FREQUENCY ANALYSIS OF THE USE OF A NUMBER OF MORPHOLOGICAL VARIANTS IN SCIENTIFIC LANGUAGE

Constantin Manea[1], Gheorghe Săvoiu[2]

[1,2]University of Piteşti, Faculty of Economic Sciences, [2]University of Piteşti, Faculty of Letters, Romania
e-mail: [1]kostea_m@yahoo.com, [2]gsavoiu@yahoo.com

Abstract. *This paper presents a statistical study of morphological variants in scientific language. The major content represents the identification of the variants, and the frequency analysis of these variants. The essential source for the beginning of such an important and modern investigation is the Internet.*

Key words: *science, scientific research, scientific language, morphological variants, descriptive statistics, frequency analysis, Internet.*

1. INTRODUCTION

In any natural language, **usage** seems to be an unwritten law, a law which will ultimately have to be observed by the current generation – or the future generations… Unfortunately, usage is simultaneously (and more often than not) the source of blatant illegality. Moreover, usage seems to be (legitimately) the very opposite of *etymology*: "*When people object to the way a word has taken on a new meaning, they usually appeal to the word's history (or etymology) for support. The older meaning, it is said, is the 'correct' meaning. (…) It is fascinating to trace the changes in meaning which have taken place, but this should not lead us to condemn new senses, and to keep old senses artificially alive. Etymology is never a true guide to meaning. To believe the opposite is to engage in the 'etymological fallacy'*".[1]

So, in the natural evolution of a language, usage, alongside the "natural" linguistic law postulated by Saussure (the phonetic body of the words in a language tends to be degraded over the course of time), is the key factor.

2. MORPHOLOGICAL VARIANTS AND FREQUENCY AN ALYSIS IN SCIENTIFIC LANGUAGE

Statistical concepts are applicable (and provably very useful) in virtually all fields of research and types of scientific approaches; this idea can be hardly challenged. In linguistics, analyzing texts (or corpora) by means of statistical factors (recurrence, frequency, weight, etc.) has always been relevant, and sometimes even crucial. Numerous instances can be evoked when the contribution of statistical analysis applied to linguistics eventually made the difference between *life and death* (David Crystal mentions one such example in his book *The English Language*: someone accused of murder and nearly sentenced to death by hanging, was finally saved from the gallows after the jury corroborated the evidence with the linguistic and statistical / frequency analysis of the text of a letter wrongly attributed to him – where frequency of individual words and phrases was compared, to make up an overall style-and-language picture…).

As far as usage and its significance in standardizing language are concerned, two main tendencies have been confronting for the past centuries, at least in Europe: absolute *tolerance* vs. absolute *regulation*, but reality faces us with a virtual infinity of grayish shades. In some countries, regulatory permissiveness is the rule (e.g. in Denmark, where the official dictionary of the national language is updated and recast every two or three years, based on data collected from real speaking subjects – by adding or changing variants, be they morphological or phonetic), while countries like France (and, to a certain extent, Romania) illustrate the opposite tendency: academics are always right… In this respect, the case of language standardization in Britain can be said to be intermediary.

On the other hand, the linguist himself/herself must justify his/her mission as an *expert*: he/she is actually an informed, knowledgeable user of the language, and, at the same time, the authorized "co-author" of the standardized system of the literary / "accepted" language, a construct in constant evolution, in permanent construction. As a matter of fact, "cultivating", "grooming" the national language, and especially the scientific and technical variety of it, is by no means an end in itself, but a collectively useful pursuit, carried out for the common benefit: the closest comparison that comes to mind is growing a plant, and, moreover, trying to guide its growth, at least to a certain human extent. One of the best, most complex dictionaries in the English-speaking world is J. C. Wells' *Pronouncing Dictionary*, where the concern for studying and glossing the words of the English language in as objective a manner as possible, based on observation and generalization of a statistical type, is paramount. A magisterial work, a feat of scrupulous lexicography work and applied linguistics, this dictionary is quite admirable for the sheer seriousness and volume of work involved. (To give only a few concrete examples, the pronunciation of words like *grievous*, *anemone*, *hyperbole* is indicated together with a conventional sign warning against their common mispronunciations; similarly, terms like: *crescent* and *increase* are accompanied by meticulous, scientifically elaborated poll panels indicating actual preference for one of the respective variant pronunciations).

Comparison of language with a living *organism* (whereby language is likened to a living entity, usually a plant or animal – v. Humboldt, etc.) may sound nice and favourable, but it is rather risky, since the living world is, in fact, the realm of indeterminism *par excellence*, since it can imply abrupt changes or leaps, unmotivated shifts and modifications, inconsistencies, even *aberrations* (v. the definition of the term *aberration*: 'an action or event that is different from what usually happens or what someone usually does' – *www.ldoceonline.com*) [3], etc.On the other hand, considering the evolution of language, one cannot but notice its systemic character. Saussure himself talked about the evolution of that system, highlighting the role of analogy, by dint of which local changes will ultimately lead to the overall alteration of the system – as many diachronic mutations in, subsequent to small, gradual accumulations along the axis of synchrony.

The present paper tries to capitalize on the constructive virtues of interdisciplinarity in harnessing statistics to linguistics analysis. It is virtually a study case relating to a batch of technical (i.e. 'learned', scientific) English words that exhibit certain structural inconsistencies, thereby posing problems for (both native and) non-native speakers of English.[3]The reason why the authors especially chose technical (or "specialized") terms, out of the quite substantial vocabulary of the corpus of learned words making up the vocabulary of modern sciences (i.e. the words of cultural relevance) may seem obvious. Anyway, the full explanation is that these words, which are supposed to belong to a lexicon segment definable, in stylistic terms, as (rather) "formal" vocabulary, i.e. an allegedly more "select" circle of terms, are comparably harder to affect by the deviations from the norm (which is most often determined etymologically, and sometimes demonstrably and analogically)… The logical inference would be that it would be normal for the use of language to have a more limited influence on them. (But is this the actual reality?)

The essential challenge for the present study is to analyze the (analogical) spreading or regular though non-etymological forms, which can show that regularization can be seen as tantamount to a (very natural) simplification of the morphological and phonematic components of the language systems. A secondary task is to study the alternative / variant spelling forms, an issue which, in the context, may prove almost as significant for (possible) generalizations in the sub-field investigated, and also demonstrating a certain attitude of the speakers towards the language, its use and evolution.

The mini-corpus that was subject to observation and analysis amounts to less than 25 items, all of which are terms belonging to the scientific/technical vocabulary of contemporary English, as shown in the table below. Two search engines were used: *Google* and *Ask*, and the corpora accessed were made up of academic writing on the net.[4]

Words (multiple plural forms)	Google search results	Ask search results	Observations
Apsides	96,300	8,330	
Apses	427,000	76,300	Relevant
Apsises	12,200	-	
Octopuses	651,000	225,000	Relevant
Octopodes	124,000	12,800	
Octopi	523,000	133,000	
Addenda	5,700,000	430,000	Relevant
Addendums	468,000	105,000	
Addendas	115,000	-	
Criteria	445,000,000	42,800,000	Relevant
Criterions	511,000	131,000	
Criterias	679,000	244,000	
Antennae	7,570,000	832,000	*Irrelevant – further search engines are needed (note+)*
Antennas	2,840,000	4,770,000	
Apexes	416,000	67,600	
Apices	607,000	193,000	Relevant
Apparatus	169,000,000	12,700,000	Relevant
Apparatuses	7,670,000	515,000	
Appendixes	2,580,000	264,000	
Appendices	17,400,000	2,150,000	Relevant
Aquariums	29,300,000	3,210,000	Relevant
Aquaria	9,790,000	1,260,000	
Automatons	528,000	1,260,000	
Automata	29,200,000	1,450,000	Relevant
Bureaux	73,700,000	1,150,000	*Irrelevant – further search engines are needed (note ++)*
Bureaus	30,600,000	3,210,000	
Cerebellums	44,800	6,510	
Cerebella	393,000	50,200	Relevant
Curricula	17,700,000	2,440,000	Relevant
Curriculums	6,860,000	694,000	
Formulas	57,000,000	8,200,000	Relevant
Formulae	13,100,000	1,940,000	
Genera	96,300,000	5,450,000	Relevant
Genuses	117,000	15,900	
Hiatuses	279,000	42,500	
Hiatus	34,900,000	4,140,000	Relevant
Maximums	2,860,000	407,000	
Maxima	131,000,000	6,640,000	Relevant
Minimums	8,030,000	1,080,000	
Minima	63,500,000	1,340,000	Relevant
Nuclei	22,200,000	4,050,000	Relevant
Nucleuses	73,200	8,270	
Phenomena	68,900,000	10,800,000	Relevant
Phenomenons	462,000	99,900	
Syllabuses	532,000	144,000	
Syllabi	4,930,000	799,000	Relevant
Strata	65,600,000	4,060,000	Relevant
Stratums	336,000	24,200	
Vortexes	486,000	113,000	
Vortices	2,610,000	448,000	Relevant

Note+: The search for **antennae** returned **110,710** results on academic data bases http://anelis1.summon.serialssolutions.com.ux4ll8xu6v.useaccess control.com/search?s.q=antennae The search for **antennas** returned**1,010,699**results. Note ++: The search for **bureaux** returned 159,398 results and the search for **bureaus**returned209,729 results.

Table no. 1. Assessment and frequency distributions of the grammatical plurals appropriate in accordance with the majority use

The charts below, which are graphic representations of the distributions studied, can help, through their better visibility, to assess the claim of such or such plural form to grammatical appropriacy, where quantitative information is relevant; frequency of occurrence could be analyzed by using multiple search engines (e.g. Google and Ask).
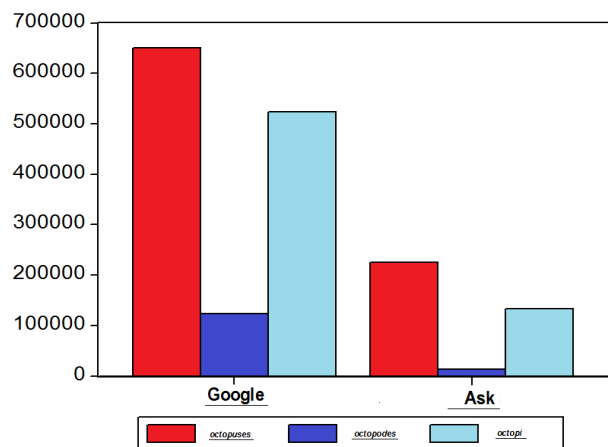

Fig. no. 1 An example of graphical analysis of frequency for the plurals variants of **octopus**

The terms included in the (parallel) list (c) illustrate a lexical class that cannot be called specifically technical or scientific terms, although they are certainly learned terms: [5]
(a) **antenna** [æn'tenə] – *pl.* **antennae** /**antennas**
apex ['eipeks] – *pl.* **apices** / **apexes** ['æpiˌsi:z / 'ei-]

apparatus [ˌæpə'reitəs, -'rɑːtəs, 'æpəˌreitəs] – *pl.***apparatus** / **apparatuses**
appendix [ə'pendiks] – *pl. anat.* **appendixes** / *pl.***appendices** [ə'pendiˌsi:z] „annexe".

automaton [ɔː'tɔməˌtɔn, -tən] – *pl.* **automata** [ɔː'tɔmətə] / **automatons**
cactus ['kæktəs] – *pl.* **cactuses** / *spec.* **cacti** ['kæktai]
calyx ['keiliks, 'kæliks] – *pl.* **calyxes** / **calyces** ['kæliˌsi:z,

'keili-]

cerebellum [ˌseri'beləm] – *pl.***cerebellums** / **cerebella** [ˌseri'belə]

cerebrum ['seribrəm] – *pl.* **cerebrums** / **cerebra** ['seribrə]

chateau / **château** ['ʃætəu] – *pl.* **chateaus** / **châteaux** ['ʃætəuz]

cicada / **cicala** [si'kɑ:də / si'kɑ:lə] – *pl.* **cicadas** / **cicalas**; **cicadae** [si'kɑ:di:] / **cicale** [si'kɑ:lei]

colloquium [kə'ləukwiəm] – *pl.* **colloquiums** / **colloquiua** [kə'ləukwiə]

contralto [kən'træltəu, -'trɑ:l-] – *pl.* **contraltos** / **contralti** [kən'trælti, -'trɑ:l-]

cranium ['kreiniəm] – *pl.* **craniums** / **crania** ['kreiniə]

criterion [krai'tiəriən] – *pl.* **criteria** [krai'tiəriə] / **criterions**

curriculum [kə'rikjuləm] – *pl.* **curricula** [kə'rikjulə] / **curriculums**

dilettante [ˌdili'tɑ:nti] – *pl.* **dilettantes** / **dilettanti** [ˌdili'tɑ:nti]

discus ['diskəs] – *pl.* **discuses** / **disci** ['diskai]

fauna ['fɔ:nə] – *pl.* **faunas** / **faunae** ['fɔ:ni:]

flora ['flɔ:rə] – *pl.* **floras** / **florae** ['flɔ:ri:]

formula ['fɔ:mjulə] – *pl.* **formulae**['fɔ:mjuˌli:] / **formulas***(chem., mat.; also fig.)*

flamingo [flə'miŋgəu] – *pl.* **flamingos** / **flamingoes**

fresco ['freskəu] – *pl.* **frescoes** / **frescos**

fungus ['fʌŋgəs] – *pl.* **fungi** ['fʌŋgai, 'fʌndʒai, 'fʌndʒi] / **funguses**

genus ['dʒi:nəs] – *pl.* **genera** ['dʒenərə] / **genuses**

halo ['heiləu] – *pl.* **haloes** / **halos**

hiatus [hai'eitəs] – *pl.* **hiatuses** / **hiatus**

iambus [ai'æmbəs] – *pl.* **iambi** [ai'æmbai] / **iambuses; iamb** ['aiæm, 'aiæmb] – *pl.* **iambs**

larynx ['læriŋks] – *pl.* **larynges** [lə'rindʒi:z] / **larynxes**

libretto [li'bretəu] – *pl.* **librettos** / **libretti** [li'breti:]

memorandum [ˌmemə'rændəm] – *pl.* **memorandums** / **memoranda** [ˌmemə'rændə]

novella [nəu'velə] – *pl.* **novellas** / **novelle** [nəu'velei]

nucleus ['nju:kliəs] – *pl.* **nuclei** ['nju:kliˌai] / **nucleuses**

palazzo [pə'lætsəu] – *pl.* **palazzos** / **palazzi** [pə'lætsi]

phenomenon [fi'nɔminən] – *pl.* **phenomena** / **phenomenons**

radius ['reidiəs] – *pl.* **radii** ['reidiˌai] / **radiuses**

radix ['reidiks] – *pl.* **radices** ['reidiˌsi:z] / **radixes**

retina ['retinə] – *pl.* **retinas**/**retinae** ['retiˌni:]

rhombus ['rɔmbəs] – *pl.* **rhombuses**/**rhombi** ['rɔmbai]

stratum ['strɑ:təm] – *pl.* **strata** / **stratums**

syllabus ['siləbəs] – *pl.***syllabi** ['siləˌbai] / **syllabuses**

tableau ['tæbləu] – *pl.* **tableaux** ['tæbləu, 'tæbləuz] / **tableaus**

tempo ['tempəu] – *pl.* **tempos** / **tempi** ['tempi:]

trapezium [trə'pi:ziəm] – *pl.* **trapeziums** / **trapezia** [trə'pi:ziə]

vacuum ['vækjuəm] – *pl.* **vacuums** /**vacua** ['vækjuə]

vertebra ['və:tibrə] – *pl.* **vertebrae** ['və:tibri:] / **vertebras**

vertex ['və:teks] – *pl.* **vertexes** / **vertices** ['və:tiˌsi:z]

vortex ['vɔ:teks] – *pl.* **vortices** ['vɔ:tiˌsi:z] / **vortexes**

(b) **bureau** ['bjuərəu] – *pl.* **bureau**/ **bureaux** ['bjuərəuz]

flamingo [flə'miŋgəu] – *pl.* **flamingos** / **flamingoes**

fresco ['freskəu] – *pl.* **frescoes** / **frescos**

grotto['grɔtəu] – *pl.* **grottoes** / **grottos**

halo ['heiləu] – *pl.* **haloes** / **halos**

manifesto [ˌmæni'festəu] – *pl.* **manifestoes** / **manifestos**

memento [mi'mentəu] – *pl.* **mementos** / **mementoes**

motto ['mɔtəu] – *pl.* **mottoes** / **mottos**

(c) **aquarium** [ə'kwɛəriəm] – *pl.* **aquariums** / **aquaria** [ə'kwɛəriə]

bureau ['bjuərəu] – *pl.* **bureau**/ **bureaux** ['bjuərəuz]

candelabrum [ˌkændi'lɑ:brəm] (*rare sg.***candelabra** [ˌkændi'lɑ:brə]*) – pl.* **candelabra**, **candelabrums**

cicerone [ˌsisə'rəuni, ˌtʃitʃ-] – *pl.* **cicerones** / **ciceroni** [ˌsisə'rəuni, ˌtʃitʃ-]

colossus [kə'lɔsəs] – *pl.* **colossi** [kə'lɔsai] / **colossuses**

focus ['fəukəs] – *pl.* **focuses** / **foci** ['fəusai]

grotto['grɔtəu] – *pl.* **grottoes** / **grottos**

gymnasium [dʒim'neiziəm] – *pl.* **gymnasiums** / **gymnasia** [dʒim'neiziə]

hippopotamus [ˌhipə'pɔtəməs] – *pl.* **hippopotamuses** / **hippopotami** [ˌhipə'pɔtəˌmai]

manifesto [ˌmæni'festəu] – *pl.* **manifestoes** / **manifestos**

maximum ['mæksiməm] – *pl.* **maxima** ['mæksimə] / **maximums**

memento [mi'mentəu] – *pl.* **mementos** / **mementoes**

millennium [mi'leniəm] – *pl.* **millenniums** / **millennia** [mi'leniə]

minimum ['miniməm] – *pl.* **minima** ['minimə] / **minimums**

motto ['mɔtəu] – *pl.* **mottoes** / **mottos**

narcissus [nɑ:'sisəs] – *pl.***narcissuses** / **narcissi** [nɑ:'sisai, -'sisi:]

persona grata [pə:'səunə 'grɑ:tə] – *pl.* **personae gratae** [pə:'səuni: 'grɑ:ti:] / **personas gratas**

referendum [ˌrefə'rendəm] – *pl.* **referendums** / **referenda** [ˌrefə'rendə]

sanatorium [ˌsænə'tɔ:riəm] – *pl.* **sanatoriums** / **sanatoria** [ˌsænə'tɔ:riə]

symposium [sim'pəuziəm] – *pl.* **symposiums** / **symposia** [sim'pəuziə]

terminus ['tə:minəs] – *pl.* **termini** ['tə:minai] / **terminuses**

ultimatum [ˌʌlti'meitəm] – *pl.* **ultimatums** /**ultimata** [ˌʌlti'meitə]

Further, special searches could be conducted, in a future contribution, along the following lines:

1. Is there any possibility for Google searches to be conducted resulting in detecting plural forms that are different from the dictionary, normative forms below?

addendum [ə'dendəm] **addenda** (maybe, forms like **addendas** or **addendums**)

alga ['ælgə] (*rarely; – pl.* **algae** ['ældʒi:]*)* (maybe, forms like **algas**)

ampulla[æm'pulə]– *pl.***ampullae**[æm'puli:] (maybe, forms like **ampullas**)

apsis ['æpsis] – *pl.* **apsides** [æp'saidi:z, 'æpsiˌdi:z] (maybe, forms like **apses** – different from the plural form of *apse* – or *apsises*)

bacterium [bæk'tiəriəm] – *pl.* **bacteria** [bæk'tiəriə] (maybe, forms like **bacteriums**)

barman ['bɑ:mən] – *pl.* **barmen** (▲*rare* **barmans**) (forms like **barmans**)

bateau [bæ'təu] – *pl.* **bateaux** [bæ'təuz] (maybe, forms like **bateaus**)

cortex ['kɔ:teks] – *pl.* **cortices** ['kɔ:tiˌsi:z] (maybe, forms like **cortexes**)

dynamo['dainəˌməu] – *pl.* **dynamos** (maybe, forms like **dynamoes**)

erratum [i'rɑ:təm] – *pl.* **errata** [i'rɑ:tə] (maybe, forms like **erratums**)

larva ['lɑ:və] – *pl.* **larvae** ['lɑ:vi:] (maybe, forms like **larvas**)

libra ['laibrə] – *pl.* **librae** ['laibri:] (maybe, forms like **libras**)

locus ['ləukəs, 'lɔk-] – *pl.* **loci** ['ləusai, 'ləuk-](maybe, forms like **locuses**)

pharynx ['færiŋks] – *pl.* **pharynges** [fə'rindʒi:z] (maybe, forms like **pharynxes**)

quantum ['kwɔntəm] – *pl.* **quanta** ['kwɔntə] (maybe, forms like **quantums**)

rondeau ['rɔndəu] – *pl.* **rondeaux** ['rɔndəu, 'rɔndəuz] (maybe, forms like **rondeaus**)

spectrum ['spektrəm] – *pl.* **spectra** ['spektrə] (maybe, forms like **spectrums**)

stimulus ['stimjuləs] – *pl.* **stimuli** ['stimjulˌlai, -ˌli:] (maybe, forms like **stimuluses**)

NOTES

(1) The only term that can be proved to have three plural forms is **octopus** (*octopuses, octopodes, octopi*). A similar search concerning the corpora distribution of *apsis* (*apsides, apses, apsises*), *addendum* (*addenda, addendums, addendas,*) and *criterion* (*criteria, criterions, criterias*), possibly also *frustrum*,would also be instructive.

(2) In some cases, usage, as captured by dictionaries, has been so decisively significant as to make the compilers gloss (as *MacMillan* does) irregular plural forms as the legitimate lemmas (e.g. *bacteria*, *criteria*, *alga*, *data*…).

(3) When the issue of meaning is also considered, the procedure used in the present contribution becomes virtually insoluble (how on earth could anyone check the meaning and use of every single occurrence as appearing in the texts searched?): e.g.

genius ['dʒi:niəs, -njəs] – *pl.* **geniuses** („musical, national etc."; *fig.*); **genii** ['dʒi:niˌai] („spirit, in Roman mythology" *also fig.*, "ghost, demon; djinn")

domino ['dɔmiˌnəu] – *pl.* **dominoes** („game") / **dominos** („article of dress")

index ['indeks] – *pl.* **indexes** / **indices** ['indiˌsi:z]

stamen ['steimen] – *pl.* **stamens** / **stamina** ['stæminə]

milieu ['mi:ljə:] – *pl.* **milieus** / *franc.* **milieux** ['mi:ljø](the French for mis very formal)

calculus ['kælkjuləs] – *pl.* **calculuses**; *med.* **calculi** ['kælkjulai]

polypus ['pɔlipəs] – *pl.* **polypi** ['pɔlipai]; *(also* **polyp***: zool., anat.)*

data ['deitə, 'dɑ:tə] *(pl., although usually considered uncontable)* → *sg.* **datum** ['deitəm, 'dɑ:təm]< or **datum** *(originally, sing.; today hardly ever used)* – *pl.* **data** (although it is provable, by empirical means, that virtually nobody ever uses *datum* today!),

agenda [ə'dʒendə] *(pl. of* **agendum***; now considered sing.)*; **agendum** [ə'dʒendəm] – *pl.*, *rare* → **agenda**

2. Terms not included in the list of the technical/scientific words, some of them having traditional or analogical irregular plurals (e.g. those in the sets "names of fish/birds/hunted or stock animals", plus the nouns *fruit* and *fish*); the asterisked words represent mere issues of spelling:

antelope['ænti ˌləup] – *pl.* **antelopes** / **antelope**
***banjo** ['bændʒəu] – *pl.* **banjos** / **banjoes**
***buffalo** ['bʌfəˌləu] – *pl.* **buffalo** / **buffaloes**
***calico** ['kæliˌkəu] – *pl.* **calicoes** / **calicos**
***cargo** ['kɑ:gəu] – *pl.* **cargos** / **cargoes**
carp [kɑ:p] – *pl.* **carp** / *rare* **carps**
catfish ['kætˌfiʃ] – *pl.* **catfish** / *rare* **catfishes**
charr / char [tʃɑ:] – *pl.* **charr** / **char**, *rare* **char(r)s**
cherub ['tʃerəb]– *pl.* **cherubs**/**cherubim** ['tʃerəbim,-ubim]
chub [tʃʌb] – *pl.* **chub** / **chubs**
coalfish ['kəulˌfiʃ] – *pl.* **coalfish** / *rare* **coalfishes**
cod/ **codfish** – *pl.* **cod** / **codfish** / *rare* **cods** / **codfishes**
coley ['kəuli, 'kɔli] – *pl.* Δ **coley**–**coalfish**.
crayfish ['kreiˌfiʃ]– Δ*AmE* **crawfish**–*pl.***crayfish**/**crayfishes**
dogfish ['dɔgˌfiʃ] – *pl.* **dogfish** / *rare* **dogfishes**
***dwarf** [dwɔ:f] – *pl.* **dwarfs** / **dwarves**
fish – *pl.* **fish** / *rare* **fishes** „varieties, kinds"
fruit – *pl.* **fruit** / *rare* **fruits** „kinds"
gazelle [gə'zel] – *pl.* **gazelles** / **gazelle**
***ghetto** ['getəu] – *pl.* **ghettos** / **ghettoes**
grouse [graus] – *pl.* **grouse** / **grouses**
hake [heik] – *pl.* **hake** / **hakes**
hoof [hu:f] – *pl.* **hoofs** / **hooves** [hu:vz]
***mosquito** [mə'ski:təu] – *pl.* **mosquitoes** / **mosquitos**
perch – *pl.* **perch** / **perches**
pike – *pl.* **pike** / **pikes**
ruff – *pl.* **ruffs** / *rare* **ruff**
saithe [seiθ] – *pl.* **saithes** / *rare* **saithe**
salmon ['sæmən] – *pl.* **salmons** / **salmon**
sardine [sɑ:'di:n] – *pl.* **sardine** / **sardines**
***scarf** [skɑ:f] – *pl.* **scarfs** / **scarves**
seraph ['serəf] – *pl.* **seraphs** / **seraphim** ['serəfim]
snipe – *pl.* **snipe** / **snipes**
springbok, *rare* **springbuck** – *pl.* **springbok** / **springboks**
squid [skwid] – *pl.* **squid** / **squids**
sturgeon ['stə:dʒən] – *pl.* **sturgeons** / Δ *rare* **sturgeon**
***tobacco** – *pl.* **tobaccos** / **tobaccos**
***tornado** [tɔ:'neidəu] – *pl.* **tornadoes** / **tornados**
trout – *pl.* **trout** / **trouts**
***wharf** [wɔ:f] – *pl.* **wharves** / **wharfs**
whiting – *pl.* **whitings** / *rar*whiting
wildebeest ['wildiˌbi:st, 'vil-] – *pl.* **wildebeests** / **wildebeest**
zander – *pl.***zander** /*rare* **zanders**
zebra ['zi:brə, 'zebrə] – *pl.* **zebras** / **zebra**
zucchini [tsu:'ki:ni, zu:-] – *pl.* **zucchini** / **zucchinis**

In addition to the purely statistical survey, a further subdivision would be in order, where the "purely phonetic vs. morpho-phonemic" subtypes of plural (mainly of the kinds: *tornadoes* / *tornados*, and *wharfs* / *wharves*) would be supplemented by the classes of the invariable plurals (for names of animals, e.g. *wildebeest*, *snipe*, *carp*, *trout*, etc.), and the bookish plurals (which are seldom regularized), vs. the class of the regularized plurals (e.g. *seraphim* / *seraphs*)

It is easily noticeable that, in such cases, usage notes are necessary even for native speakers of English, e.g.: ● "In Latin, **data** is the plural of **datum** and, historically and in specialized scientific fields, it is also treated as a plural in English, taking a plural verb, as in *the **data were** collected and classified*. In modern non-scientific use, however, despite the complaints of traditionalists, it is often not treated as a plural. Instead, it is treated as a mass noun, similar to a word like **information**, which cannot normally have a plural and which takes a singular verb. Sentences such as ***data was*** (as

well as ***data were***) *collected over a number of years* are now widely accepted in standard English"; • "Although **agenda** is the plural of **agendum** in Latin, in standard modern English it is normally used as a singular noun with a standard plural form (**agendas**)"; • "The word **media** comes from the Latin plural of **medium**. The traditional view is that it should therefore be treated as a plural noun in all its senses in English and be used with a plural rather than a singular verb: *the media **have** not followed the reports* (rather than 'has'). In practice, in the sense 'television, radio, and the press collectively', it behaves as a collective noun (like **staff** or **clergy**, for example), which means that it is now acceptable in standard English for it to take either a singular or a plural verb"; • "Based on the Latin forms, the correct singular is **candelabrum** and the correct plural is **candelabra**. However, these forms are often not observed in practice: the singular form is assumed to be **candelabra** and hence its plural is interpreted as **candelabras**. In nearly 50 per cent of the examples in the British National Corpus the singular is incorrectly given as **candelabra**". (At other times, the problems posed by the form (and grammar) of certain foreign nouns are even more complex, *e.g. octopus*: "The standard plural in English of **octopus** is **octopuses**. However, the word **octopus** comes from Greek and the Greek plural form **octopodes** is still occasionally used. The plural form **octopi**, formed according to rules for Latin plurals, is incorrect)". All the dictionary quotations are from *The* NEW OXFORD *Dictionary OF* ENGLISH.

3. CONCLUSION

In the same context of applying statistics to linguistic research, we think it would be worth studying, in point of usage and frequency of use, such hybrid semantic variants (or 'barbarisms') occurring in Romanian, which are either borrowed from English or not, as: *oneros*, *intrepid*, *vocal*, *versatil*. Unfortunately, it would be impossible to search for pronunciation variants as phonetics cannot be recorded in net texts; so, the only option for that kind of study is to *methodically* conduct linguistic surveys.It is apparent that standardization, very much like explaining, detailing and systematic description of natural languages, also possesses an important conventional/conventionalized component, which often relies on imposing constraints, simplifications, or even forcing common logic, and sometimes on grossly denying historical and etymological evidence and / or flagrant abuse.

In this country at least, everything seems to be sacrificed for the idea of clarity, nice round (scientific) conformity and parallelism. (Here are just a few examples: *aşază*, *înşală*, *ceapă* ['tʃapə] / *cepe* ['tʃepe], *Coreei*, *o imparicopitată*, *chimen* (considered to be masculine), *datorită* (usually described today as having a "positive" implication, vs. *din cauza / pricina*), spelling the Romanian sound [î] in two ways, as *î* and *â*, using the ghost-word *sunt*, etc.).

Thus, statistical distribution studies like the example above should be undertaken for such controversial issues, in order to be able to reach really objective, reliable results (let us compare them with the rules issued from valid analogies, e.g. why the imperative *nu fi* has only one *i*, unlike *să nu fii*). And then, what should the common user of the language believe – especially if they happen to make their own judgment, operating with some basic comparisons, i.e. analysis based on analogy? He/she is left practically at the mercy of chance, allegedly guided by a coterie of self-proclaimed linguistic *gurus*.Obviously, dictatorship of usage (or "common" use), under the guise of maximum tolerance (which proves to be, in fact, a kind of laxity drifting to absolute values or implications), is not likely to be a wise option. Therefore, we believe, based on experiments like the one illustrated above, that a well-advised, moderate, objective intervention by the linguist should be welcome, provided that it is made in good faith, based on the rules of logic (especially analogy) and common sense, and true respect for the inherent practical use of the language (which actually represent reality, life itself), a use that has to be observed, recorded and analyzed by such desirable scientific methods as the methods provided by statistics. The final result should be placed somewhere in the area traditionally called *the golden mean*.

REFERENCES

[1] Crystal, D., (1989), *The English Language*, pp.42-43.
[2] Longman Dictionary of Contemporary English – *www.ldoceonline.com*
[3] Manea, Constantin, Florentina Enescu, (2013).*A proposal for developing a text editing checking software material based on a complex contrastive lexicographic database*, in the proceedings of the international conference*Electronics, Computers and Artificial Intelligence* (ECAI), 2013, University of Piteşti
[4] Săvoiu G., (2013), *Modelarea economico–financiară* (*Economic and Financial Modelling*), University Publishing House, Bucharest
[5] *The* NEW OXFORD *Dictionary OF* ENGLISH.