Abbreviations and Acronyms in English Word-Formation
Author(s): Garland Cannon
Source: American Speech, Vol. 64, No. 2 (Summer, 1989), pp. 99-127
Published by: Duke University Press
Stable URL: http://www.jstor.org/stable/455038

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

The American Dialect Society and Duke University Press are collaborating with JSTOR to digitize, preserve and extend access to American Speech.
ABBREVIATIONS AND ACRONYMS IN ENGLISH WORD-FORMATION

GARLAND CANNON

Texas A&M University

This is the fourth article in a series about categories of English word-formation—the first three articles treating functional shift, blends, and back-formations (Cannon 1985, 1986; Cannon and Bailey 1986). Collections of abbreviations and acronyms and the considerable scholarship on these two categories will be reviewed, possible solutions to presently untenable definitions and other problems will be offered, and a large corpus of recent items in the two categories will be analyzed. This study has the advantage of drawing on vastly more numerous written data and building on a comprehensive earlier study—Algeo’s article (1975), which summarizes the overlapping terminology and the scholarship to that date. This study will update his data and findings, especially in light of the large corpus that we will describe. It will simultaneously update the overall analysis of all 21 categories in Cannon’s total 13,683-item corpus (1987), and make some comparisons with other categories. The results may shed light on lexicology and lexical theory, the dynamic relations between writing and speech, and the seeming state of the art, as it were, for abbreviations and acronyms.

History and Sources

We will begin with the history and the sources of our data, before turning to the scholarship, terminology, and taxonomy. It is well known that such items go back several millennia, with abbreviations even occurring in Sumerian. The desire to economize is seen in numerous Hebrew examples like MILH ‘Mi Iolh Lnv Hshmilh (Who shall go up for us to heaven?)’ and Roman ones like SPQR ‘Senatus populusque Romanus’ and INRI ‘Jesus Nazarenus Rex Judaeorum’. Old English borrowed from Latin what became the modern ampersand. The seventeenth-century English cabal was associated with five committee members’ names, of which their first letters could be arranged to spell a variant of the borrowing kabbalah (Clifford, Arlington, Buckingham, Ashley, and Lauderdale—see Algeo 1975, 217–18).

Early forms of dictionaries of initialisms (a variously defined term that we will employ as a general rubric for acronyms and abbreviations) appeared in the fifteenth century. Perhaps the first was the famous Modus
Legendi Abbreviaturas (1475?), 277 folio leaves of two 38-line columns that recorded many visual devices rather than true lexemes/words, as they were evidently pronounced as the whole word instead of as a reduced form. Thus there are gle ‘generale’ and nobe ‘nobiscum’. There were Walther’s 459-page folio Lexicon Diplomaticum (1745) and Feutry’s Manuel tironien (1775), consisting of 424 pages of visual devices like agrsn ‘agression’, with clippings like Aar ‘Aaron’ but few true abbreviations like c ‘ce’. In his Preface he noted that his purpose was to correct “le défaut de leur uniformité,” which was “un très grand inconvénient.” Chassant’s 136-page Dictionnaire des abréviations (1846) continued use of both the ultimately Latin word abbreviation and the classical practice of treating clippings as abbreviations. Like Modus Legendi Abbreviaturas, it was widely used in Europe, going into a fifth edition in 1884. The first book-length collections of English initialisms may have been Courtenay’s 3,000-item, 53-page pamphlet and Macgregor’s 40-page one. Both published in 1855, they included numbers of clippings like Brit. and chron. T.W.W.’s general English collection (1873) is noteworthy because it is one of the few abbreviations dictionaries where entries are given citations. Next came Fallows (1883) and the anonymous Dictionary of Abbreviations (1886) containing nearly 2,500 items. Abbreviations were beginning to serve specialized groups, as shown by Martin’s Latin and French abbreviations used in English historical manuscripts (1892), and by Cordingley’s mercantile collection (1902).

So many abbreviations were being used so frequently in England that De La Rue (1867) compiled a brief list for the Smithsonian Annual Report, including common items like A.B., D.C.L., and J.P. American abbreviations like C.O.D., N.G., O.K., and P.D.Q. were appearing, a procedure which Mencken (1919, 22) proudly described as “the characteristic American habit of reducing complex concepts to the starkest abbreviations.” By the late nineteenth century, additional hundreds were appearing. One of the earliest American collections was Wilson’s, which added “Abbreviations and Representative Letters” in the second edition (1850). Though the list was in the appendix to this very popular book on punctuation, it was still prominent as of the thirty-first edition in 1899. So many collections were appearing that Rogers (1913) commented in his preface on the “many smaller but most useful works, both English and foreign,” which he incorporated into his 30-year project.

The number of abbreviations was expanded more systematically during World War One, as Americans had long been using reduced forms to name business, governmental, and daily activities (Riordan 1947, 108). Kirby (1918) divided his military abbreviations into American, British,
French, German, and general ones. Among many clippings were abbreviations like A.B. & Q. ‘appropriation for barracks and quarters’. Contemporary word lists in Dialect Notes contained a sprinkling of abbreviations like U ‘university’, C ‘cocaine’, and M ‘morphine’ (e.g., Wittman 1914, 125; Wells 1922, 182). Long (1915) collected 32 abbreviations like w.c. There were a few British acronyms like D.O.R.A., Waac, and Anzac. The 1930s witnessed more deliberate, larger supplies from F.D.R.’s alphabet agencies; but acronyms were still comparatively few. Contemporary college initialisms included Y, Y.M.C.A., R.O.T.C., and S.A.E. amid many clippings like dorm and trig, but few like Deke ‘D.K.E., Delta Kappa Epsilon’ (Schultz 1930).

The real explosion in systematically created initialisms began with World War Two. The many new items, including informal and obscene ones like F.O. and S.O.S., prompted the U.S. War Department’s two official lists, particularly Technical Manual 20-205. The Manual became part of Army Regulations, with a succession of updated revisions entitled Authorized Abbreviations, Brevity Codes, and Acronyms (1985). This lists the official initialisms and the policy and places for their use. The most regulated of the many specialized collections, it includes numerous long items like USAMSMA DHS (which surpasses the longest item in our corpus) ‘United States Army Medical Service Meat and Dairy Hygiene School’. Meanwhile, American Speech became a treasure house of the new items in World War Two. The U.S. Army’s seven-page “Glossary of Army Slang” (1941, 163–69) recorded a scattering of items like A.A. About half of the Army and Navy Journal’s radar nomenclature was abbreviations (1945, 309–10). Hornstein’s two collections (1945, 150–51; 1946, 73–75) had a few like KIA and others like DP’s, as did Dunlap’s “G.I. Lingo” (1945, 147–48). Berger (1945, 258–64) included numerous abbreviations in an illuminating article, as did Riordan’s “G.I. Alphabet” (1947, 108–14) and Fleece’s compounds made from FU (1946, 70–72). Russell’s “Among the New Words” (1946, 137–45, 220–26) listed many more. Howson (1945, 125) regretted the prematurity of Shankle’s 207-page Current Abbreviations (1945), as the war was generating such items daily. Clearly, F.D.R.’s alphabetical names had prepared Americans for understanding and creating initialisms in large numbers.

Russell (1946, 220) noted that Merriam added 72 initialisms to the 1945 supplement to Webster’s Second (1934) that were not in the 1939 one, and speculated that they had now become part of the general American lexicon. He extrapolated from several sources: Webster’s Second’s table of 5,000 abbreviations was less than 1% of the 550,000 entries stated to be in that dictionary, Mencken’s Supplement I contained only about .8%
abbreviations, and Parry's *War Dictionary* (1942) contained 1.8%. Yet the two 1945 Yearbooks contained 14.5%. Russell concluded that if even half of these survived, such items would be playing the largest role in word formation to date (1946, 138–39). In 1943–66, the Merriam Company and an American Dialect Society committee compiled lists of new words and new meanings, which was published annually in the *Britannica Book of the Year* and later classified according to date (Versand c. 1973, 179–221). The total 202 acronyms included 71 in 1943–47, 44 in 1948–52, 37 in 1953–57, 30 in 1958–62, and 20 in 1963–66.

Robbins (1951) collected 126 oral and written initialisms from aviation alone. Greatly larger numbers have since belied the several assertions that initialisms are a minor lexical process which mainly produces proper names that are a minor source of new words peripheral to the general vocabulary (e.g., Barber 1964, 97; and Huddleston 1984, 229). Certainly some older items like *Ala* and *secy 'secretary'* shed no light on morphology or lexical theory; they belong in encyclopedias, atlases, and other reference works. However, Thomas Matthews Pearce expansively predicted that the atomic age will make English a language of acronyms (quoted in *Time*, 11 Aug. 1947, 22); and 17% of Bryant's space-exploration collection (1968, 172–81) were initialisms. Most were abbreviations like *AAP* and *AVCS*; her 37 initialisms were in second place behind the 103 compounds. Crowley and Sheppard's 1987 edition of Gale's *Acronyms, Initialisms, & Abbreviations Dictionary* had the subtitle "A guide to more than 400,000 acronyms, initialisms, abbreviations, contractions, alphabetic symbols, and similar condensed appellations," most of which were identified with the United States. Today, Everyperson is skillfully creating initialisms with gusto.

Moreover, they have become productive, though their affixed forms, by comparison with derivations from other kinds of words, often produce slang or a very informal tone in school and family groups, usually gaining -y or -ie, -er, -o, or -s (Quirk et al. 1985, 1584). Cannon's corpus (1987) tabulates 177 new items formed from initialisms utilizing most word-formation categories. So initialisms enjoy the same freedom as do noninitialisms. The affixation may be initial (non-*U*) or terminal (C*Ber*). There are simple compounds (*H-Y antigen, complementary DNA, A-OK, PG-13*) and more complicated varieties (*gamma-ray laser, single-copy DNA, blood-CSF barrier*), as well as back-formations (*CAT scan*) and variant forms (*Feebie 'FBI').

Over the centuries, the purpose of creating initialisms also has changed from the original medieval need for economy and efficiency (scarce paper and time). As Hamilton noted in his preface (1918): "The use of
ABBREVIATIONS IN WORD-FORMATION

abbreviations and signs is often a convenience and sometimes a temptation. It is a saving of time and labor which is entirely justifiable under certain conditions." A few early individuals had envisioned initialisms as an unstigmatized, vital part of technical vocabularies. Thus Parkhurst philosophized in his preface (1917):

Symbols really constitute a language in themselves. A few characters, suggestive to a marked degree, replace from six to many times six the number of letters that would ordinarily be required to describe the same thing, or combination of things, in the usual words. One memorizes these characters. . . . After the mind becomes trained to the habit of thinking in the symbol language, and this is an exceedingly easy and quick thing to master, it becomes instinctive to talk in the symbol language.

Today we find updated collections of initialisms in most subject fields and a recognized need for efficient items to serve the ever-growing business community and organizations. As initialisms became fashionable in various contexts, items like US(A), USSR, and MIT superseded the full forms (see Partridge and Clark 1951, 221; Adams 1973, 201; Barber 1964, 97; and Malkiel 1968, 373).

The originally condemnatory attitude was slower to change. Wilson (1856) tagged some abbreviations that appeared "unsuitable, either in consequence of their being already employed for other words, because they are less intelligible than they should be, or have but slight authority for their adoption." He warned that few could be used in ordinary composition, just as contractions were eschewed; but he recommended them for catalogues, directories, tables, and family registers. His list (272–300) totaled about 1,200 abbreviations. Boss (1880) also objected to the indiscriminate use of such devices. Years later, Daniel (1946) and also the New York Times (28 Nov. 1946) attacked the use of initialisms, which was said to produce a sensation like eating dehydrated food; F.D.R.'s AAA was bad enough, but now Russian-style AGRADAD and UOPWA were taking over. Time next took up the cudgel. The originally playful sport was said to have become contagion and verbal smog, and the horrible ultimate was still nowhere in sight ("Acronymous Society" 1961). The last straw was EPDOPAC 'Enlisted Personnel Distribution Office—Pacific Fleet'. Such items were condemned as hazards, as when Gale listed eighteen different interpretations for the sequence AID; people do not want to clutter their memories with clusters of letters ("Agonies of Acronymymania" 1971). Some Britishers were also alarmed. Jamieson (1968, 473–74) said that the disease had reached epidemic proportions, into every field of human endeavor. His attack was carefully argued. Acronym-producers claim that the item can be "a convenient code for some
particularly lengthy or cumbersome phrase,” whereas it is likely to be meaningless, often ambiguous, sometimes unpronounceable, and ugly instead of euphonic.

The twentieth-century popularity of initialisms is demonstrated by the increasing numbers and size of dictionaries, some of which have gone into profitable later editions, not to mention the expanding number of specialized dictionaries. Following the nineteenth-century collections, there were three early twentieth-century ones (Latham 1904, Dobbs 1911, Rogers 1913). Dobbs’s book was noteworthy because he remarked in his Editor’s Note that multimeanings were already so well established that his 10,000 entries required him to list about 50,000 meanings. Five collections appeared during the World War Two period (Partridge 1942, Stephenson 1943, Shankle 1945, Allen 1946, Matthews 1947). Fourteen dictionaries of general initialisms have since appeared, most of which used the traditional term *abbreviations* in their titles (Buttress 1954, Schwartz 1955, De Sola 1958, *Acronyms Dictionary* 1960, Mayberry 1961, Fawcett 1963, Goldstein 1963, Moser 1964, Gurnett and Kyte 1966, Wilkes 1966, Kleiner 1971, Paxton 1974, Webster 1985, Miller 1988). De Sola’s book has reached an “augmented, international” seventh edition (1986). Gale’s *Acronyms Dictionary* began with 1,200 items in 1960 and has steadily expanded, to 130,000 in 1976, 300,000 in 1984, and 425,000 in Crowley and Sheppard’s twelfth edition (1987). Such collections are necessary, as Roosevelt’s “alphabet soup” would now pale in comparison even to the much-thicker “soup” used by computer professionals. Insiders at almost every computer hardware and software company use many acronyms and numerous abbreviations.

There are now general regional dictionaries like Wilkes’s British one (1966) and Jones’s Australian one (1977). Also, since World War Two, numerous items from languages with alphabetic writing systems have moved into worldwide use. We find *GUM* ‘Gosydarstvennyi Universalni Magazine’, *TASS*, the German counterpart *DANA* ‘Deutsche Allgemeine Nachrichten Agentur’, the French *CERN* ‘Conseil européen de recherche nucléaire’, and others (Potter 1969, 80). Measurements are often international, as in *kg*.

The number of dictionaries of general foreign initialisms has also expanded, as in Portuguese (Froes 1961), French (Dubois 1977), Russian (Scheit 1986), German (Wendt 1967), general foreign language (Jung 1985), and international (Sheppard and Towell 1987), plus other collections for Finnish, Flemish, Greek, Hebrew, Italian, Latin, Macedonian, Slovenian, Spanish, and Swedish. Spillner’s extensive bibliography (1970–72) lists many German dictionaries. In just 75 years, the number
of initialisms in numerous languages worldwide has exploded from tens of thousands to perhaps 800,000 recorded in dictionaries, besides those that are still unrecorded.

**Scholarship, Terminology, and Taxonomy**

Scholars first concentrated on recording these growing numbers in brief lists that (as noted above) particularly appeared in *American Speech* during the World War Two period. Scattered comments continued to appear in linguistic books and articles, but initialisms received little serious attention until Baum’s four short articles (1955–62) and Hargreaves’ ESL applications (1957). Like blends, initialisms were considered to have no real theoretical import, and certainly visual devices like *bldg.* ‘building’ proved the point. The result was that the many legitimate initialisms in the numerous dictionaries were essentially ignored. Most works in general linguistics still refer only fleetingly to the subject, and histories of the English language provide little new insight. The chief exceptions are Bolton (1982, 362–63) and Pyles and Algeo (1982, 275–77), which has an historical and categorial discussion and differentiates items like *bra, BO,* and *CORE.* Even their treatment is brief, as is that in the most comprehensive book on English word formation (Marchand 1969, 452–54), which relegates initialisms to the role of word manufacturing because supposedly few of them belong to the general vocabulary. Bauer (1983, 237–38) similarly downplays them because they are unpredictable and heavily based on orthography.

There were only four long, theoretical studies of initialisms by 1975. Wells (1956) compared them with their sources, concluding that such “pairs” are different morphemes but nonetheless grammatically related. Makkai (1974) used this comparison to argue that the word-formation process which produces initialisms fits within a stratificational framework. Third, Malkiel (1968, 357–98) provided a worldwide perspective by citing numerous foreign-language examples within a theoretical description. Algeo’s state-of-the-art study (1975) appeared too late for his extensive bibliography to be used in Stein’s word-formation bibliography (1973). The latter listed only 18 works devoted to abbreviations and acronyms, of which 7 were dictionaries, constituting only one percent of the total 1,839 titles.

One reason for the lack of systematic study has been the considerable overlapping and inconsistency within a general taxonomy, on which linguists may still disagree. I will sketch the overlap of initialisms with blends and clipping, before proceeding to the troublesome matter of definitions.
To Kennedy (1935, 282–83), reduced compounds like Duco ‘Du Pont Co.’ were as much a blend as was Arco ‘American Radio Co.’. Bauer (1983, 236) described new-classical compounds where a blend is a reduction resembling a compounding of old Greek/Latin roots into a new item like stagflation (from stagnation + inflation), and other blending shades off into acronymy. Actually, such items differ from typical acronyms like GARP ‘global atmospheric research program’, where the initial letters of the constituent words of the phrase are sequentially formed into a pronounced word. Adding to the confusion is the interpretation by Gale, which publishes the largest collection of initialisms and from the first has described the blends motel, brunch, and smog as acronyms. The influential computer world, an extremely fertile creator of initialisms, generally terms all initialisms as acronyms, even when the items are true abbreviations.

A second reason is the dictionary practice. Since the fifteenth century, European (and American) dictionaries have been listing as abbreviations legitimate abbreviations like D.B.S. ‘de bonis suis’, clippings, contractions (items like can’t, which are pronounced as a shortened form), and visual devices. Chassant even formulized devices like ips. ‘imprimis’ (first, middle, and last letters, “dans l’ordre des lettres qu’elle a conservées”). Today, abbreviation often denotes any kind of shortening including contractions. It is used confusingly for both the item and the word-formation process which produces the item (as in Hockett 1958, 313–16; Fromkin and Rodman 1988, 140, 153). Noah Webster (1850) confusingly offered three kinds of “abbreviations”: Gen. ‘Genesis’, USA, and $ ‘dollar’. The Encyclopaedia Britannica (1876, 26) also defined the term so generally that it might be any shortening: “A letter or group of letters, taken from a word or words, and employed to represent them for the sake of brevity.” The OED used App. and other citations to construct the meaning ‘A shortened form of a spoken word, or written symbol; a part of a word or symbol standing for the whole’. When linguists like Bloomfield (1933, 488) and Wells (1956, 662) adopted this historical practice and described items like prof, lab, vet, and ad as abbreviations, the popular confusion was continued into modern scholarship.

At this point we will stipulate two conditions for initialisms. First, except for an infrequent one or two letters inserted for orthoepic purposes, every constituent in the initialism must have a known lexical source, thereby excluding “shape” items like A-line because A has no such source. Second, no constituent word in the source can be preserved intact, thus excluding French L’heure H ‘L’heure heure’ (see Malkiel 1968, 380). Next, we will propose replacing abbreviation with the common term short-
ening as the name of the division that produces blends, acronyms, abbreviations, and other reduced items. (A division is a taxonomic term meaning 'a grouping of similar word-formation categories'.) Until 1943, no coined term posed a serious challenge to abbreviation as the name of a word type like BA: factitious word (Haldeman 1877), letter-word (OED), Initialwörter (Koziol 1972), and alphabetic shortening or alphabetism (Jespersen 1942, Potter 1969). But a Bell Laboratories researcher's coining of the attractive term acronym (evidently first printed in American Notes and Queries, Feb. 1943, 167/1), and its recording in dictionaries in 1947, began to cause terminological confusion. Three mutually exclusive views have developed: (1) abbreviations and acronyms are separate categories, and (2)–(3) abbreviations are a subclass of acronyms, or vice versa. Webster's Third did not clarify matters by defining acronym as 'a word formed from the initial letter or letters of each of the successive parts or major parts of a compound term'. Their three kinds of examples were jato (initial letters of all words), radar (one or two of the "initial" letters of most words), and Anzac (initial letters of the "main" words). At least there was no illustration like OK.

As the abbreviations and acronyms among our initialisms will reveal formal differences to require their categorial separation, we will continue to assume this differentiation until we describe our data. Meanwhile, we will summarize seven recent, notable efforts to place initialisms within a general scheme of English word-formation, after which we will suggest an overall taxonomy. Heller and Macris (1968) suggested several new terms for a typology of shortening devices, which classed initialisms as a kind of shortening. Algeo (1975, 1978, 1981) successively refined his arrangement into a five-division taxonomy embracing the whole lexicon: loanwords, shortenings, composites (affixations and compounds), blends, and shifts. Fourth, Kreidler (1979) proposed three types of reductions: "acronyms" (his term for abbreviations and acronyms), back-formations, and clippings like Trib and Met (with blends defined as multiple clippings). Next, Quirk et al. (1985, 1580–84) defined abbreviation as the general term for clippings (often reduced to a single syllable), acronyms (embracing acronyms and our abbreviations), and blends (compounds made by blending one word with another). Gramley and Pätzold (1985, 330–32) refined Algeo's taxonomy into four divisions. Their shortenings (Kurzungen) include three categories: clippings (Kurzformen), back-formations (Ruckableitungen), and three kinds of initialisms (Initialwörter—acronyms, abbreviations, and items that can be pronounced as both).

The seventh effort was Cannon's revision (1987) of Algeo's 1981 taxonomy into four divisions. The principal change was a demotion of
blends from being a separate division to becoming the only category of shortening that involves reduction of at least two preexisting items (the other categories involve reduction of a single source-item). Besides blends and back-formations, the shortenings include four types of clipping: initial (fiche), medial (binocs 'binoculars'), terminal (limo), and mixed (prepreg 'pre-impregnated'). Of the four, only limo might be called a traditional clipping, but it would probably be unrealistic to hope that editors might discard the practice of two millennia and not describe limo as an abbreviation. Thus Miller (1988) includes items like max. 'maximum'.

The boundary between clippings and acronyms requires further defining, which we can do by starting with straightforward items. Recent English includes numerous items like kbar 'kilobar'. An acronym preserves only the initial part(s) of a multiword source, whereas kbar came from a derivation instead of a compound, and preserves the base bar; so kbar will be classified as a clipping. But Autoland 'automatic landing' is a bit fuzzy, raising the vital question of how many letters/sounds/syllables must be lost before an item can be classed as an acronym. It has lost 8 of its 16 letters, 7 of its 14 sounds, and 3 of its 6 syllables. This 50% reduction still seems inadequate, not to mention that traditional acronyms preserve at most only two initial letters/sounds, whereas Autoland retains the first four letters of each constituent word, and the morpheme land has not been reduced. Comsat 'communications satellite' retains a much smaller percentage, but still keeps three letters of each constituent. In fact, none of our two dozen borderline creations from two-word sources retains as few as two letters/sounds of each constituent word (e.g., Algol 'Algorithmic Language', sci-fi 'science fiction'). So our interpretation of relevant items in the total 13,683-item corpus would seem to require that an acronym must come from a source with at least three constituents, where a combining form can be a constituent (ASP 'Anglo-Saxon Protestant'). Not more than two initial letters/sounds of some or all of the constituents can be retained, though an exception of three or even four is permitted if the majority of the reduction typifies acronymy. For example, except for the COM- portion, COMUSMACV 'Commander United States Military Assistance Command, Vietnam' is an acronym. Despite having three constituents, Cogas 'coal, oil, gas' is not an acronym because gas is not reduced.

The Corpus

When we add to Cannon's data the initialisms in 12,000 Words (1986), the total becomes 501 abbreviations and 130 acronyms. We will begin
ABBREVIATIONS IN WORD-FORMATION

with abbreviations, describing the source words and the item itself, with attention to names, form classes, usage labels, phonology and graphemics (including punctuation), and subject areas. We will consider the possible manner(s) in which the word formation took place, the item's contemporaneity with its full originating form, and its utility and place in general international English.

The structure of the source words can tell us whether abbreviations come from a wide variety of structures, as do most of the other categories composing Cannon's total data. Though all but 40 of the sources are nouns, 429 of these are compounds of various sorts. There are modifying prepositional phrases (Bachelor of Liberal Studies), compounded abbreviations (cyclic GMP) and numbers (Simian Virus 40), compounded compounds (front-wheel drive), compounded names (Creutzfeldt-Jacob disease), more complicated compounds (Committee to Reelect the President, human T-cell leukemia virus, magnetic ink character recognition), and inflections (lines per minute, cataloging in publication, elapsed time, citizens band, letter-sorting machine, diagnosis-related group, self-addressed stamped envelope, luteinizing hormone-releasing factor, consciousness-raising, million years, Special Drawing Rights, zero-based budgeting). Ten compounds involve conjunction (Death and Dying). The noun noncompounds include 20 derivations (sado-masochism), 7 free forms (cocaine), and 5 coinages (arbitrary combinations of nouns as in Parent, Adult, Child → PAC).

The 40 nonnouns (8%) include 21 adjectives (14 varied compounds as in very superior old and 7 derivations as in Restricted). The other 19 are 3 prepositional compounds (Over-the-Horizon), an adverb compound (as soon as possible), 6 bound forms (deka-), 4 sentences (bring your own bottle, do-it-yourself, draw-a-Person, rank has its privileges), and 5 syntactic groupings that might be described as partial sentences (abort guidance system, also known as, did not finish, driving while intoxicated, material unaccounted for). Only a few dozen of the constituents are names or come from names (Japan, Siemens; Carson-Stouton → CS).

The abbreviations that they provide also reveal large diversity, with many "groupings" containing only two members. A first question concerns the proportion of proper nouns. As the great majority are usually written in capitals, without periods, one might anticipate that these are proper nouns. Actually, some like CAM 'computer-aided manufacturing' may have been capitalized to differentiate the item from the existing cam and cannot be described as a proper noun. Many of the others function as common nouns by taking modifiers, plurals, and possessives. Perhaps only 63 items (13%) function as proper nouns. Twenty-eight of these name organizations, committees, governmental departments, and the like (ACP, AIM); and the others are a widely scattered miscellany (ELT,
MMPI, etc.). PCP is the only trademark. Overall, the only items that are not nouns are 5 adjectives (U, D & D, and the movie-ratings trio—G, PG, R). Four are plural entries (bpi, bps, by, and my; only bps has a plural form).

The proportion of items with restrictive stylistic labels is among the lowest of all the categories, as only four are marked as slang (C, D & D, j., TV ‘transvestite’). The 14 British items (AONB) comprise 2.8% of the abbreviations, as compared to the 3.6% of Cannon’s total data (1987, 248); and the 13 “U.S.” items (MVP) are proportionately much smaller (2.6% vs. 9.2%). An interesting aspect of our abbreviations’ form is that they are almost always homonyms of identical sequences recorded in other dictionaries, a tradition dating back to Walther (1745). Thus Barnhart’s ATV is ‘all-terrain vehicle’, but is ‘Associated Television’ in the OEDS. The multiple meanings may be numerous. For example, Crowley and Sheppard record 11 items spelled as ACSC.

Graphemically, our abbreviations consist of 1–5 letters. Earlier abbreviations were sometimes a bit longer; Fallows (1883) recorded C.Y.M.T.A.&B.S. ‘Catholic Young Men’s Total Abstinence and Benevolent Society’. Our maximum is 5 and may be about the maximum today. Only 4 have 5 letters (EPNdb, HnRNA, V/STOL, VVSOP). The average is 2.8 letters, and 314 (60%) contain 3. There are 455 all-capital items (90%), 28 lower-case (6%), and 18 mixed (4%), with many of them sometimes containing periods. Eighteen of the all-capital items contain other elements—6 with & (H & I), 5 with a number (M²), 5 with a virgule (/O), and 2 with a hyphen (LH-RH). The lower-case ones include r. & r. The 18 mixed ones include 13 with an incorporated abbreviation (kHz, PNdB), plus BUdR, Ig, Jfn, Lr, and Pa. Some involve two steps, as in transfer + RNA → transfer RNA → tRNA.

As orthographic statistics provide little information about patterns, we will consider the representation of the constituent words in the source items. Eighty-nine percent (447 items), exhibiting 5 patterns, contain representation for every constituent word. The first pattern occurs in 357 items (71%) that utilize the first letter of every constituent word, and all but 15 of these sets of constituents are semantically substantive. Fifteen articles, prepositions, or particles are represented: 6 items with of (DOE ‘Department of Energy’), 2 with in (MIA ‘missing in action’), 2 with off (OTB ‘off-track betting’), 2 with the (OTC ‘over-the-counter’), and rtw ‘ready-to-wear’, MUF ‘material unaccounted for’, and DUV ‘data-undervoice’. BUdR ‘bromodeoxyuridine’ exhibits a reordering of the letters. A few of the 357 contain representation of a compound by only the first free form’s first letter, without representation of its terminal
free form \((ULMS \text{ 'underwater long-range missile system'})\). Surely all such items as these probably originated from the written (not oral) full form.

Combining forms are almost invariably represented if all the free forms are also represented by only their first letter. Thus 68 items exhibit the second pattern \((QSO \text{ 'quasi-stellar object'}, pg \text{ 'picogram'}, RIA \text{ 'radioimmunoassay'})\). The third is found in 4 items that lack representation for the base to which the combining form was affixed \((CN \text{ 'chloroacetophenone'})\). The fourth occurs in 13 items, which may be considered the exceptions among our abbreviations. Utilizing 2–3 letters of a constituent, some of these are visual devices like Hungarian Szt ‘Szent’. In English \(vb\) the first and last letters were retained, and pluralization creates a problem. Yet these 13 resemble abbreviations much more than they do any other category. The fact that a given abbreviation or acronym is spelled more often today without periods than with them suggests that it is now a more direct part of the vocabulary. Three of the exceptions use the first two letters \((Pa \text{ 'pascal'}, ID \text{ 'identification [station break]}, FeLV \text{ 'feline leukemia virus'})\). Five utilize the first and last letters of a constituent \((da \text{ 'deka-'}); 4, \text{ the first and medial ones (sr \text{ 'steradian'})}; \text{ and Jpm, three letters. The latter 5 are pronounced as the whole word, not as shortenings. Fifth, 5 items utilize a replacement for a first letter in varying substitutions: pronunciation letter \((BX \text{ 'base exchange'})\), medial letter \((K \text{ 'strikeout'})\), abbreviation \((lbf \text{ 'pound force'})\), symbol \((r. \& r. \text{ 'rock’n’roll'}, \text{ and number to indicate a repeated letter (}L^2L \text{ 'integrated-injection logic')})\).

The remaining 54 items lack representation for at least one unaffixed free form, which is almost invariably a function word. First, the most frequent pattern is that of 41 items that lack representation for only one unaffixed free form, and represent all other words by their first letter. A preposition, conjunction, or article is unrepresented in 35 items: \(of\) in 23 (including a set of 14 names of college degrees like \(DA\), \(by\) in 3 \((AA \text{ ‘Accompanied by Adult’}), in in 2 names of certificates or degrees \((BSEE), to \text{ in 2 (A/D ‘analog -to-digital’), and in 4 conjunctive phrases (CBW ‘chemical and biological warfare’), and the in OJT ‘on-the-job training’. SSL \text{ ‘Licentiate of Sacred Scriptures’ also lacks representation of a preposition, but is reordered from the original Latin. A preposition is unrepresented in CETI and SETI, but both extra- and terrestrial are represented (‘search for extraterrestrial intelligence’). One or two substantive unaffixed free forms are unrepresented in 3 items: \(U \text{ ‘upper class’}, OTC \text{ ‘one-stop inclusive tour charter’}, \text{ and DBA ‘dihydro-dimethyl-benzopyranbutyric acid’ (where dimethyl- is also unrepresented). This trio violate the usual definition that only function words can be ignored. Second, of\}
is unrepresented in 2 names of degrees (DED), where one constituent word is represented by its first two letters. Finally, 11 items lack representation of two words—preposition, conjunction, and/or article. Their 8 patterns include 3 items formed from the structure Master of X in Y(Z), as in MAT 'Master of Arts in Teaching'. Two come from the structure W of X(Y) and Z, as in OMB 'Office of Management and Budget'. Three have no representation for two varying sequential words (ESL); and three have none for two varying nonsequential words (AFDC).

Semantically, our abbreviations are scattered across the spectrum, with 99 (20%) concerned with chemistry, biology, and/or health. Thirty-three refer to computers, 31 to transportation, 30 to the military, 28 to educational certificates (usually college-level), 26 to a variety of organizations, and 18 to space. As the other 236 items fall into perhaps 100 groupings, the subject matter apparently has little or no relationship to whether the new item is an abbreviation.

Our 130 acronyms considerably differ from our abbreviations. All but 3 of the source items are nouns (98% vs. 92% for the abbreviations), and all are compounds by definition (vs. 84%). There are modifying prepositional phrases (COLA 'cost of living adjustment'), existing abbreviations (KREEP 'k [potassium], rare earth elements, p [phosphorus']), compounded compounds (ROM 'read-only memory'), conjoined compounds (AWACS 'Airborne Warning and Control System'), inflected forms (ESOL 'English for speakers of other languages'), and affixed forms (Imp 'in-determinate mass particle'). Unlike abbreviations' sources, the prefixed bound form is usually a prefix rather than a combining form. There are 6 coinages (arbitrary combinations of nouns—4.6%, vs. 1%), as in fido 'freaks, irregulars, defects, oddities' and ANZUK 'Australia, NZ, UK'. These source items contain numerous names, usually geographical (Boston).

The nonnoun sources include an adjective compound (Generally Recognized as Safe) and two assembled sets of adjectives (YAVIS 'young, attractive, verbal, intelligent, successful' and TOW 'tube-launched, optically tracked, wire-guided'). There are no sentence sources. Though the sources are less structurally varied than are those of the abbreviations, some of our acronyms are still probably too varied to permit generation by a phonological rule of the type A → B, as is also true for abbreviations. The several dozen irregularities may prevent the items from fitting into a rule-governed theory of word formation.

Perhaps 70 acronyms (54%) are proper nouns, and their subject areas are more clearly defined and less scattered than are those of our proper-noun abbreviations. Seventeen name organizations, committees, govern-
mental departments, and the like (BOSS ‘Bureau of State Security’, MAC ‘Municipal Assistance Corporation’), and 15 name systems or programs (MEDLARS). Ten refer to the military (MACV), 6 to language (Ameslan), and 6 to computers (PLATO). There are 3 trademarks (Optacon, Taser, VASCAR). While the common nouns are usually capitalized, they take plurals and modifiers (REITs ‘real estate investment trusts’). The only nonnouns are the interjection shazam and two inflectible verbs (MARV and Taser) joined the language at about the same time as their noun homonyms did, and so are classified as acronyms rather than as functional shifts). The regional labels for acronyms are proportionately different from what we saw for abbreviations. Nineteen items are tagged as “U.S.” (FIDO ‘flight dynamics officer’—10.5% vs. 9.2% for U.S. ones in the entire Cannon data). Only 3 are British (Possum—2.3% vs. 3.6%). So our initialisms are more likely to be American English rather than British English, and roughly coincide with the percentages in the entire data. As the acronyms have no stylistic labels, this category is tied with bound-morpheme items (e.g., biocrat) as having the lowest stylistic dimension of all categories. Few appear in other dictionaries with a different meaning, a fact that further differentiates them from abbreviations.

Graphemically, they are longer, having 3–9 letters. Earlier, a few military acronyms had 10 letters, which may be a principled maximum. Only 2 of our items have 8 letters (Intelsat), and 2 have 9 (COMUSMACV). The average is 4.4 letters, and the 59 with 4 letters constitute 45% of the total. Their 7 possibilities (vs. 5) make them less predictable in length. A lower percentage is all capital (OAPEC—103 or 80% vs. 90%), with 16 mixed (Iris—12% vs. 4%) and 11 lower case (ecu—8% vs. 6%). Of the all-capital items, only L-PAM contains any other punctuation except for occasionally varied forms with periods; and none of the all-lowercase does. The “mixed” variety is usually a lower case item that begins with a capital (Acas); if it contains two “words,” the second one is also so punctuated (Pave Paws). Some of these statistics are variable, as in AIDS (American) vs. Aids (British), though the latter might be explained as acceptance of the original AIDS as an ordinary word in British English.

Eighty-six acronyms, which can be divided into two broad groups, have representation for every constituent “word” in the source item (66% vs. 89% of the abbreviations). First, 53 contain only the first letter of every “word” regardless of whether some “words” have initial combining forms or are compounded (TOW). Thirty-nine of these contain no function words (SWAPO). Eleven have such words—4 with of (SOMPA ‘System of Multicultural Pluralistic Assessment’), and 1–2 each with as, and, before,
in, out, and per (GIGO ‘garbage in, garbage out’). Three “coinages” are the least straightforward of the 53. Fido and shazam are true coinages, where a sequence of balanced nouns that fits orthoepic practice is chosen, without particular attention to its phonological-semantic impact. Meow ‘moral equivalent of war’ is a pseudocoinage; an existing item was pre-selected for its ironic connotation, before its “source words” were selected as though it were produced by the usual acronyming process. Here the semantic quality was paramount.

Second, 33 items exhibit 5 different patterns. Sixteen have representation of a bound form (FLIR ‘forward-looking infrared’, so that infrared is represented by two letters). Possum ‘Patient Operated Selector Mechanisms’ has two extra letters needed for orthoepic purposes, Fido has secured the needed i from Flight in ‘Flight Dynamics Officer’, and PLSS ‘Portable Life-Support System’ orthographically lacks the oral vowel needed to pronounce the acronym. The fifth kind includes 14 items with a first-letter representation of at least one constituent word, but also the first two or even three letters of at least one constituent word (APEX ‘Advance purchase excursion’, tacan ‘tactical air navigation’).

The 44 acronyms that do not contain representation for every constituent word exhibit 5 patterns. The majority pattern accounts for 26 items, where all words except one unrepresented word are represented by their first letter. Except for the substantive omission in MIRV ‘Multiple Independently Targeted Reentry Vehicle’ and BASIC ‘Beginners All-Purpose Symbolic Instruction Reentry Card’, the omission is a function word: 10 of (OPEC), 9 and (OSHA), 3 for (NOW), and 2 to (SAM). However, except in MIRV and BASIC, both first letters are retained when the word is a compound (STOL ‘short takeoff and landing’); and only the initial bound form is represented when it is a derivation (ASCII ‘American Standard Code for Information Interchange’). Second, 8 items reveal various patterns of multiple omissions of function words, with all other words contributing their first letter (TEFL, where of and as a are unrepresented). Other omissions are as a (TESL), 3 items with two unrepresented prepositions (ESOL), and TESOL with three. Third, 4 items constitute a set where the first letter is pronounced as though it were an abbreviation, and the remaining three or four letters are the first letters of represented words except for an unrepresented and (CTOL, RTOL, VTOL, QSTOL). Fourth, 5 items have no representation for one word, but represent one or more words by two or even three first letters. The only substantive omission is the radio in LOFT, and cumecs ‘cubic meters per second’ gains its c discontinuously from cubic. Last, CREEP ‘Committee to Reelect the President’ lacks representation of 2 function words, and represents reelect by 3 letters.
ABBREVIATIONS IN WORD-FORMATION

Semantically, our acronyms are distributed in quite different subject areas and proportions from those of our abbreviations. Though they appear across the semantic spectrum, they are concentrated in 7 areas. Twenty items refer to primarily political organizations, 15 each to systems and the military, 14 to computers, 11 to space, and 10 each to chemicals and transportation. Like the abbreviations, they vary trivially or not at all from the meaning of their source items. No variation approaches the magnitude of the shifted meaning of an older abbreviation not in our corpus, STP 'Scientifically Treated Petroleum', which today also means 'a hallucinogenic drug'. CREEP has technically not pejorated, as peoples' pronunciation of the sequence as an acronym instead of as the intended abbreviation caused the categorial and thus connotative change.

Acronyms exhibit the most unpredictable forms of all of Cannon's 21 categories (1987); abbreviations may be the second most unpredictable. Function words are represented only slightly more than half the time, and some substantive words are unrepresented. Numerous other differences that necessitate categorical separation have been seen, but we might recapitulate the major ones. No acronym comes from a noncompound source, whereas 7 abbreviations come from a free morpheme, and 101 abbreviations come from a two-constituent source. As our acronyms require at least a three-constituent source, they are longer. Abbreviations particularly utilize writing, whereas a few acronyms in our corpus originate as "coinages" that are pre-formed for speech purposes. Thus various antipollution and antismoking groups chose the existing gasp for its sound and connotation, a procedure followed almost a century ago for Waac. But this time the multiple creators concocted supposed sources like 'Group/Gals against Smoke and Pollution', to the degree that one cannot know what GASP actually represents (Barnhart et al. 1980, 7). Actually, many more of this sort may have been created than our source material indicates; there seem to be a considerable number in the computer world. Thus the computerization of the OED led to OEDIPUS 'OED Integrated Proofing and Updating System'. A good many organizations have names dependent on this process, as in MADD 'Mothers Against Drunk Driving'.

Such a process is very different from that used in the making of abbreviations. Some acronyms and abbreviations are true coinages, when a sequence of words is assembled to create a desired collocation, in what may be a new pattern for both categories. Sometimes, the first letter of an acronym is pronounced as an abbreviation composing the acronym's first syllable (VTOL). Most importantly, acronyms are pronounced as words rather than as a sequence of letters, and letters are sometimes
inserted in the graphemic shape (or an oral vowel is added in the pronunciation). For instance, ARVN requires an additional vowel to pronounce it as the acronym that it is, and this fact explains the variant spelling Arvin. Acronyms look more like words than abbreviations do, with fewer varying forms containing periods, and they are more likely to be proper nouns (54% vs. 16%).

Now we can refine the relevant definitions so as to fit our large, modern corpus. Initialisms consist of abbreviations and acronyms, 2 of the several categories of shortening. They are the most writing based of all categories of English word-formation, and about half function as common nouns (some even taking possessives). An abbreviation is an item created from one or two first letters of all or most of the 1–5 constituents of an existing item. Medial free forms and bound forms may be constituents, and the resulting shortening is pronounced letter by letter; but there is a handful of troublesome exceptions. Excepting perhaps numerous pseudocoinages, an acronym is created from the first letter (and infrequently the second or even third letters) of all or most of the 3–9 constituents of an existing compound. Initial bound forms and free forms prefixed by a bound form may be part of the representation, and the resulting shortening is pronounced syllabically according to orthoepic practice.

The nature of written data obscures the manner in which the acronymizing or abbreviating took place. To posit a written abbreviation as coming from a written source, we need a source that is a good bit earlier. Barnhart et al. (1980) gives a date of 1971 for AI as coming from the written artificial intelligence, dated 1966. Is this source early enough? We can be surer of the 1975 AI as coming from the written Amnesty International, dated 1961. But what about ASC as coming from the written altered state of consciousness, when both are dated 1972? Dozens of similar situations suggest that the written abbreviation may come from a perhaps undatable oral full form, unless we are willing to accept the marvelous idea that a long form and a short synonym come into being simultaneously. If oral forms are being shortened into writing, there is a dynamic intermixing of speech and writing. Or does the oral full form provide the oral reduction, which then supplies a written initialism that is more likely to be an acronym? If so, this would provide modern evidence that speech still generally comes earlier and might eventually move into a written form. It is possible that the long and the short form may come into use simultaneously. AIDS seems to have been used in both ways within a few weeks, and may well have been used in both ways from the beginning, though the acronymic sound of AIDS was much in the minds of the doctors who created the term, in rejecting GRID, ACID, and others (see Butters 1984).
As Barnhart et al. (1973, 1980) and the OEDS include dates, there is adequate chronological information for the majority of at least the Barnhart items. Information about the contemporaneity of the item with its source may tell us whether it is competing almost immediately with its source (and so might have a better chance of supplantation), or else is competing only after decades of the full form’s monopoly. If both forms appear as new items in the same dictionary, they are presumably contesting contemporaneously. We find 145 abbreviations (29%) so contesting, in contrast to 6 acronyms (4.6%).

A second measure of contemporaneity is the inclusion of the source item in Webster’s Third (1961), whereby we know that it was enjoying some frequency in a variety of writing by 1961. Recognizing that date as a chronological touchstone, Barnhart et al. used post-1961 as a starting point for their collecting. Fifty-two of our abbreviations’ sources (10%) appear in Webster’s Third, in contrast with one acronym’s source (1%). A similar preponderance holds true for appearance of the abbreviations’ sources in the OEDS—122 (24%) vs. 5 of the acronyms’ sources (4%). Clearly, an acronym’s source seldom appears in writing frequently enough to gain admission to dictionaries, though nonappearance might also be explained by some dictionaries’ policy of excluding proper nouns of geographical or biographical information. Thus the nonappearance of North Atlantic Treaty Organization vs. the inclusion of NATO may provide a misleading conclusion. Overall, these statistics provide conflicting indications about abbreviations. That is, 29% evidently go into immediate competition with their only slightly earlier sources, whereas only 10% engage in much-later competition with their established sources, but about 30% if we include appearance in the OEDS. Infrequently the oral or written abbreviation itself became the source for a longer, phoneticized form, as in deejay, emcee, jayvee, kayo, okay, and teevee (see Baum 1957). Or, perhaps uniquely, the original source recreation vehicle provided the longer reduction recevee to compete with RV.

A related consideration is our initialisms’ place in general international English, since they might be distinguished chiefly for their newness and might be ephemeral. Crowley and Sheppard (1987) record all but 21 of our abbreviations, and all but 5 of the acronyms, in identical proportions of 96% inclusion. If an item is listed in a Barnhart dictionary, as well as in Merriam and in the OEDS, we can assume a greater currency than if it appears only in one or two of these. Only 45 abbreviations appear in all three dictionaries (9%), versus 109 in two (20%), and 347 in one (71%). Nineteen acronyms appear in all three (15%), 34 in two (26%), and 77 in one (59%), so that they seem to have comparable currency. Inclusion in the new Random House unabridged dictionary (1987) provides an-
other index: 221 abbreviations and 57 acronyms appear, in an identical but low 44%.

A better index is appearance in the latest desk dictionaries, where the limited space makes word frequency a critical determinant, although scholars have traditionally ignored this matter. Thousands of initialisms are alphabetized within these lists or else are grouped in special Abbreviations sections. There are four American sources: 1982 *Webster's New World Dictionary*, 1983 *Webster's Ninth New Collegiate Dictionary* and *Random House College Dictionary*, and the 1987 *Concise American Heritage Dictionary*. The British equivalents are the 1982 *Concise Oxford Dictionary*, 1983 *Chambers 20th Century Dictionary*, and 1984 *Longman Dictionary*. A considerable proportion of abbreviations is included: 211 multiple occurrences (42% vs. 42 acronyms or 31%), and 97 single ones (19% vs. 25 or 19%). So 61% of our abbreviations appear at least once, in contrast to 50% of the acronyms. Seven acronyms appear in all seven dictionaries (*COBOL, OPEC, SALT, SAM, STOL, VTOL, WASP*), whereas 9 abbreviations are thus recorded (*ABM, CB, DMZ, EDP, IUD, MHz, p, PLO, U*). Many of the multiply-appearing items are straightforward, rather than being the most irregular (cf the prediction of Bybee 1985 and Zipf 1949 for items with higher frequency and lexical strength).

We find this regional indication for items recorded in at least three dictionaries (either all British or all American): 8% of our abbreviations are regional (26 U.S., 13 British), as are 8% of the acronyms (6 British, 4 U.S.). This identical percentage may not be surprising, though 54% of the acronyms are proper nouns (vs. 13% of the abbreviations). Our proper nouns are not reference-book types like the Spanish *EE, UU* for *U.S.A*. Our initialisms have been collected from continuous texts, not from lists, commercial documents, and registers (which particularly utilize peripheral items like *Gk* and *secy*). Many, if not most, of ours are as much a part of the general language and writing as are items from any other categories. A final measure of their place in general English is inclusion in Miller (1988), the first collection designed for popular use. A fourth have already gained entry into this restricted reference dictionary: 115 abbreviations, or 23%; and 33 acronyms, or 25.3%.

The history of English initialisms up to 1961 has indicated that most have not become common nouns, whereas the majority of ours evidently came into existence as common nouns and so were potentially more ready to move into general English. They may have initially belonged solely to the rather technical vocabulary of the particular discipline, but now about half of them also are evidently already a part of general English. The bulk of our corpus are deliberate creations. They are not
as important to word formation in an abstract way as is, say, the new derivation *holophone* from two combining forms. Word formation generally covers the making of words by regular, rule-governed processes like derivation and compounding, whereas initialisms are not necessarily bound by such rules because they are crucially dependent on orthography. The formation of some of them is thus closer to word creation than to word formation, and so perhaps at least these should not be lumped together indivisibly with ordinary word-formation processes. Yet initialisms cast light on some ways that the English lexicon is expanding. When *unidentified flying object* is shortened, we recognize that *UFO* is closely related semantically, but formally is phonologically and morphologically different. When *UFO* becomes a stem to produce *ufological* and *ufologist*, temporarily becoming an acronym in the derivational process, morphological theory is vitally involved. Of course, our 631 items are generally too new to have had a chance to serve as stems.

**Other Conclusions**

All evidence suggests that the already vast number of initialisms is expanding at ever-higher rates and that they are moving ever more easily and quickly into general English. One naturally wonders whether any may be dying. Various checks on viability might be made. Comparison of earlier and later editions of the Gale dictionary would be unrevealing, as Gale's current policy is to retain items for historical value. One revealing comparison is Gale's 1987 edition and Stephenson's 1943 collection, where extrapolation from the first page of each letter discovers a 10.4% omission, or 89.6% continuation in 45 years. Items dropped are like *F.A.M. 'Federal Air Mail*, a service which was abandoned two decades ago. A second indication that initialisms do not massively go out of date is the reprinting of Martin's 1892 collection (London: Stevens, 1949) and Rogers' 1913 one (Detroit: Gale, 1969), and particularly the reprinting of Chassant's fifth edition (New York: Franklin, Burt, 1973), probably the most famous of all dictionaries of abbreviations.

A second aspect of viability concerns whether an initialism may replace its source. Our acronyms have apparently often come into being soon after creation of the full form, which appeared too late for inclusion in *Webster's Third*. As they moved into quick, direct competition before their sources could be well established, they may enjoy greater utility in relation to their sources than do abbreviations. Users of items like *radar* may not even know that there was once a full form. Lexicographers do not accord a separate entry to such full forms, which have been
superseded if not eliminated from the vocabulary. But all of our acronyms need to be observed over the decades, to see if there is any trend or formal reasons for one of the competing pair eventually to dominate. By contrast, about a third of our abbreviations seem to have appeared years after the full form was well accepted. They, too, should be observed, though older examples like DDT parallel radar, in that dichloro-diphenyl-trichloro-ethane is not likely to be resurrected and is seldom given a separate entry in dictionaries.

Interestingly, lexicographers seldom disagree about whether an item is an abbreviation or an acronym. As there are only two such examples in our corpus, we have not adopted Gramley and Pätzold’s proposal (1985, 332) for a category containing initialisms that are pronounced by some speakers as abbreviations, and by others as acronyms. One example is CAD ‘computer-aided design’, which structural engineers use as an acronym (and Random House so records it), whereas Barnhart et al. classifies it as an abbreviation. The other is VAT, which Gramley and Pätzold classifies as “Abkürzung und Akronym” (Oxford provides both etymologies). To avoid a double tabulation, we have arbitrarily classified CAD and VAT as abbreviations, as Barnhart et al. does on the basis of their data; but it is true that there are a few other such items (RAF and ROTC, though the latter has a somewhat pejorative meaning).

Though an item that looks like a word is likely to be pronounced as such, we have found numbers of acronyms like TESL, BMEWS, and PLSS, and several abbreviations like NAD, DOD, and MAT. While orthoepy undoubtedly helps to determine the category, there is still a quality of arbitrariness, as illustrated by the graphemic pair ASP in Barnhart et al. The abbreviation means ‘American Selling Price’, but the acronym means ‘Anglo-Saxon Protestant’. Most initialisms evidently come into the lexicon as abbreviations and remain as such, with only two of ours having changed to acronyms—REM and the curious example of CREEP. UFO is still not an acronym despite ufology. Scholars have cited only one reverse change in half a century, the old UN, which was an acronym before becoming an abbreviation.

Our corpus includes numerous abbreviations of the kind when someone else’s writing system is adapted to a language. English borrowed both the Latin full form and its abbreviation, but has usually preferred the abbreviation (e.g., i.e., etc., and &c.—see Bloomfield 1933, 288). Walther’s practice (1745) of having a single capital letter represent several words (C ‘cum, civis, civitas, carus, curavit, contra,’ etc.) has been expanded and massively extended into multiletter overlapping in the twentieth century.
ABBREVIATIONS IN WORD-FORMATION

Cannon’s borrowings (1987) include a few abbreviations from other languages, but English acronyming and abbreviating have not reached the stage where an item that is being borrowed might be simultaneously made into an English initialism. The closest examples would be SI unit from Système international d’Unites, Department D from dezinformatsiya, and OVIR from the Russian loan translation Office of Visas and Registration. This trio is not in our corpus; they must be etymologized as borrowings, since their full form is not English. An interesting question is whether a language might be more likely to borrow the full form or its initialism when they both exist in a language that uses an alphabetic writing system. In at least two cases, English borrowed the acronym rather than the full form (Persian SAVAK and Portuguese UNITA), and also borrowed KHAD from the Afghan Persian Khidamati Aetilaati Daulati. The order was not changed to fit the English grammatical system, unlike what happened to AIDS in Spanish (el SIDA) and French (le SIDA).

As is clear, continuing change in English acronyms over recent decades has widened the parameters of that category. Only 40% are like the traditional jato. By contrast, 71% of our abbreviations are still of the traditional type like MVP, which emulate the patterns found in 150-year-old French and British dictionaries, which in turn emulate classical and medieval patterns. Though this influence continues to dominate, other, widely varying patterns are emerging; and it is these that present most of the irregular features. A glance at some initialisms in other languages reveals that our redefinitions based on recent English initialisms are inadequate for them and would need reworking in order to describe them. Similarly, a glance at Russian newspapers discovers proportionately far more acronyms than appear in Cannon’s total corpus (1987), as in НПП ‘Pol’skaja Narodnaja Respublica’ and УВА ‘Sojedinennye Staty Ameriki’.

Algeo (1975, 231–32) concluded that English initialisms are easier to make than a word of any other category, letting Everyperson be creative, but also secretive and exclusive. But in the last decade or so, in a subjective sense, this part of neologizing has become more of a mere reductive process rather than of individuality and imagination. This development has not conventionalized or automated patterns of reduction, but, rather, dramatically expanded the capacity for reducing almost any source in many ways. The metaphorical quality (if it ever existed) is being replaced by the pure economizing that first motivated the appearance of initialisms on the linguistic scene, and thus the clock has turned full circle. The creation of an abbreviation as a password within a particular group who know the item is not of much advantage today,
since several identical sequences may make that particular meaning ambiguous elsewhere. As even our 211 abbreviations that appear multiply in desk dictionaries have such competing meanings, one almost necessarily must define the item the first time it is used to prevent miscommunication. One can imagine the confusion that might result in using *TV* to mean ‘tranvestite’, or *IRA* to mean ‘Individual Retirement Account’ without first defining the items. Obviously, this problem is not insurmountable, as our abbreviations outnumber our acronyms. It should be noted that a particular initialism seldom shifts in meaning, though it may become grammatically productive in the usual kinds of derivations, compounds, and functional shifts. It seldom shifts in social status and scope, as in pejoration and specialization.

Cannon’s study of English blends (1986, 750) concluded that oral slips and deliberate creations are pouring out from vast numbers of businesses and people who are enjoyably indulging in linguistic creativity, and that this dynamism and individuality may be matched by few if any other categories of word formation.\(^1\) Perhaps 800,000 written initialisms, primarily abbreviations, have been checked for purposes of this paper and may surpass the total number of blends; and these do not include probably many oral, unwritten forms. While the majority of the written initialisms may be restricted to the subject fields in which they are mainly used, this still leaves evidently large numbers in general use. Almost all seem to have been deliberately created as items which utilize writing, whatever the role of oral sources in that formation. The number of pseudocoinages like *GASP* is evidently growing, but still constitutes a small percentage of the total number of new acronyms; this type may be unique in word formation. A few dozen initialisms that violate the general patterns of our corpus make them among the most unpredictable of Cannon’s 13,683 data (1987), though a few of these are visual devices like those in *Modus Legendi Abbreviaturas* (1475?). Instead of relegating initialisms to a peripheral process of word manufacturing, we should study them for possible insights into language change. Finally, the many, proliferating, continuously varying initialisms have considerably complicated the construction of any overall theory of even English word-formation, much less of universal word-formation.

**Notes**

1. Lev Vladimirovitch Malakhovski, of the Institut Jazykoznaniia of Leningrad, has supplied examples of oral, spontaneous blending of Russian items like
/mama/ + /papa/ \rightarrow /mapa/ from his children. John Algeo and Stuart B. Flexner have provided valuable suggestions in the revising of this article, as did attendees at the EURALEX Third International Congress in Budapest, and at the 1988 annual meeting of the American Dialect Society. The British Museum, as always, was supremely helpful in finding and fetching books, and the Bibliotheque Nationale identified an elusive medieval book.

References


Butters, Ronald R. 1984. Review of 9,000 Words: A Supplement to Webster’s Third


*Modus Legendi Abbreviaturas.* 1475? Weidenbach?


Scheitz, Edgar. 1986. *Dictionary of Russian Abbreviations; Containing about 40,000 Abbreviations.* New York: Elsevier.


