

Максимів О. Й.

ПРОБЛЕМА МЕЖ СЛОВА В УКЛАДАННІ ЧАСТОТНИХ СЛОВНИКІВ ПЕРСЬКОЇ МОВИ

Статистична лексикографія перської мови репрезентована невеликою кількістю частотних словників та конкордансів. Головним чином, це частотні словники мови письменників (напр., Унсури [2], Гафеза [27], Дакікі Тусі [23]) чи окремих творів (напр., “Шахнаме” Фердоусі [8], “Святого Послання” [24], “Послання Духа” [29], творів з поетичних збірок різних авторів [20] тощо). Укладання частотних словників перської мови вважається “справово дуже необхідною і важливою” [18, с. 140]. Проте на сьогодні у перській мові не існує частотного словника “мови в цілому”, як, наприклад, словники Штейнфельдт [7] чи Засоріної [6] у російській лексикографії, або словники Кучери і Френсіса [9], чи Торндайка і Лорджа [11] в англійській. Щодо різних функціональних стилів перської мови, то для більшості з них частотних словників також немає. Краща ситуація у галузі художнього стилю, бо словники мови письменників та окремих творів індивідуально відображають його частотні характеристики. Проте і ці словники є, переважно, тільки конкордансами: значення слів у них не наводяться, оскільки їхні завдання переважно історико-літературні, а не лінгвістичні (див. [1, с. 3]).

Мета даної розвідки – унявнити існування проблеми меж слова як неподоланого етапу в укладанні частотних словників перської мови.

Єдиним словником, який може претендувати на статус частотного словника перської мови “в цілому” є словник “Уживані слова сучасної перської мови” Гаміда Гасані [21]. Розглянемо його детальніше та визначимо основні характеристики.

Названий словник укладений на базі корпусу обсягом приблизно 1 мільйон слововживань з різних функціональних стилів сучасної перської мови. Основна мета словника – створити навчальне джерело, реєстр базових, основних слів сучасної писемної перської мови для укладання навчальних словників. Основні джерела словника – тексти художнього, публіцистич-

ного та науково-популярного стилів мовлення, більшість з них опубліковані між 1379–1383 рр. сонячної гіджри (2001–2005 р. християнського літочислення відповідно). До джерел увійшли вибірки з 62-х оригінальних перських та перекладних художніх і науково-популярних творів (усього 608 350 слововживань), а також статті з 11 повних чисел газет і журналів і 7 сторінок з 7 інших номерів (усього 396 240 слововживань). Сам словник складається з двох списків – частотного і алфавітного, до яких увійшли слова корпусу з частотою 10 і більше, всього 8 438 слів.

Корпус для словника ретельно укладався з дотриманням правил формування репрезентативної вибірки: було обрано 80 великих текстів (книг, журналів і газет), або понад 500 уривків (розділів книги чи статей) [17], зверталася увага на те, щоб до корпусу не входило більш, ніж два твори одного автора чи перекладача [19].

У передмові до словника автор характеризує досвід укладання частотних словників і корпусів англійської мови і зазначає, що “існує три суттєві відмінності при укладанні корпусу текстів перської мови порівняно з європейськими: 1) тексти джерел європейських словників з погляду достовірності написання, цитації, орфографії і особливо межі і відстані між словами мають набагато більший ступінь довіри, ніж перські джерела¹; 2) на думку дослідників, характеристика різновидів граматичних конструкцій перської мови є складнішою від подібної характеристики такої мови, як англійська; 3) стильова відстань перського писемного і розмовного мовлення іноді викликає такі труднощі, що навіть писемно слабке, непідготоване перське мовлення потрапляє під вплив розмовного, і у багатьох випадках структура окремих речень цілком розпадається” [17, с. 17–18]. З першим пунктом цілком можна погодитися, оскільки має місце неусталеність орфографії перської мови, незважаючи на активну нормалізаторську роботу Академії перської мови (існують правописні (орфографічні) словники (напр., [25; 26]), видано низку окремих випусків правил перського

¹ Очевидно, це стосується, передусім, дотримання правил орфографії та написання разом чи окремо частин складних слів.

письма (напр., [22])). Але більші труднощі тут становить нечіткість меж слова у перській лінгвістичній традиції. Другий пункт також переважно стосується меж слова, а не складності граматичної будови перської мови. Щодо третього пункту, то тут має місце цілком природний процес взаємопроникнення явищ розмовного мовлення до писемного, і це притаманно не лише, виключно чи більшою мірою, перській мові порівняно з іншими.

Питання про визначення меж складних слів і словосполучень ставили перед собою як укладачі перських тлумачних словників, так і перекладних (див., напр., [5, с. 9]). Проте у тлумачних і перекладних словниках перської мови немає одностайності, напр., у поданні складних слів, утворених за допомогою префікса та іменника, слів, утворених шляхом лексикалізації ізафетних словосполучень тощо (див. [там само, с. 13]). Укладачі таких словників найчастіше вирішують цю проблему таким чином, що ті чи інші сполучення слів подають і як самостійне слово, і як словосполучення всередині відповідної словникової статті з відсиланням до основного варіанта. Для частотних словників таке рішення не є прийнятним, оскільки їхня цінність полягає саме у визначенні місця слова у ранговому списку, незалежно від того, просте це слово чи складне.

У словнику Г. Гасані письмовим словом вважалася “одиниця, що складається з однієї або кількох літер, з незалежним значенням або функцією, і має щонайменше одну морфему. Слово може бути простим, похідним або складним” [17, с. 20]. Під час диференціації слів даного корпусу укладач намагався здійснювати максимальне розчленування на складові, наскільки це не суперечить граматичним законам. На жаль, часто таке розчленування було штучним.

Більшу частину передмови (36 сторінок) автор присвятив переліку випадків, у яких приймалося рішення на користь розчленування або не розчленування складних слів чи словосполучень. У Таблиці 1 ми спробували систематизувати ці випадки і показати, що причиною багатьох рішень є невизначеність меж слова у перській граматиці і орфографії, а саме порушення принципів цільнооформленості лексичної одиниці.

Таблиця 1. Написання разом і окремо складних слів та словосполучень у “Словнику вживаних слів сучасної перської мови” Г. Гасані.

№	Розділено:	Не розділено:
1	аналітичні форми майбутнього категоричного часу дієслів; дієслова-зв'язки (нескладові часових форм, навіть якщо вони виражені особовим закінченням);	усі граматичні і часові форми дієслів; частини складних дієслів і дієслівні звороти; дієприкметники; розмовні та орфографічні варіанти усіх слів;
2	складні іменники, утворені від числівника зі словом «سالگی» sālegi чи «ماهگی» māhagi, бо не мають словникового характеру (друга частина фіксувалася зі знаком *) опущено займенники, які стоять після суфіксів множини, а також після артикля невідзначеності, проте є винятки;	усі граматичні форми (числа, особи) іменників; орфографічний та фонетичний варіанти написання займенникових енклітик після слів, які закінчуються на голосний; масдари і відносні прикметники, утворені за допомогою «ی» -yāu та «ی» -i-невизначеності; частини віддієслівних іменників (похідних чи складних);
3	конструкції зі сполучником و va 'i', обидві чи одна частини яких мають значення і вживаються окремо; якщо одна з частин не має значення або окремо не вживається, вона фіксувалася зі знаком *;	сполучникові конструкції такого типу, якщо сполучник обтяганий; якщо до таких конструкцій додавався словотворчий афікс; якщо усі їх частини не мають значення або в сучасній перській мові самотійно не вживаються;
4	означальні ізафетні словосполучення, навіть якщо вживаються зі слабкою касрою і одним наголосом, із суфіксом множини чи з частиною-активним словотворчим словом, несловникові інверсовані означення з вивідною структурою; якщо одна з їх частин не має значення або окремо не вживається, фіксувалася зі знаком *. Є ряд винятків;	складні ізафетні та безізафетні прикметники, остання частина яких походить від дієслова; складні багаточленні прикметники з ізафетом, жодна частина яких не є похідною від дієслова;

№	Розділено:	Не розділено:
5	словосполучення з двох елементів, між якими стоїть дефіс;	схожі на них дво- і тричленні словосполучення, вжиті без дефіса; скорочені і складноскорочені слова, записані літерами перського алфавіту;
6	складні прикметники та прислівники, утворені з переліченими словами у функції прийменника як першої частини чи з іншими переліченими словами як останньої частини; утворені з прийменника і одного чи двох інших слів; частини прислівників, таких як «چطور/چهطور» četowr ‘як’, «تاکنون» tākonun ‘досі’ та ін.; складні слова, перша частина яких є одним із прикметників «کدام» kodām, «هر» har або «هیچ» hič; відносні прикметники, які складаються з числівника або слова «تک» tak чи «چند» čand і пов’язаного слова, які не мають словникового характеру. Також використано символ * для певних слів і частин слів;	складні прикметники та прислівники, теж перелічено першу та останню частини; прикметники, утворені від іменника, простого дієслова чи дієслівного словосполучення; утворені з першою частиною «به» be» і «نا» nā» перед ними; які складаються з прислівника і прикметника чи іменника; частини прислівника «چگونه/چهگونه» čegune ‘як’; складні прислівники, утворені з двох нетотожних, але подібних частин; утворені від складного дієслова чи іменника і дієслова; відносні прикметники, утворені з числівника або перелічених слів і пов’язаного слова, у разі, якщо мають словниковий характер; прислівники і прикметники, які складаються з числівника і іменника, і після іменника стоїть суфікс; прикметники, утворені від числівника чи прикметника «چند» čand і іменника з прикметникотворчим словом (переважно дієслівним похідним); усі прикметники та прислівники, вжиті у формі ступенів порівняння; утворені за допомогою суфікса «ی» -i чи «ه» -e-відносності; з прикметниковим префіксом «بی» bi;

№	Розділено:	Не розділено:
7	складні слова, утворені внаслідок повторення однієї частини, яка має незалежне вживання (якщо зустрічалися з суфіксом множини або займенником, друга частина фіксувалася зі знаком *);	такі ж складні слова з додаванням однієї чи більше морфем, які творять іншу граматичну категорію або нове значення;
8	складні слова, утворені внаслідок повторення основи теперішнього часу дієслова фіксувалися зі знаком *, навіть якщо зустрічалися з показником множини чи займенником; так само частини слів, другий елемент яких є залежним і не має значення (виняток – 9 слів, через те, що вони словникові); утворені в результаті повторення однієї частини і прийменника всередині; несловникові складні слова, утворені з двох нетотожних частин і прийменника чи афікса «ـا» між ними;	складні слова, утворені в результаті повторення елемента (слова), який не має незалежного вживання; якщо до вказаних розділованих слів додавалися інші слова, які надають їм нового граматичного значення чи ى uā-відносності або масдара (віддієслівного іменника); якщо між такими словами стоїть афікс «ـا» –ā- чи «ـوا» -vā- (є 2 винятки, розчленовані і зафіксовані зі знаком *); словникові складні слова, утворені з двох нетотожних частин і прийменника між ними;
9	складні прийменники (один з них розчленований і друга частина зафіксована зі знаком *) та сполучники, які починаються з прийменника;	–
10	частини складних числівників і десяткових дробів; дробові із-афетні числівники – кінцева їх частина зареєстрована зі знаком *; числівники-нумеративи, де не мали словникового характеру, або де їхній словниковий характер був слабким; групи слів типу «یگرشته» uekrešte ‘ряд’ і под., які складаються з числівника (переважно «یک» uek ‘один’);	числівники «دویست» devist ‘двісті’, «سیصد» sisad ‘триста’ і «پانصد» pānsad ‘п’ятсот’; складні слова, які складаються з числівника або слова «تک» tak і іншого слова або слів, у разі, якщо мають словниковий характер чи є власною назвою; частини назв днів тижня;

№	Розділено:	Не розділено:
11	словосполучення і неперські звороти, які у структурі неперського тексту вживаються окремо;	арабські словосполучення з «الـ» al-; якщо у перському тексті вжиті без нього, то зберігалася орфографічна та фонетична форма слів; окремо фіксувалися усі варіанти написання і вимови арабських слів; арабські словосполучення і звороти, які у структурі перського тексту вживаються як кліше; складні слова, запозичені перською мовою з європейських чи інших мов; складні перські слова, які складаються з першої частини європейської і основної неєвропейської;
12	групи слів, таких як «بچه‌یتیم» baččeyatim ‘дитина-сирота’, «سسته‌گل» daste gol ‘букет квітів’ та ін.; «چند» čand і «چندین» čandin; слова типу «استاد» ostād ‘професор’, «خانم» xānom ‘пані’ на початку та в кінці власних імен, де не є частиною цих імен (наведено ряд прикладів); слово «ماه» māh ‘місяць’ після назви місяця; слово «روز» ruz ‘день’ після назви дня тижня; слова «رنگ» gang і «فام» fām ‘колір’ після назви кольору; якщо суфікс у прізвищі є географічною назвою чи словом, включеним до географічної назви.	слова, такі як «آن‌جا» ānjā ‘там’, «چنانچه» čopānče ‘як, оскільки’ та ін.; групи слів, таких як «درصد» darsad ‘відсоток’, «یکدیگر» yekdigar ‘один одного’ та ін.; слова – частини імен «آقاجان» āqājān, «آقاخان» āqāxān, «سادات» sādāt та ін. як словникові звороти; слова «رنگ» gang і «فام» fām ‘колір’ після слів-не назв кольорів (напр., «تیرمرنگ» tīrēgang ‘темного кольору’); двослівні і більше власні імена; усі слова, утворені за допомогою суфікса «ی» yāu-відносності чи іншої морфемі.

Як видно з таблиці, укладання частотного словника Г. Гасані висвітлило “на повний зріст” нерозв’язану проблему визначення меж слова. Ще більше, її слід вважати однією з причин недостатньої активності укладання корпусів текстів перської мови і, відповідно, частотних словників різного скерування. В обговорюваному словнику, на нашу думку, вирішення цієї проблеми є дещо штучним. Так, нечіткими є правила розчленування / нероз-

членування складних слів у пунктах 3, 4, 5, 6, 7, 8, 10 і 12, недостатньо визначені поняття “словниковості” слова (قاموسى بودن کلمه *qāmusi budan-e kalame*), “складного слова” (کلمه‌ی مرکب *kalame-ye morakkab*), “морфемі” (تکواژ *takvāž*) та “звороту” (عبارت *‘ebārat*) чи “словосполучення” (ترکیب کلمات *tarkib-e kalamāt*), що суттєво впливає на якісний і кількісний склад реєстру словника. Наприклад, у реєстрі словника з зірочкою зафіксовані як реєстрові слова *211) گُو (*211), (گُوها (52), (گویی *12), а також *277) گفت. Сумарна частотність перших трьох слів (275) і останнього (277) показує, що всі вони відокремлені від одного слова گفتگو *goftogu* ‘діалог’, а дві кінцеві форми не перевищили порогу частотності 10, і тому не увійшли до опублікованої частини словника. Також, наприклад, відповідно до наведених правил, до реєстру увійшло слово شرح حال نویس *šarh-e hāl-nevis* ‘біограф’, але не увійшло شرح حال *šarh-e hāl* ‘біографія’, бо його розчленовано і підраховано під окремими реєстровими формами شرح і حال, або є реєстрове слово قابل اعتمادترین *qābel-e-e ‘temādtarin* ‘найбільш вартий довіри’, але немає початкової форми قابل اعتماد *qābel-e-e ‘temād* ‘вартий довіри’.

Необґрунтованим залишається різний підхід до окремих складних слів. Наприклад, بی شک *bi-šakk* ‘безсумнівно’, رنگوارنگ *rang-o-vā-rang* ‘різнокольоровий’, حداقل *hadd-e-aqal* ‘принаймні, мінімум’, چطور / چطور *četowr* ‘як’, آبی رنگ *ābi-rang* ‘блакитного кольору’ (друга частина вживається у функції нумератива) і под. розчленовано на окремі слова, а بی اختیار *bi-exteyār* ‘мимоволі’, رنگارنگ *rang-ā-rang* ‘різнокольоровий’, لاقفل *lā-aqal* ‘принаймні, мінімум’, چهگونه / چهگونه *čegune* ‘як’, تیره رنگ *tire-rang* ‘темного кольору’ і под. – не розчленовано. Наведені слова відрізняються за вживанням і граматичною функцією (іноді походженням) їх складових, але в сукупності у мовленні (чи тексті) мають схоже лексичне і абсолютно однакове граматичне значення, тому їх не варто розділяти.

Відповідно до пунктів 2, 4, 6, 7, 8, 10 та 12 таблиці, формотворчі і словотворчі афікси сплутані, і в одних випадках (напр., приєднання суфікса множини до частин сполучникових зворотів, обидві з яких мають значення і можуть вживатися окремо) їх наявність дає підстави розділяти складне слово, а в інших (напр., приєднання суфіксів форм ступенів порівняння прикметників та прислівників) навпаки – фіксувати окремо і не розділяти. При-

єднання займенникової енклітики до слова дає підстави вважати його окремою словоформою, але якщо вона приєднується після суфікса множини чи артикля невизначеності, то опускається зовсім, і таке слово окремою словоформою не вважається.

Велика кількість словотворчих морфем фіксується окремо чи підсумовується до омографічних їм словоформ відповідно до пунктів 2, 3, 4, 5, 6, 7, 8, 9, 10 і 12 таблиці (напр., зафіксовані у словнику як морфеме *داد کم، سوء* та ін.), що призводить до відсутності у реєстрі словника частовживаних складних слів (напр., *سوء استعماله su'estefāde* 'зловживання', *درباره darbāre-ye* 'про'). Довгі списки винятків із зазначених правил також свідчать на користь неефективності такого підходу до розчленування / нерозчленування слів у частотному словнику. Ігноруються регулярні типи творення складних слів перської мови, зафіксовані як у академічних (див., напр., [4, с. 139–167, 204–232, 293–299; 28, с. 124–179, 193–208, 272–274, 278–309, 320–321; 31, с. 145–157, 219–225] тощо), так і в навчальних (див., напр., [3; 29, с. 8; 12, с. 15–27, 122–123, 184–190] тощо) граматиках. Загалом, не дотримано єдиного підходу до визначення меж слова, хоча перські мовознавці виокремлюють шість ознак та функцій слова: 1) фонетична єдність, 2) синтаксична єдність (те саме слово виконує певну синтаксичну функцію), 3) семантична єдність, 4) орфографічна єдність, 5) єдність вживання (певне експресивне забарвлення) та 6) єдність словозміни (див. [15]).

У лексикографічній традиції перської мови немає словника омографів, не зустрічали ми також їх списків в інших працях, навіть у граматиках про існування такого явища згадується лише декількома словами (див., напр., [13, с. 85; 14, с. III]). Передмова до словника Г. Гасані є першою відомою нам працею, у якій подано список омографів перської мови. Проте тут розмежовано тільки омографи, які відрізняються вимовою (парафони), напр. *پار par* 'непо' і *پور por* 'повний'. Повні омоніми не розмежовувалися, напр. *دوش duš* 1. перське 'плече', 2. 'вчора ввечері', 3. французьке 'душ' зафіксовані під одним заголовним словом і частота їхня підсумована. На нашу думку, це методологічно некоректно, адже під час величезної роботи ручного розписування текстів була можливість повністю врахувати омонімію, яка є основною невирішеною проблемою укладачів словників автоматичним і напівавтоматичним способами.

Усі слова у словнику Г. Гасані, зокрема омографи і запозичення, супроводжуються необхідними діакритичними знаками, вказівками на належність до певного слова чи транскрипцією. Очевидні помилки виправлено, однак, такі, що пов'язані зі вживанням слів (коли це не завдає шкоди структурі речення), залишено.

Таким чином, з одного боку, словник “Уживані слова сучасної перської мови” Г. Гасані не враховує достатньою мірою стилістичної диференціації перської мови, а з іншого – не дотримується єдиного підходу до визначення меж слова, і поряд зі словоформами фіксує морфему (що не завжди чітко можна диференціювати), а складних слів не реєструє. Усе сказане підводить до висновку, що даний словник є частотним словником словоформ і морфем, укладеним на базі декількох функціональних стилів сучасної перської мови, а не “частотним словником” (див. [16]), у європейському розумінні терміна. Як видно з таблиці, авторське (Г. Гасані) трактування цього поняття (підсумоване у передмові до його словника) відрізняється від європейського. Не викликає сумнівів цінність словника Г. Гасані для навчально-методичних потреб, проте це не є частотний словник лексем перської мови. Так само безперечним є і те, що джерельна база даного словника заслуговує на увагу, і в разі потреби дослідник, який має у своєму розпорядженні усі матеріали (а не тільки словоформи з частотою 10 і більше) і відповідні тексти, може доопрацювати їх і створити повноцінний частотний словник перської мови “в цілому”. Проте, це окреме і трудомістке наукове завдання.

Стрімкий розвиток комп'ютерної лінгвістики, укладання великих за обсягом електронних баз даних перської мови, робота над автоматизованими комп'ютерними програмами зумовили потребу вироблення чітких правил лематизації перських слів. На жаль, у цьому напрямі зроблено небагато.

Єдиним частотним словником, у якому провадилася лематизація перських слів є “Частотний словник Унсурі” М.-Н. Османова [2] (що пояснюється його формуванням як науковця в європейській лексикографічній традиції). Тут складні слова без потреби не розчленовано (напр., رنگارنگ *rangārang* ‘різнокольоровий’, چگونہ *čegone* ‘як, яким чином’ тощо), дієслова зведені до початкової форми масдара, форми множини іменників і ступенів порівняння прикметників до їхніх вихідних слів, розмежовано омогра-

фи, повні і стягнені форми того самого слова подаються в межах однієї словникової статті тощо. Безперечною позитивною рисою словника, якої на сьогодні не вдалося досягнути автоматичним і напівавтоматичним способами укладання є те, що зафіксовані окремо складені прийменники.

Вдала спроба створити алгоритм лематизації перських слів була здійснена дослідницею Т. М. Мійангаг (див. дет. [10]). Було укладено списки перських кореневих морфем (684 одиниці) та афіксів (25 одиниць) (доступні з сайту Вищої ради інформатики <http://www.scict.ir>) і розроблено алгоритм автоматичного розкладання перських слів на морфему з метою укладання програмного забезпечення перської мови (скажімо, програми автоматичної перевірки орфографії, розпізнавання слів із їх зображень тощо). Проте у цій роботі також омографи не розрізнені, основне завдання – відокремити основу від афіксів.

Отже, статистична лексикографія перської мови ще чекає своєї низки словників (щонайменше, для кожного з п'яти функціональних стилів окремо), а правила лематизації перських слів достатньо розроблені для процесу укладання частотних словників у ручному режимі. Можемо зробити висновок, що труднощі укладання корпусів і частотних словників перської мови пов'язані, передусім, з невеликим досвідом роботи у цій галузі, з невеликої кількістю досліджень, а не з “труднощами” граматики перської мови. На сьогодні не існує жодного лематизованого частотного словника сучасної перської мови чи котрогось з її функціональних стилів. На нашу думку, можливим шляхом укладання такого словника є формування реєстру словоформ з будь-якого репрезентативного для певних цілей доступного електронного корпусу, лематизація за певними стандартними правилами одержаного списку з подальшою перевіркою у конкордансах можливих омографів. Такий спосіб укладання частотного словника можна назвати напівавтоматичним.

ЛІТЕРАТУРА

1. *Османов М.-Н. О. Принципы и цели составления частотного словаря и конкорданса “Дивана” Унсури / М.-Н. Османов // Османов М.-Н.О. Частотный словарь Унсури / Магомет-Нури Османович Османов. Москва, 1970.*

2. *Османов М.-Н. О. Частотный словарь Унсури* / Магомет-Нури Османович Османов. Москва, 1970.
3. *Пейсиков Л. С. Лексикология современного персидского языка* / Пейсиков Л. С. Москва, 1975.
4. *Рубинчик Ю. А. Грамматика современного персидского литературного языка* / Ю. А. Рубинчик. Москва, 2001.
5. *Рубинчик Ю. Предисловие* / Ю. Рубинчик // **Персидско-русский словарь** / [ред. Рубинчик Ю. А.]. В 2-х т. Т. 1. Москва, 1970.
6. **Частотный словарь русского языка** / Под ред. Л. Н. Засориной. Москва, 1977.
7. *Штейнфельдт Э. А. Частотный словарь современного русского литературного языка. 2500 наиболее употребительных слов* / Э. А. Штейнфельдт. Таллин, 1963.
8. http://fa.wiktionary.org/wiki/پيوست_فهرست_بسامدی_شاهنامه_ [Электронный ресурс].
9. *Kučera H., Francis W. N. Computational Analysis of Present-day American English*. Providence, 1967.
10. *Miangah T. M. Automatic Lemmatization of Persian Words* / Tayebeh Mosavi Miangah // **Journal of Quantitative Linguistics**. 2006. Vol. 13. No. 1.
11. *Thorndike E. L., Lorge I. The Teacher's Word Book of 30 000 Words*. New York, 1944.
12. احمدی گ. دستور زبان فارسی ۱ / حسن احمدی گیوی، حسن انوری. [ویرایش ۲]. - تهران: فاطمی، ۱۳۷۴. - ۲۵۵ ص.
13. احمدی گ. ح، انوری ح. دستور زبان فارسی ۱. - تهران: فاطمی، ۱۳۷۴. - ۲۵۵ ص.
14. احمدی گ. ح، انوری ح. دستور زبان فارسی ۲. - تهران: فاطمی، ۱۳۷۶. - ۳۸۴ ص.
15. بنادکی م. س. واژه چیست؟ [Электронный ресурс] / مهدی سعید بنادکی، <http://linguist87.blogfa.com/post-448.aspx> : **Режим доступу** // ۱۳۸۸
16. حسنی ح. تشابهات و تفاوت‌های «پیکره» و «فرهنگ بسامدی» از دیدگاه فرهنگ‌نگاری / حمید حسنی // گزارش فرهنگ‌نگار زبان و ادب فارسی. دفتر اول. - تهران، ۱۳۸۴. - ص. ۸۹-۹۲.
17. حسنی ح. شیوه‌نامه پیکره / حمید حسنی // حسنی ح. واژه‌های پرکاربرد فارسی امروز (بر مبنای پیکره ی یک میلیون لغتی؛ شامل بیش از ۸۰۰۰ لغت قاموسی و غیر قاموسی) / تألیف حمید حسنی. - تهران: کانون زبان ایران، ۱۳۸۴. - ص. ۹-۵۵.
18. حسنی ح. ضرورت تهیه فرهنگ‌های بسامدی برای زبان فارسی / حمید حسنی

- // مجله کتاب ماه (ادبیات و فلسفه). - ۱۳۸۰. - سال پنجم، شماره ۱ (پیاپی: ۴۹). - ص. ۱۴۵-۱۴۰.
۱۹. حسنی ح. فهرست کلان‌متن‌های پیکره / حمید حسنی // حسنی. ح. واژه‌های پرکاربرد فارسی امروز (بر مبنای پیکره‌ی یک میلیون لغتی؛ شامل بیش از ۸۰۰۰ لغت قاموسی و غیرقاموسی) / تألیف حمید حسنی. - تهران: کانون زبان ایران، ۱۳۸۴. - ص. ۳۱۵-۳۲۲.
۲۰. حسنی ح. نرم‌افزار فارسی / حمید حسنی // مجله کتاب ماه (ادبیات و فلسفه). - ۱۳۸۱. - سال ششم، شماره ۴ (پیاپی: ۶۴). - ص. ۱۲۸-۱۲۹.
۲۱. حسنی ح. واژه‌های پرکاربرد فارسی امروز (بر مبنای پیکره یک میلیون لغتی؛ شامل بیش از ۸۰۰۰ لغت قاموسی و غیرقاموسی) / تألیف حمید حسنی. - تهران: کانون زبان ایران، ۱۳۸۴. - ص. ۳۲۱.
۲۲. دستور خط فارسی / مصوب فرهنگستان زبان و ادب فارسی. - تهران: فرهنگستان زبان و ادب فارسی (نشر آثار)، ۱۳۸۱، چاپ دوم ۱۳۸۳، چاپ سوم و چهارم ۱۳۸۴، چاپ پنجم ۱۳۸۵، چاپ ششم ۱۳۸۵، چاپ هفتم ۱۳۸۶. - ص. ۵۸.
۲۳. دیوان ابو منصور محمد بن احمد دقیقی طوسی بانضمام فرهنگ بسامدی آن / به اهتمام دکتر محمد جواد شریعتی. - تهران: انتشارات اساطیر، ۱۳۶۸. - ص. ۵۳۶.
۲۴. سنج د. واژه‌نامه بسامدی رساله القدس / پایان‌نامه حوزه زبان و زبان‌شناسی / داریوش سنج. - تهران: ۱۳۵۵. - ص. ۲۴۵.
۲۵. شعار ج. فرهنگ املائی و دستور خط و املائی فارسی بر پایه مصوبات فرهنگستان زبان و ادب فارسی / تألیف جعفر شعار. - تهران: سخن، ۱۳۷۸. - ص. ۲۵۰.
۲۶. صادقی ع. فرهنگ املائی خط فارسی بر اساس دستور خط فارسی مصوب فرهنگستان زبان و ادب فارسی / علی‌اشرف صادقی، زهرا زندی‌مقدم. - تهران: فرهنگستان زبان و ادب فارسی (نشر آثار)، ۱۳۸۵. - ص. ۶۱۹.
۲۷. صدیقیان م. د. فرهنگ واژنمای حافظ به انضمام فرهنگ بسامدی / دکتر مهین دخت صدیقیان. - [چاپ دوم]. - تهران: روزانه، ۱۳۷۸. - ص. ۱۳۵۹.
۲۸. فرشیدورد خ. دستور مفصل امروز بر پایه زبان‌شناسی جدید: شامل پژوهش‌های تازه‌ای درباره آواشناسی و صرف و نحو فارسی معاصر و مقایسه آن با قواعد دستوری / نگارش خسرو فرشیدورد. - تهران: سخن، ۱۳۸۲. - ص. ۷۰۳.
۲۹. مدرسی تهرانی ی. واژه‌نامه بسامدی رساله نفس: تألیف شیخ‌الرئیس ابو علی سینا: پایان‌نامه حوزه زبان و زبان‌شناسی / یحیی مدرسی تهرانی. - تهران، ۱۳۵۱. - ص. ۳۲۸.
۳۰. نگارش و دستور زبان فارسی. سال دوم. آموزش متوسطه عمومی. ادبیات و علوم انسانی / مؤلف علی سلطانی‌گردفرامروزی. - تهران: شرکت چاپ و نشر ایران، ۱۳۷۳. - ص. ۱۲۳.
۳۱. یوسفی ح. ع. دستور زبان فارسی ۱ و ۲ / تألیف حسین علی یوسفی. - تهران: نشر روزگار، ۱۳۷۸. - ص. ۲۸۰.