

id ipsum, Italien *desso*, comparez aussi le roumain *dinsul, dinaa*.

in directum, roumain *in dreptul* : CIL VI, 36126 *a maceria... in der(e)c(tum) ad villam. inquietare*, intransitif, roumain *a înceta* « cesser ».

machinari, à la forme active *machinare*, roumain *a măcina* « moudre » : RUF1 *De podagra 35 herbum maccinatum* ; ANTONIN. PLACENT. 34, p. 181, 12 *habentes unum asellum, qui illis macinabat*.

mica « miette », roumain *nimica* « rien » < *ne mica*.

neque unus, roumain *niciunul, niciuna* « personne » : *Vitae patrum* VI, 3, 13 (PL 73, 1012 A) *sicut anima ista non me fecit quiescere neque una hora in se* ; VII, 11, 1 (PL 73, 1033 D) *neque uno die suffere poterit*.

per diem, roumain *pe zi* : THEOD. PRISC. I, 39 ; *per noctem* roumain *pe noapte* : CGL III, 219, 57.

totum adverbe, roumain *tot* : Itin. Eg. 15, 2 *totum per vallem*, roumain *tot pe vale* ; 3, 1 *totum ad directum*, roumain *tot de-a dreptul*.

Septembre 1965

H. Mihăescu

Institut d'études sud-est européennes
Bucarest, I. C. Frimu 9

ALPHONSE JUILLAND, P.M.H. EDWARDS, ILEANA JUILLAND,
Frequency Dictionary of Rumanian Words, Mouton & Co., London, The
Hague, Paris, 1965.

The *Frequency Dictionary of Rumanian Words* belongs to a more comprehensive series of works consecrated to the lexical quantifying of the Romance languages ; the other volumes are dedicated to the Spanish, French, Italian and Portuguese languages, respectively. The linguistic investigations for all the above mentioned languages have been carried out by the same technical means and according to the same principles of method.

The dictionary is, at the same time, the *first* frequency dictionary of the Romanian language.

The volume consists of three parts :

I- an ample introduction expounding the various theoretical and practical problems to be solved when elaborating this kind of works ; II- types and tokens in alphabetical order ; III- types : usage, frequency, dispersion.

The introduction gives an ample discussion of the criteria observed on choosing the lexical material and the description of the mathematical apparatus employed.

The authors have dealt with a vast material which they have proposed themselves to be at the same time homogeneous and diverse, so that it should be really *representative* for the Romanian literary language : their computations have been carried out on a number of 20,000 sentences, totalling 500,000 words. To ensure the diverse character of the material they have proceeded to its division into five sections, as follows : (1) dramatic literature (only theatrical plays), (2) fictional literature (novels and short-stories), (3) essayistics (essays, memories, correspondence), (4) periodicals, (5) technical literature.

In connection with this classification it should be mentioned on the one hand that we don't consider as particularly useful the separation of the first two sections which are, generally, grouped together as belonging both to the belletristic genre; on the other hand, the fact that the essayistic literature is looked upon as a genre in itself is again imperfect as far as the Romanian language is concerned, bearing in mind the great diversity of works which it includes (both works with a literary character proper and works with a scientific character). Treating the essayistic genre as an independent section, the authors have grouped together works as diverse as N. Bagdasar's *Filosofia contemporană*, E. Lovinescu's *Critice*, A. Vlahuță's *România pitorească a.s.o.*

As regards the choice of texts, the authors have established a selective bibliography for the years 1890—1940, from which they have selected *at random* a series of writers and works, considered to be the most significant for the respective sections. By doing this they have probably supposed that they would obtain as objective data as possible; actually, the selected names are not always the most representative ones. For the dramatic genre they have worked for instance with texts by A. de Herz (whereas such names as Camil Petrescu or Mihail Sebastian do not even appear on the list on which one can find, instead, minor names as S. Ionescu or M. Rădulescu); for the narrative genre they have worked with texts by A. L. Moldovanu, a. s. o. The same method of selection had as a result the fact that some authors appear with more than one work within the same section. Thus, for instance, of a total number of 20 dramatic works taken into consideration, six plays belong to L. Blaga, four to B. Delavrancea and three to D. Pătrășcanu; out of a total of 24 fictions, three novels belong to M. Sadoveanu alone, Ionel Teodoreanu and V. Vlahuță appear with two each, O. Goga with two essays out of a total of 23, N. Iorga with four, etc.

A final remark concerns the fact that although it has been underlined that no work in Romanian, published since 1940 has been included in the bibliography of the dictionary, yet Zaharia Stancu's novel *Desculț* figures under the section of narrative literature, notwithstanding the fact that as it is known, it has been published after that date (1948).

The lexical units obtained from the excerption of the texts have been structured keeping in view both the criterion of frequency and of dispersion. *The frequency* indicates the number of appearances of each word within the whole of the material taken into consideration. *The dispersion* represents the distribution of the words into each of the constituent sections of the literary material. The frequency and the dispersion are in a relationship of correspondence: it is thus to be observed that words with high frequencies tend towards classes with a high dispersion. For a still more complete characterization of the words, the authors have resorted to another factor also — *the usage*. The signification they give to it is: $usage = frequency + dispersion$, indicating thus that frequency is corrected by the dispersion or else, that dispersion is corrected by the frequency. It follows that a word with a high frequency may be looked upon as less usual if its distribution is low and vice-versa, a word with low frequency may be looked upon as more usual if its distribution is high. By applying these criteria, lists of words may be drawn up which can be divided into classes ranging from the highest frequency, dispersion and usage respectively down to the lowest frequency, dispersion and usage.

Successive conditions have been set to the words in turn, in order to enable them to be included in the dictionary, for instance: frequency = at least 4, distribution = at least 3, usage = at least 2, and almost 5,000 words have been selected in this manner out of a total of 50,000 taken into consideration.

According to the authors these 5,000 words form the basic vocabulary of the Romanian language. For clarifying the notion, it has been observed that it is not the statistics of frequen-

cies which determines the structure of the basic vocabulary, and that the basic vocabulary does not separate itself from the rest of the vocabulary by means of distinctly different coefficients of frequency or distribution. The quantitative distinctions between the two units do not permit their objective isolation. The basic vocabularies may be and are selected according to external criteria, to the practical aim they are meant to, the number of their elements being variable. Henceforth, the 5,000 most frequent lexical units of the Romanian language included in the *Frequency Dictionary* may be considered as components of the basic vocabulary as they permit the understanding of any Romanian text in a ratio of almost 97 per cent.

The theoretical exposition is followed by lists of words. These latter have suggested us some observations which do not aim, of course, at exhausting the whole material of facts.

We mention from the very beginning that we cannot effect absolute quantitative comparison, between the lists of words contained in the *Frequency Dictionary* and the words entered in other lists concerning the Romanian language (A. Graur, *Inecare asupra fondului principal lexical al limbii române*, București, 1954; V. Șuteu, *Observații asupra frecvenței cuvintelor în operele unor scriitori români*, SCL 3/1959; Gh. Bolocan, *Unele caracteristici ale stilului publicistic al limbii române literare*, SCL 1/1961): The lists by A. Graur, Member of the Academy, are not based on statistical criteria, while those by V. Șuteu and Gh. Bolocan are limited to belletristic works and publicistic style respectively. We shall therefore confine ourselves, when needed, to some observations. We shall note, for instance, that in all cases, the prepositions, conjunctions, auxiliary verbs, pronouns and indefinite article, register maximum frequencies and small ranks; this is quite natural if we think that without these words no text may be formed.

In the *Frequency Dictionary*, in the series of words arranged according to their frequency, the first adjectives appearing are *acești* — R¹22, *tot* — R.30; the first verbs, *da* — R.37, *vedea* — R.38; the nouns appearing only at R.44 — *om* ($F^2 = 1.069$, $U^2 = 895.31$) and R.49 — *domn* ($F = 994$, $U = 728.01$) — and we are struck by the fact that they are entered with very close frequencies and usage percentages.

In the group of maximum frequency ($F = 1 - 500$) we point out a series of words which we consider of a much smaller circulation: *doamnă*-, *biserice*-, *dumnezeu*-, *boier*-, *județ*-, *popă*-, *zeu* —, *călugăr* —, *potfiie* —, *religios* —, *secară* —, etc.

Both in this first group and in the others, the terms belonging to the semantic spheres "army" and "religion" seem excessively represented. Here are the words (in their decreasing frequency): *biserice*-, *dumnezeu*-, *popă*-, *călugăr*-, *religios*-, *sfânt* — adj. —, *mănăstire*-, *căpitan*-, *ostăș*-, *ortodox*-, *oaste*-, *creștinesc*-, *inger*-, *general*-, *icoană*-, *sfânt* — noun —, *episcop*-, *jandarm*-, *mitropolit*-, *apostol*-, *comandant*-, *sfântă*-, *scriptură*-, *oștean*-, *colonel*-, *religie*-, *paște*-, *boteza*-, *evanghelie*-, *creștinism*-, *biblie*-, *dumnezeiesc*-, *potir*-, *sinod*-, *călugăriță*-, *luteran*-, *sacru*-, *călugărie*-, *post*-, *artilerie*-, *blagoslov*-, *catolic*-, *chilie*-, *creștină*-, *ecumenic*-, *schil*-, *teologie*-, *ortodoxie*-, *santinelă*-, etc.

We think that this fact may be explained particularly by the way in which the texts have been selected. In the very same manner do we explain also the absence in the dictionary of the words designating current notions and objects, as: *cu* and *femeie* (the most obvious absence in the *Dictionary*!) *acru*-, *băț*-, *cană*-, *creion*-, *dulap*-, *ladă* a.s.o.

Noticing, from a first reading, that the *Dictionary* does not include such terms, we have considered it as useful to compare the lists drawn up by its authors, with that of A. Graur,

¹ R — Rank

² F — Frequency

³ U — Usage

which include 1, 419 words. In this respect, we shall confine ourselves to observing which of the words considered as being part of the basic word-stock of the Romanian language (and some even of its nucleus) are not entered in the *Frequency Dictionary*: *acru, agăța, ager, alde, alintă, alloi, alună, bade, baieră, bășică, băș, boboc, bocî, boldă, brîncă, bubă, bucă, burduf, burete, burie, burîdă, bufe, calic, camion, coș, caznă, căldare, cerne, cerși, cheag, chingă, chior, cîneșă, cîrlig, cîrmă, cîrn, clate, clăți, clei, clește, clipi, cloci, coajă, colindă, comandă, coperi, crae, creș, croi, cu, cunună, curea, democrat, despica, despula, destrămă, dezmierda, dicta, doagă, dogori, duos, duminică, făg, fală, falcă, fasole, fașă, fărîmă, fâta, femeie, ferice, foarfece, friu, furnică, garant, găti, geamăn, ghem, ghimpe, greajă, hai, tapă, idee, instrui, împunge, încinge, îndemna, îngina, îngrădi, jaf, joi, june, lac, lacăt, lene, lin, lînge, lopată, luci, lumina, lumnare, lună, luntre, maimușă, mană, marfi, maș, măcelar, măcina, mălestru, mălai, măsea, mătură, mesteca, mercuri, mijl, mîneacă, mînz, moară, moare, most, monta, muc, murg, mustra, nadă, năduș, năraș, neted, nînge, oboși, obște, ogor, opera, osteni, otrăvi, ofel, papură, par, pască, pată, păcură, păduche, păgîn, păianjen, păpa, păpușă, perie, pestriș, petec, peși, pic, pieptene, pîper, pisa, piscică, pișca, piti, pîrlî, plăcea, poale, praștie, prăji, prăvălie, primeni, pripi, probă, procopsi, propti, prund, puș, pupa, purice, puși, rade, radio, raporta, rașie, răchită, rezema, rîma, rîonă, rouă, rugini, rumega, rumen, rușine, sanie, saf, sădî, sălta, sătul, scai, scaldă, scăpăra, scăpăta, scărmană, scribă, scobi, scuipe, scul, seceră, sfii, sfredel, sindicat, sine, sită, stîmătă, sirg, slut, smintî, soroc, spin, spinteca, spoi, spurca, spuză, stavilă, sterp, stingher, stîmă, stoarce, strădui, strîmî, strang, strungă, stuf, stup, suci, suge, sul, surpa, șa, șolm, șoptrî, știrb, șuiera, șurub, tăbăci, tăciune, tălmăci, tălmîie, tăvăli, telegraf, toarce, toartă, tobă, loca, toci, tocmî, topi, trăsni, treaz, treiera, trîmbișă, trudi, tufă, tuna, tunde, turba, turna, turidă, țap, țarc, țeară, țearpă, țese, țîșar, țîșă, țoi, uđ, uimi, unge, unghie, urși, urzică, ustura, vad, valma, vargă, văduș, vătămă, venin, vierme, viespe, vîneri, viscol, vișin, vltă, vina, vlnă, vlnă, vrtos, vrtă, vtagă, vopsi, vrăbie, vrajă, vulpe, zăbavă, zăcea, zătrei, zăuciuma, zburda, zdruncina, zeamă, zgtrei, zori, zvrli, zvon.*

Although such terms are absent in the *Frequency Dictionary*, the conditions set to the words considered as components of the Romanian basic vocabulary have been very large. That is why certain terms have been entered in the lists though never appearing in some sections (as it is known, a section contains a comprehensive number of texts) — ($F = 0$), or having but a sporadic presence. For instance ⁴:

	I	II	III	IV	V
<i>legifera</i>	0	2	0	1	3
<i>legiune</i>	1	0	3	2	0
<i>legună</i>	0	1	1	0	2
<i>magistrat</i>	0	1	1	3	2
<i>mobîlat</i>	0	1	1	2	0
<i>savant</i>	0	1	2	2	2
<i>spor</i>	0	1	1	1	3
<i>stîrni</i>	3	0	1	1	2
<i>uifat</i>	0	1	0	2	1
<i>valah</i>	1	1	3	0	0
<i>zăgaz</i>	2	2	2	1	0
<i>zidi</i>	1	1	1	0	3

R.S.O.

⁴ In order to simplify things, we shall use the notations: I = dramatic literature, II = narrative literature, III = essayistics, IV = periodicals, V = technical literature.

The authors have aimed at "correcting" the possible errors by taking into consideration the frequency only on the distribution alone, introducing the criteria of usage. But the very high frequency of certain words in but one section *alone* (and being terms of speciality we are inclined to think that in the respective sections also, they do appear particularly in one single work) has led to rather unexpected results. Thus, for instance :

	Frequencies					Usage
	I	II	III	IV	V	
<i>acid</i>	1	7	0	5	13	14.37
<i>ecumenic</i>	0	1	1	0	6	2.39
<i>mi</i> [the musical note]	2	1	0	8	0	3.52
<i>secară</i>	0	25	1	0	27	21.51
<i>sulfat</i>	0	2	0	1	10	3.57
<i>oxid</i>	0	5	0	1	12	6.53

A.S.O.

The figures reveal quite clearly the presence of the words above with a high frequency, particularly in technical literature; the high frequency of the musical note *mi* in periodicals is probably due to musical chronicles.

We have to mention the fact also that *after* these words with much smaller usage coefficients there occur : *camion* (2.36), *incomplet* (2.36), *salut* (2.36), *veșnic* (2.36), *coase* (2.33), *incălzi* (2.33), *asemăna* (2.26), *aplicatie* (2.13; *copie* (2.13), *lărgi* (2.13), *mihnit* (2.13), *răspîndi* (2.13), etc.

As regards the organizing of the words after the first letter, we do not see why a pseudo-phonetical criterion has been used concurrently to the graphic one. The words noted down with *c* have been divided by the authors into two groups : group *c* and *k'* together, group *ĉ* — separately (therefore, *cărare*, and *chema* are entered under the letter *c* whereas *cecușiu* under "a letter" *ĉ*); the same for *g* — *g'*/*ĝ* (*gazetă* and *ghici* together under the letter *g* and *geme* under the "letter" *ĝ*). Similar distinctions have been drawn for vowels : *i*/*î* (*indica* noted with *i*, *lepure* with *î* and *el, ea* with *e*); *o*/*o'* (*oră* noted with *o*, *oaste*, with *o'*).

This division, which is as a matter of fact incorrect, is not applied in other cases but for initials; within the words, the authors have grouped the letters observing the alphabetic order.

A similar inconsistency may be observed in fact at the semantic level as well, where the homonyms have not always been treated in the same manner; sometimes explanations are given for both the homonymous terms whereas, at other times only one of them is being defined. For instance : *vară* noun "cousin" and noun "summer", but : *părinte*, noun "—"; noun "priest", *șiret*, adj. "—"; noun "shoe lace", etc.

The *Dictionary* comprises also a series of obsolete forms : *laboratorii* (pl. of *laborator*), *recomandez*, *salele* (pl. of *sală*), nouns which nowadays appear as incorrect.

The *Orthographic and Orthoepic Handbook* of the Romanian language has not been consulted for some words the result being therefore their being written as : *alcool*, *alcolic*, *delega*, *fică*, *ștofă*, *turbure*, *vrășmaș*.

To these, printing mistakes have been added of the type : *draguțe*, *judecătaria*, *tarmvai*, or incomplete drawing up to the word *șoarece*.

By making these observations, we have aimed at presenting the reader some aspects from the *Frequency Dictionary of Rumanian Words*. It has not only the merit of being the first more comprehensive statistical study on the Romanian word-stock, but it comprises also a rich material which, to be sure, may be used, at least partially, in subsequent investigations. It was a good idea that of the authors' giving for each word the image of its distribution. It would have been no doubt preferable to dispose of precise indications concerning the texts (and not only

the sections) in which the respective terms are being attested, though this would have led to a sensible increase in the volume of the dictionary. Our wish may be justified in a single way: we consider the main shortcoming of the present work to be the selection of the literary material, selection engendering both words with surprising frequencies and, as a consequence of these frequencies, usage coefficients which do not represent the real importance of the terms of the language.

The *Frequency Dictionary* is nevertheless an interesting work which has yet to be used with reserves in order not to be led to false linguistic conclusions on the basis of the statistical data established by the authors, along the criteria we have pointed out above.

January 1966

Alexandra Roceriu-Alexanărescu

Center for Phonetical and Dialectological Research
Bucharest, Calea Victoriei, 194

A. DAUZAT, J. DUBOIS, H. MITTERAND, *Nouveau dictionnaire étymologique et historique*. Paris, Librairie Larousse, 1964, XLIX + 804 p.

Après avoir atteint la 7^e édition « revue et argumentée », le *Dictionnaire étymologique de la langue française* se présente au public sous un aspect inédit: *Nouveau dictionnaire étymologique et historique*, dont la couverture enregistre à côté du nom de l'auteur initial, Albert Dauzat, les signatures des linguistes bien connus par d'autres ouvrages remarquables: Jean Dubois et Henry Mitterand.

Par rapport à l'ancien *Dictionnaire étymologique...* ce nouvel ouvrage présente, sans pour autant faire accrotre le nombre des pages du dictionnaire proprement dit, de nombreuses augmentations et remanements. Chaque mot-souche et en général chaque mot à l'intérieur d'un article est suivi de la date et de la source où il se trouve attesté, de son orthographe et de son sens primitifs (si ceux-ci présentent des différences par rapport au stade actuel de la langue), de ses sens, par ordre de leur importance (datés eux aussi), des sources de ceux-ci et, enfin, de l'étymologie.

Du fait que les auteurs ne considèrent pas comme unique préoccupation de l'étymologie d'identifier et de dater les formes les plus anciennes du mot et d'en expliquer l'origine, mais qu'ils entendent qu'elle est appelée aussi à tracer, brièvement, l'évolution sémantique du mot, leur œuvre est un dictionnaire historique.

Ainsi que les auteurs le précisent dans la préface, l'espace consacré aux termes d'usage régional et aux termes archaïques a été réduit, pour enregistrer, par contre, « la totalité du vocabulaire général employé de nos jours aux divers niveaux de l'idiome, du français académique au français populaire, ainsi que l'essentiel des lexiques techniques et scientifiques contemporains, répertoriés jusque dans leurs termes les plus récents » (p. V). Laisant de côté le fait qu'il est impossible pour un dictionnaire de la langue vivante d'être « exhaustif », c'est-à-dire de consigner la totalité du vocabulaire général, on pourrait observer que le présent dictionnaire a enregistré, avec une minutie digne d'éloges, les termes les plus récents. Si, par exemple, l'édition précédente donnait pour *électrique* 8 formes dérivées et composées, rédigées dans l'espace de 11 lignes, la nouvelle édition consacre à *électrique* environ trois fois plus de place (soit 31 lignes) embrassant 25 formes, dont certaines très nouvelles (comme *électronique*, *électronicien*, etc.). Il en est de même pour *télé*: l'édition précédente s'arrêtait aux formes enregistrées jusqu'en 1882 et, en fait d'appareils, à *téléphone* (1834), tandis que dans