

ISI's Master Dictionary
Aids Scientific Etymology
And Reflects Changes in Science

Number 4, January 23, 1978

Between 1973 and 1976 *happiness*, *ethics*, and *self-image* increased. *Violence* doubled; *terrorism* and *soldiers* more than tripled. *Apollo* went down. *Famine* and *cannibalism* increased. But we had more *beef*, *chocolate*, *yogurt*, *horseradish*—and, naturally, more *garbage*.

Has *happiness* really increased? Only in an abstract sense. The fact is that from 1973 to 1976 occurrences of the word *happiness* more than doubled in article titles covered by ISI®'s data base.

One of the ways changes in the activities of scientists are reflected is by the words used in journal articles. Since ISI uses title words—as well as other bibliographic descriptors—to index each article added to our data base, we are able systematically to observe and quantify these changes. It is becoming increasingly clear that this data has practical applications in lexicography.¹

The vocabulary of science, like that of all living languages, is constantly changing. New words are coined to describe new substances,

improved processes, and previously undiscovered phenomena.² Extant words fall into disuse as interest wanes in the areas they describe.

In our quality-control procedures we use several manual and computer routines to verify each title word before it becomes a part of the data base.^{3,4} One of these routines makes use of a machine-readable file dubbed the *Unique Word Dictionary (UWD)*. Actually, the choice of this name illustrates the process by which new words are coined. No one at ISI remembers who first used the phrase "Unique Word Dictionary" to describe this particular word list. But apparently someone who worked with the list coined the name, someone else began using it, and before long "Unique Word Dictionary" had become a part of ISI's in-house jargon.

The Unique Word Dictionary can be considered a kind of subject heading authority list. It is essentially a "master list" of correctly spelled terms that have been certified as being "real" words. Real words are defined as words that have been used in the titles of pre-

vious journal articles and have been verified as accurate and authentic by our editors.

As we add new articles to the data base, the computer compares the words in the new titles to the words in the *UWD*. If a new word does not appear on the list, it is tagged for further evaluation. Some of the non-matching words are simply spelling errors. Some are old terms that can be found in standard reference works. Others can only be classified as valid words after our editors contact the author of the article or the editor of the journal in which it appeared. Eventually, all tagged words are either corrected or added to the dictionary.

During 1977 about 300,000 different terms were used in the titles of articles we indexed. Of these, we estimate that about 500 were used for the first time.

Some of these terms are new or unusual combinations of prefixes and suffixes with known stems. For example, *paracollege* was used in an education journal to describe an experimental satellite institution at St. Olaf's College, Minnesota. While the average reader could not know the precise meaning of *paracollege* without a proper definition or explanation, a new word like *pathobiochemical* is readily understood.

Some of the words, however, are completely new combinations of letters, such as *waxicon*. A "nonlinear waxicon" is an optical component used in lasers. It consists of a complementary pair of cone-like mir-

rors. On the other hand, *thermvision* is a temperature-controlled camera system. It's one of those words that only makes sense after you've been told what it means. *Eutonia* refers to good muscle tone, and requires no elaborate explanation. However, *ethphon*, used in horticulture, is 2-chloroethylphosphonic acid. Like many "trivial" names in chemistry, it is meaningless by itself.

A dictionary of new scientific words, assembled with the help of ISI's Unique Word Dictionary, might prove to be a valuable tool for researchers as well as for publishers of scientific books, librarians, and laymen. Such a dictionary might be updated monthly to keep up with the changing vocabulary of science.

Apart from words that occur for the first time, a dictionary of "new" scientific words should eventually take into account new uses of existing words, whether alone or in combination. One can identify such uses by examining the new contexts in which words appear. For automatic indexing and translation systems, new uses for old words probably cause even more grief than the use of idiomatic expressions, which are in fact less ambiguous. Deciphering the phrase "on the beam" is less of a problem than deciding whether the word "beam" alone refers to a beam of light or the beam of a roof!

The figure on pages 8-9 contains 172 words selected from the *UWD*. The frequency of these words has changed significantly from 1973 to 1976. When examining the list,

however, it is important to keep in mind that some variations may be the result of adding and deleting new journals to the ISI data base. The size of the data base increases each year, so frequency can be expected to increase for many words. Overall, the number of source articles in the ISI data base increased by approximately 30% between 1973 and 1976. Therefore, for this highly selective list we have chosen mainly words whose change in frequency is greater than 30%.

For example, *charm* increased from 6 occurrences in 1973 to 109 in 1976. This seems to indicate increasing interest in *charm*—a property of subatomic particles. But we cannot be certain that the word *charm*—or any other word on the list—was always used as we might expect. *Charm* could have been used not only in physics titles but also in titles dealing with physical attractiveness, good luck amulets, music, or enchantment.

Another such word is *nude*, which increased from 33 occurrences in 1973 to 129 in 1976. We strongly suspect that the increase in frequency reflects increasing experimental use of the nude mouse, a hairless mutant. The nude mouse has no thymus and no cell-mediated immune system. Thus it easily accepts skin and tumor grafts. Since it was first described in 1966, the nude mouse has become increasingly popular in cancer and immunology research, according to Norman Reed of Montana State University, Bozeman. Obviously, the use of the

word *nude* might also refer to people who don't wear clothes! But apparently that is not a popular area of research.

Between 1973 and 1976 *fusion* increased while *fission* decreased, as did *resistivity*. In other areas of physics, *lepton*, *hadron*, *monopole*, and *positron* all increased significantly.

The dramatic decline in the U S space program between the late sixties and early seventies is reflected in the decreasing frequency of the various Apollo designations. *Apollo 11*, the flight that landed man on the moon for the first time, took place in July 1969. *Apollo 12* occurred in 1969, *Apollo 14* and *15* in 1971, and *Apollo 16* and *17* in 1972. Altogether, the word *Apollo* with various flight numbers was used 231 times in 1973—but only 46 times in 1976. *Viking 1*, the Mars probe launched on August 20, 1975, was, of course, not mentioned in 1973; but occurred 18 times in 1976.

In medicine, such words as *cancer*, *mastitis*, *nephropathy*, *scurvy*, and *spasm* at least doubled in frequency. *Herpes* almost doubled, and *diabetes* significantly increased. Surprisingly, *sickle-cell* declined. *Euthanasia*, which was used only 13 times in 1973, occurred 62 times in 1976. *Cryosurgery*, *myeloblastic*, *neonate*, and *neurosurgical* at least doubled. *Histocompatibility*, which refers to the capacity of tissues to accept or reject grafts, increased from 114 to 223.

In immunology and biochemistry, *counterimmunoelectrophoresis* in-

Figure 1. Selected words from ISI's master dictionary, which includes all words used in the titles of articles covered by ISI's data base. Each word's frequency of occurrence in 1973 and 1976 is indicated.

WORD	1973	1976	WORD	1973	1976
aerosol (ized) (s)	454	630	earthquake (s)	238	407
affirmative	13	74	ethic (al) (s)	263	470
algorithm (ic) (s)	505	737	euthanasia	13	62
amniocentesis	30	46	fallout	50	31
androgyny (ous)	0	15	famine	18	55
Apollo	85	21	feminism (ist) (ists)	35	75
Apollo 11	9	0	fiber-optic (s)	55	193
Apollo 12	5	0	fission	505	423
Apollo 14	10	0	fusion	460	743
Apollo 15	23	0	garbage	4	16
Apollo 16	29	8	gene (s)	1405	930
Apollo 17	70	17	genocide	5	21
apricot (s)	10	24	grass	130	299
backache	6	14	hadron (s)	127	181
BCG (Bacillus Calmette-Guerin)	106	259	handicapped	158	289
beef	339	622	happiness	14	33
bibliography (ies)	236	556	herbicide (al) (s)	83	309
biofeedback	39	98	herpes	272	509
biopharmaceutical	2	13	hexachlorophene	68	33
biorhythm (s)	0	11	histamine	381	530
boycott (s)	2	33	histocompatibility	114	223
braille	2	16	hockey	2	25
breastfeeding	14	36	horseradish	50	147
cancer	1547	3257	insemination	43	128
cannibalism (istic)	7	22	integration	440	669
charm (ed)	6	109	invention	14	58
chocolate	11	27	kelp	8	20
clone (s) (ing) (ed) (al)	194	339	laetrile	0	19
cocaine	44	91	L-dopa	226	162
constipation	12	23	lemming (s)	4	19
copyright (s)	16	63	leprosy	100	191
counterimmunoelectrophoresis	13	38	lepton (s)	57	117
cryosurgery	22	58	linguistic (s)	167	378
cybernetic (s)	31	67	manure	67	142
deoxyribonucleic (ase) (ease) (s) (eate) (des-) (DNA)	3100	3767	marijuana	99	160
desegregation	27	70	mastitis	63	142
diabetic (s)(es)	1383	2140	microcirculation	60	128
diarrhea	188	303	microsurgical	17	55
dopa	75	38	mitogen (ic) (s)	133	225
dopamine	289	620	monopole	27	65
			morphometric	64	127
			myeloblastic	14	67

WORD	1973	1976	WORD	1973	1976
myeloid	111	259	seminar (s)	121	338
nephropathy	16	159	sexism	10	46
neonate (s)	69	148	sex-role	18	46
neurochemical	21	58	sexual	492	815
neurosurgical	23	74	sexuality	65	216
neurotoxicity	15	37	shellfish	9	33
nude	33	129	shrimp (s)	60	147
olefin	72	132	Siamese	5	20
parity	82	141	Siberia (n)	39	99
pharmacokinetic	64	157	sickle-cell	63	37
plasmid (s)	69	231	silage	78	161
positron	96	170	simplex	166	353
prostaglandin (s)	706	1661	sister (s)	43	107
quark (s)	115	219	sludge	134	282
radioimmunoassay	458	724	smog	24	61
rehabilitation	312	768	snail (s)	134	245
renal	2784	4494	snake (s)	124	220
renin	359	650	socialism (ist) (istic) (ists)	211	684
renin-angiotensin	35	61	sociobiology	0	34
renomedullary	6	24	socioeconomic	150	298
renormalization	59	124	soldier (s)	30	91
resistivity	515	361	somatomedin	20	78
reticulum	195	295	somatostatin	11	218
retinopathy	72	143	song (s)	42	97
retrograde	125	233	spasm	51	115
rheumatic (toid)	600	929	spectrin	0	23
rhizosphere	18	48	sporozoa	7	21
ribonucleic (case) (s) (eates) (RNA)	2388	2228	squamous	84	143
ringspot	43	20	stereopsis	5	17
rosette	46	108	sterilization	164	253
rural	345	730	subpopulation (s)	49	267
saccharine	9	26	sulfobromophthalein	18	30
sarcoplasmic	100	193	terrorist (s) (ism)	10	68
science	1339	2056	thalidomide	27	18
scientific	462	864	transactional	12	47
scintigraphic (y)	171	416	transcendental	15	45
sclerosis (ing)	324	559	transsexual (s, ism)	18	40
scurvy	8	16	tumor (s)	3496	5116
seawater	95	160	vasectomy (ies) (ized)	119	80
secretin	99	176	vasopressin	124	217
security	202	392	Viking I	0	18
segregation (ed)	169	334	violence (violent)	143	308
self-awareness	4	14	winter	211	435
self-help	13	47	woman ('s) (en) (en's)	1139	2312
self-image	7	23	yogurt (yoghurt)	9	35
self-sufficiency	2	25	Zambia	35	83
semantic (s)	161	305	Zen	6	25

creased from 13 occurrences in 1973 to 38 in 1976. *Radioimmunoassay*, a technique for measuring concentrations of substances developed by 1977 Nobel Prize winner Rosalyn S. Yalow, increased from 458 to 724. In 1973 the word *somatostatin*, the name of a chemical which inhibits the release of growth hormone, was used only 11 times. But in 1973 it occurred 218 times. Strangely, *dopa* and *L-dopa* decreased significantly from 1973 to 1976, but *dopamine* increased from 289 to 620! *Neurochemical*, *neurotoxicity*, and *prostaglandin* at least doubled; *renin*, *secretin*, and *vasopressin* almost doubled.

Looking at the drug names on the list, it is interesting to observe that *laetrile*, which did not occur at all in 1973, occurred 19 times in 1976. *Apricot*, from which *laetrile* is made, also more than doubled. *Biopharmaceutical* increased from 2 to 13 occurrences, and *cocaine* more than doubled. *Thalidomide*, the drug responsible for the birth defect tragedies between 1959 and 1962, decreased as a term in article titles from 27 in 1973 to 18 in 1976. *Hexachlorophene* also decreased significantly.

In genetics, *clone* almost doubled in frequency, but *DNA* "only" increased from 3100 to 3767. Surprisingly, *gene* decreased from 1405 to 930, and *RNA* decreased from 2388 to 2228.

In biology, *kelp*, *mitogen*, *rosette*, and *sporozoa* at least doubled. *Ringspot*, a plant virus, decreased by more than half.

In the area of ecology, *smog*, *sludge*, and *garbage* all increased dramatically.

Among words that seemed relevant mainly to the social sciences, we found that the use of the term *segregation* almost doubled, from 169 in 1973 to 334 in 1976, and *integration* increased significantly, from 440 in 1973 to 669 in 1976. Of course, these words are also used in mathematics and in other areas. The word *affirmative* increased in use from 13 to 74, possibly because of increased interest in affirmative action programs. *Feminism* increased from 35 to 75, and *sexism* from 10 to 46. *Woman* increased from 1139 to 2312, and *sister* from 43 to 107.

Socialism, *genocide*, *violence* and *terrorism* at least doubled. So did *linguistics* and *cybernetics*. *Bibliography* and *semantics* almost doubled. *Copyright* more than tripled, as did *fiber-optics*, a novel method for transmitting electronic signals by using clear glass fibers to conduct laser light.

According to the data presented in the figure, several of the listed words were never used at all in 1973. This is not completely accurate, however. At the end of each calendar year, words which occurred less than three times are purged, or removed, from the Unique Word Dictionary. But even if such words as *androgyny*, *biorythm*, and *sociobiology* did occur once or twice in 1973, their use increased enormously by 1976.

Changes in the frequency of oc-

currence of a new word give us one indication of scientific activity. It is important to note, however, that the citation record of the publication in which the word was first used is often a better indicator of activity.

Consider the word *quark*, which refers to the subatomic particles which constitute hadrons. From 1973 to 1976 occurrences of the word *quark* almost doubled. But even before this increase was noted, one could have determined the tremendous amount of interest in the field by examining the citation record of the primordial paper in which Murray Gell-Mann introduced the term.⁵ This classic paper, with a total of 637 citations to date, was being heavily cited as early as 1965. Incidentally, Gell-Mann borrowed the term from the line, "—Three quarks for Muster Mark!" in James Joyce's *Finnegan's Wake*.⁶

If asked to name the basic "unit" of scientific communication, I think

that most scientists, myself included, would mention the journal article or the book. About twenty years ago I described the article as a "micro-unit of thought" and the book as a "macro-unit of thought."⁷ Somehow my use of *macro* was picked up by Webster's *Third New International Dictionary*.⁸ So I suppose that by extension a word, a phrase, or even a cited reference might be regarded as submicro- or submacro-units of thought, depending upon their frequency. In any case, the communication of scientific results is accomplished mainly by the written word.

Scholars have long studied the evolution of words and language, not only as a joy in itself, but as a window on culture and society. In the same way, scientific etymology can help us gain a better insight on the development of science.

REFERENCES

1. Garfield E. Jabberwocky, the Humpty-Dumpty syndrome and the making of scientific dictionaries! *Current Contents*[®] No. 41, 10 October 1973, p. 5-6.*
2. ———, On routes to immortality. *Current Contents* No. 21, 22 May 1974, p. 5-7.*
3. Garfield E, Koenig M & DiRenzo T. ISI data-base-produced information services. *IEEE Trans. Prof. Commun.* PC-20(2): 95-9, September 1977. Reprinted in: *Current Contents* No. 45, 7 November 1977, p. 5-17.
4. Garfield E. The *Permuterm*[®] Subject Index: an autobiographical review. *J. Amer. Inform. Sci.* 27(5/6):288-91, 1976. Reprinted in: *Current Contents* No. 12, 21 March 1977, p. 5-10.
5. Gell-Mann M. A schematic model of baryons and mesons. *Phys. Lett.* 8:214-5, 1964.
6. Joyce J. *Finnegan's wake*. New York: Penguin, 1967, p. 383.
7. Garfield E. Citation indexes for science: a new dimension in documentation through association of ideas. *Science* 122:108-11, 1955.
8. Gove P B et al., eds. *Webster's third new international dictionary of the English language, unabridged*. Springfield, MA: Merriam Co., 1961, p. 1354.

*Reprinted in: Garfield E. *Essays of an information scientist*. Philadelphia: ISI Press, 1977. 2 vols.