن‌واژه است و ثانیاً بسامد کاربرد «ماده» از «ماده» بیشتر است.

۲) به واژه دوم، یعنی «ماهی»، هم این اشکال وارد است که تنها واژه‌ای بسیط نیست و ناظر به کاربرد صورت‌واژه «ماه [اسم] + ی [پسوند نکره]»<sup>(۱۰)</sup> نیز هست.

۳) واژه سوم، یعنی «شیر»، در این جعبه، یعنی در فهرست حیوانات، در معنای حیوان (*lion*) معرفی شده؛ در حالی‌که، در مدخل شماره ۱۵۹۳، در معنای مایع خوراکی (*milk*) هم هست. «شیر» در این دو معنی از دو ریشه متفاوت است؛ بنابراین، در هنگام برچسب‌گذاری واژگانی، باید این دو را از هم جدا کرد، و مؤلفان چنین نکرده‌اند.

۴) واژه چهارم «کره» است. کدام «کره»؟ در این جعبه، یعنی در فهرست حیوانات [و واژه‌های مرتبط]، تلفظ واژه را korre داده‌اند و معادل انگلیسی آن را foal نوشته‌اند؛ حال آن‌که، در مدخل شماره ۱۶۷۵، با این تلفظ‌ها نیز آمده: kære, kore. معادل‌های انگلیسی آن، در متن اصلی فرهنگ، یعنی در فهرست بسامدی، به ترتیب، عبارت‌اند از: butter, Korea/globe, foal؛ یعنی این چهار تا: کره (اسم، فارسی)، کره<sup>[۱]</sup> (اسم خاص، نام دو کشور کره جنوبی و کره شمالی)، کره<sup>[۲]</sup> (اسم، شکل هندسی [و معنای مرتبط، یعنی سیاره]، عربی)، کره (اسم، توله چهارپایان، به ویژه اسب و خر). این بن‌واژه‌ها و واحدهای واژگانی سه تلفظ متفاوت دارند و به چهار ریشه گوناگون متعلق‌اند؛ بنابراین، در هنگام برچسب‌گذاری واژگانی، باید این چهار بن‌واژه را از هم جدا کرد، و بازم مؤلفان چنین نکرده‌اند.

۵) واژه پنجم «پروانه» است. باز کدام «پروانه»؟ در این جعبه، یعنی در فهرست حیوانات، معادل انگلیسی «پروانه» را butterfly نوشته‌اند؛ حال آن‌که، در مدخل شماره ۳۰۵۰، با هویت دستوری اسم خاص نیز آمده؛ یعنی نام زنانه/ دخترانه. پیکره‌نگار<sup>۳۷</sup> حرفه‌ای و کوشا، در هنگام برچسب‌گذاری واژگانی، باید این دو بن‌واژه را از هم جدا کند، و البته باز مؤلفان چنین نکرده‌اند.

### سخن پایانی

در این مقاله از پرداختن به بسیاری دیگر از خطاهای مصداقی کتاب چشم پوشیدیم. اگر مجال دست دهد، در مقال دیگری به سایر خطاهای محرز اثر خواهیم پرداخت. به جرئت می‌توان اذعان داشت که، در کتاب مورد بحث، صفحه‌ای نیست که از یک، دو، و بلکه چند یا چندین خطا خالی باشد. این کتاب، به این صورتی که هست، اصلاح‌شدنی یا ویرایش‌پذیر نیست؛ زیرا مؤلفان در کاری تخصصی، هم از نگاه روش‌شناختی و هم از نظر موردی، به شیوه‌ای مکانیکی عمل کرده‌اند و کار را سهل و سبک شمرده‌اند. بنابراین، در مجموعه «فرهنگ‌های بسامدی راتلج»، برای زبان فارسی باید کسی یا کسانی دیگر به تدوین فرهنگی دیگر دست برند.

### پی‌نوشت

- (۱) در ادامه مقاله، هر جا «مثال» نوشته‌ایم، منظورمان «شاهد‌مثال/ شاهد - مثال» است؛ یعنی جمله‌ها و عبارت‌هایی که به‌عنوان «شاهد» [= citation] از پیکره به‌دست داده‌اند نه «مثالی» خودساخته [= example] خودساخته.
- (۲) مؤلفان پیکره‌های مورد استفاده خود را در سرآغاز ص ۲ و در جدولی فهرست کرده‌اند. پیکره اصلی، «پیکره معروف موسوم به «پیکره همشهری» (با بیش از ۱۴۸ میلیون رکورد) است. پس از آن، «پیکره بی‌جن‌خان» (با بیش از ۴,۲ میلیون رکورد) و سپس چند پیکره کوچک‌تر را نام برده‌اند.

(۳) مقاله «فرهنگ بسامدی فارسی» در ویکی‌پدیای فارسی را [سید] حمید حسنی نوشته و ایجاد مقاله در تاریخچه آن با نام کاربری وی (به تاریخ ۱۳ اسفند ۱۳۹۴) ثبت شده است. شباهت‌های آماری در این گفتار به همین دلیل است.

(۴) این عدد بر روی جلد چاپ نخست (۲۰۱۷) ۲۰ نوشته شده، اما در چاپ دوم (۲۰۱۸) آن را به ۳۱ اصلاح کرده‌اند.

در مقایسه دو چاپ (نه دو ویرایش) تنها اصلاح و تغییری که به چشم می‌آید، اصلاح همین عدد است.

(۵) گرچه این فهرست‌ها را به‌شکلی نامعیاری/ غیر استاندارد و کاملاً ماشینی و مکانیکی به‌دست داده‌اند. برای مثال، «آن»، که در فهرست بسامدی (در ص ۸) در رتبه ۱۳ است، در این نمایه، هم در فهرست نام‌ها آمده، هم در فهرست ضمایر، و هم در فهرست نام‌های خاص (Anne). در بند ۳-۹: «ب» به این موضوع پرداخته‌ایم.

(۶) «بن‌واژه» را ابتدا منصور اختیار (۱۳۴۸)، ص ۲۴۹: «[فهرست] واژه‌های علمی در معنی‌شناسی و زبان‌شناسی» (به‌عنوان برابر‌نهادی برای morph ساخت. سپس دیگران، چند دهه بعدتر، آن را برابر‌نهاد «لِما» (انگلیسی: lemma، فرانسوی: lemme [لم]؛ جمع: lemmata, lemmas, lemmes) قرار دادند. ما نیز «لما» را به‌کار می‌بریم. «بن‌واژه» عبارت است از صورت واژگانی و دستوری اصلی واژه که در فرهنگ لغتی معیار/ استاندارد، یک مدخل (entry) جداگانه به آن اختصاص داده می‌شود. این صورت را انتزاعی/ مجرد در نظر می‌گیریم. «صورت‌واژه» (En.: word-form/ word form, variant; Fr.: forme lexicale, variant) یا «صورت‌کلمه» یا صورت‌تصریفی نیز عبارت است از هر یک از صورت‌های عینی یا صرف‌شده بن‌واژه، به‌عنوان هسته مشترک آن‌ها، که عملاً در گفتار و نوشتار به‌کار می‌رود و نماینده صورت‌های انتزاعی ذخیره‌شده در ذهن گویشوران زبان است. بر این اساس است که بن‌واژه صورت انتزاعی دارد؛ زیرا خود آن نیز صورت‌واژه خودش است؛ مثال‌هایی از بن‌واژه و صورت‌واژه را بیشتر در تعریف «واژه» آوردیم (درباره مفهوم بن‌واژه و صورت‌واژه، نه لزوماً یا صرفاً با این عنوان‌ها، نک. بدیه‌ای، ص بیست‌وسه - سی و دو؛ Crystal: Hartmann and James: "variant"; Trask (1); ibid (2); عاصی؛ Bauer؛ کوچرا، ص ۵۲ و ۵۳؛ حسنی (۱)، ص ۱۴۳؛ همو (۲)، ص ۱۰-۱۱؛ Atkins & Rundell, pp. 162-163؛ شَقَاقی؛ طباطبایی؛ میرزانی).

(۷) از این به‌بعد، هر جا آندراسکور آمده، کار ماست و معرف فاصله نابه‌جاست. نکته: «آندرلاین»، در عالم ویرایش، خطی است که زیر حروف می‌کشیم؛ underline؛ اما «آندراسکور» نشانه «\_» است و آن را میان کلمه‌ها یا حروف می‌گذاریم. البته، کاربرد «آندرلاین» به‌جای «آندراسکور» در همه‌جا رایج است.

(۸) «انتصاب گشتن» مثال مدخل «انتصاب» است: «اعضای شورای ملی می‌توانند به‌عنوان وزیر انتصاب گردند.» این شاهد مثالی طبیعی نیست؛ گرچه ظاهراً به رسانه‌ای مثل روزنامه تعلق دارد. «انتصاب شدن/ انتصاب گشتن» در فرهنگ بزرگ سخن (انوری (۱)) نیامده؛ اما، در سندبادنامه (از ظهیری سمرقندی، تألیف‌شده در حدود ۶۰۰ هـ.ق، به‌کوشش احمد آتش، استانبول، ۱۹۴۸، ص ۳۲۵) «انتصاب نمودن» به‌معنی اقدام کردن (به) و قیام کردن (به) به‌کار رفته است (نک: دهخدا و دیگران: «انتصاب نمودن»).

(۹) مؤلفان، در ذکر معادل (های) انگلیسی، مانند برخی دیگر از تک‌نگاره‌های همین مجموعه، از لفظ gloss، یعنی توضیح و تأویل، استفاده کرده‌اند.

(۱۰) از آنجاکه، در قیاس با مصداق‌های بسیار — که پیشتر به نمونه‌هایی از آن‌ها اشاره کردیم — احتمال می‌رود برچسب‌گذاری واژگانی و دستوری در پیکره مبنا یا صورت نگرفته و یا، اگر چنین شده، این کار به‌دقت انجام نشده، «ماهی»، بالفعل و بر پایه کاربرد، چهار صورت‌واژه می‌تواند باشد: یک. «ماهی» (اسم)، دو. «ماه + ی: پسوند نکره»؛ سه. «ماه + ی: پسوند اسم‌ساز (/ در دستور سنتی: ی: پسوند حاصل‌مصدرساز)؛ چهار. «ماه + ی: فعل ربطی دوم‌شخص مفرد».

## منابع

- اختیار، منصور، معنی‌شناسی. دانشگاه تهران، تهران ۱۳۴۸.
- انوری (۱)، حسن (سرپرست)، فرهنگ بزرگ سخن، سخن، تهران ۱۳۸۱.
- \_\_\_\_\_ (۲) فرهنگ روز سخن، سخن، تهران ۱۳۸۳.
- بدره‌ای، فریدون، واژگان نوشتاری کودکان دبستانی ایران، فرهنگستان زبان ایران، تهران ۱۳۵۲.
- بی‌جن‌خان، محمود، فرهنگ بسامدی (بر اساس پیکره متنی زبان فارسی امروز)، دانشگاه تهران، تهران ۱۳۹۱.
- حسنی (۱)، حمید، «ضرورت تهیه فرهنگ‌های بسامدی برای زبان فارسی»، در کتاب ماه ادبیات و فلسفه، سال پنجم، ش ۱ (پیاپی ۴۹)، آبان ۱۳۸۰، ص ۱۴۰-۱۴۵.
- \_\_\_\_\_ (۲) واژه‌های پرکاربرد فارسی امروز (بر مبنای پیکره یک‌میلیون لغتی؛ شامل بیش از ۸۰۰۰ لغت قاموسی و غیر قاموسی)، کانون زبان ایران، ۱۳۸۴.
- دهخدا، علی‌اکبر و دیگران، لغت‌نامه، چاپ دوم از دوره جدید، دانشگاه تهران، تهران ۱۳۷۷.
- دیهم، گیتی، فرهنگ آوایی فارسی، فرهنگ معاصر، چاپ دوم (با اصلاحات)، تهران ۱۳۸۷.
- شَقایق، ویدا، فرهنگ توصیفی صرف. علمی، تهران ۱۳۹۴.
- طباطبایی، علاء‌الدین، فرهنگ توصیفی دستور زبان فارسی، فرهنگ معاصر، تهران ۱۳۹۵.
- عاصی، مصطفی، «پایگاه داده‌های زبان فارسی»، در مجموعه مقالات سومین کنفرانس زبان‌شناسی، به‌کوشش یحیی مدرّسی و محمد دبیرمقدم، دانشگاه علامه طباطبائی و پژوهشگاه علوم انسانی و مطالعات فرهنگی، تهران ۱۳۷۶.
- کوچرا، هنری، «ریاضیات زبان»، ترجمه بهروز صفرزاده، در کتاب ماه ادبیات و فلسفه، سال چهارم، ش ۷ (پیاپی ۴۳)، اردیبهشت ۱۳۸۰، ص ۵۰-۵۳.
- میرزائی، آزاده، فرهنگ توصیفی زبان‌شناسی پیکره‌ای، علمی، تهران ۱۴۰۰.
- Atkins, B. T. "Sue" & Rundell, Michael, *The Oxford Guide to Practical Lexicography*, Oxford University, UK 2008.
- Bauer, Laurie, *A Glossary of Morphology*, Edinburgh University, UK 2004.
- Crystal, David, *A Dictionary of Linguistics and Phonetics*. Blackwell, USA, 1<sup>st</sup> (1980) & 6<sup>th</sup> (2009) editions.
- Davies, Mark, *A Frequency Dictionary of Spanish; Core vocabulary for learners*, Routledge, USA/UK 2006.
- Davies, Mark & Gardner, Dee, *A Frequency Dictionary of Contemporary American English; Word Sketches, collocates, and thematic lists*, Routledge, USA/UK 2010.
- Miller, Andrew [Corey] & Aghajanian-Stewart, Karineh (2017) *A Frequency Dictionary of Persian; Core Vocabulary for Learners*, Routledge, USA/UK 2017, 2<sup>nd</sup> imprint: 2018.
- Trask, R. L. (1), *A Dictionary of Grammatical Terms in Linguistics*, Routledge, USA/UK 1993.
- \_\_\_\_\_ (2) *The Penguin Dictionary of English Grammr*, Penguin Books, UK 2000.



Ullmann, S., *Semantics. An Introduction to the Study of Meaning*. Oxford/Blackwell, UK/USA  
1962.

<https://www.english-corpora.org>

Ullmann, S., *Semantics. An Introduction to the Study of Meaning*. Oxford/Blackwell, UK/USA  
1962.

<https://www.english-corpora.org>

