

English Abbreviations and Acronyms in Recent New-Words Dictionaries*

Garland Cannon

This paper will review the history of initialisms (a rubric term including abbreviations and acronyms), with stress on 631 recent English ones. It will consider the scholarship, terminology, and taxonomy within overall word-formation. Description of the 631 items, an updated number from my 13,683-item corpus described in *Historical change...* (Cannon 1987), may shed light on lexicology and lexical theory, the relations between speech and writing, and the seeming state of the art in English initialisms today.

Initialisms have been found in Sumerian, classical Hebrew and Latin, and many modern languages. *Modus legendi abbreviaturas* (1475?) collected Latin items in 277 folio leaves of two 38-line double columns listing visual devices, symbols, contractions, and some abbreviations, as did Walther (1745) and Feutry (1775). Until the mid-19th century, the chief motivation for abbreviations seems to have been economy, and their status in the given writing system was usually stylistically shaky. The many abbreviations in French led Chassant (1846) to publish a collection as famous as *Modus legendi abbreviaturas*, for which he continued use of the ultimately Latin word *abbreviation* and the classical practice of treating clippings like *fil.* 'filius' as abbreviations. His *DICTIONNAIRE DES ABBREVIATIONS...* was widely used in Europe and went into a 5th edition in 1884. The earliest 'book length' English collections may have been Courtenay's 3,000 items and Macgregor's 40-page pamphlet<, both published in 1855 > (Courtenay 1885, Macgregor 1855).

By the mid-19th century, Americans had joined Europeans as creators of abbreviations. There was a sprinkling of acronyms, a term that will be tentatively defined as 'an item pronounced syllabically as a word rather than as a succession of letters'. Items like *A.B.*, *J.P.*, *O.K.*, and *C.O.D.* were apparently beginning to move into the oral vocabulary. Numbers of military abbreviations were systematically developed during World War I, with the various collections culminating in the constantly updated U.S. Army's *Authorized abbreviations, brevity codes, and acronyms* (Washington: Department of the Army, 1985). Roosevelt's 'alphabet agencies' caused a further explosion. So Americans and Britishers were well prepared to create and accept the vast numbers of initialisms that appeared during World War II and that have continued to expand perhaps in all languages. At least 28 dictionaries of English initialisms have appeared by 1988. The Gale Company's three-volume collection (GALE 1987) now lists about 425,000 primarily American items, and general regional collections for British and Australian items are available. Additionally, there are numbers of specialized dictionaries of initialisms (usually abbreviations) for trade, manuscripts, the sciences, etc., not to mention general dictionaries for Arabic, Finnish, Flemish, French, German, Greek, Hebrew, Italian, Latin, Macedonian, Portuguese, Russian, Slovenian, Spanish, Swedish, and other languages.

* The full form of this paper, with different examples, appeared in *American Speech* 64 (1989): 99—127.

Meanwhile, many English initialisms were becoming productive, as in the derivations *non-U* and *CBer*, the compound *CAT scan*, and the functionally shifted verb *RSVP*. Many were losing their stigma and superseding their original full forms, as in Spanish *EE. UU.* and English *USA*, *MIT* and *DDT*. The condemnatory attitude was slow to change, particularly in the dropping of the period that traditionally marked an abbreviation, as in *secy* 'secretary'. Not only are initialisms fashionable today, but many like *TASS* have moved into the international vocabulary. Some languages change the ordering and add the needed grammatical gender, as in French *le SIDA* 'AIDS'.

Scholarship on the subject has continued to expand. Major theoretical studies were done by Wells (1956), Malkiel (1968), Makkai (1974), and especially Algeo, whose state-of-the-art study in 1975 appeared too late to expand the 18 entries for initialisms in Stein's bibliography (1973). Menzel (1983) published his dissertation on the subject at Köln. The overall scholarship reveals considerable overlapping and inconsistency within a general taxonomy, particularly the overlap of initialisms with blends and clippings. The major dictionaries of initialisms continue to follow the classical practice of including clippings and visual devices like *bldg.* that are pronounced as the whole word. From the first, GALE has described the blends *motel* and *smog* as acronyms. *Abbreviation* is used to name both the shortening process and the item that is shortened. Once the term *acronym* was introduced in 1943, it further confused the terminology differentiating the kinds of *Initialwörter*. Three divergent views have developed: (1) abbreviations and acronyms are different categories; and (2—3) one is a subclass of the other. Gramley and Pätzold (1985) have offered a fourth, composite view, where some initialisms are abbreviations to some people, but acronyms to others. Such views naturally confuse the place of abbreviations and acronyms in word formation, with scholars like Kreidler (1979), Algeo (1981), Quirk (1985), and myself (1987) presenting taxonomies. Algeo's has been the most influential.

Now we can turn to our corpus, which totals 501 abbreviations and 130 acronyms when we add the initialisms in Merriam's *12,000 words* (1986) to my 1987 corpus. We will begin with abbreviations, describing the source words and the item itself, with attention to proper nouns, form classes, usage labels, phonology and graphemics, and subject areas. Abbreviations come from a wide variety of structures, of which all but 40 are nouns, and 429 of these are compounds. There are modifying prepositional phrases (*Bachelor of Liberal Studies*), compounded compounds (*front-wheel drive*), more complicated compounds (*human T-cell leukemia virus*), and inflected forms (*consciousness-raising*). 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 nonnouns include 21 adjectives (14 varied compounds as in *very superior old* and 7 derivations as in *Restricted*). The other nonnouns are 3 prepositional compounds (*Over-the-Horizon*), an adverb compound (*as soon as possible*), 6 bound forms (*deka-*), 4 sentences (*bring your own bottle*), and 5 partial sentences (*did not finish*). A few dozen of these constituents contain names (*Siemens*).

The abbreviations that they provide reveal large diversity, with many groupings containing only two members. The great majority are usually written in capitals, and some like *CAM* 'computer-aided manufacturing' may have been capitalized to differentiate them from the existing *cam*. Many function as common nouns by

taking modifiers, plurals, and even possessives. Perhaps 63 items are proper nouns. Twenty-eight of these name organizations or governmental groups (*ACP*). *PCP* is the only trademark. There are 4 plural entries, with only *bps* having a plural form. The only nonnouns are 5 adjectives (*U*). Only 4 abbreviations are marked as slang (*C*), with 14 as British (*AONB*) and 13 as American (*MVP*). They are almost always homonyms of identical sequences recorded in other dictionaries. Thus our *ACSC* appears in *GALE* along with 10 other interpretations. As our items contain 1—5 letters, 5 may be the maximum today. Only 4 have 5 letters (*EPNdb*); the average is 2.8 letters.

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. All but 15 of these sets of constituents are all semantically substantive. The 15 include articles, prepositions, or particles, as in *rtw* 'ready-to-wear'. A few of the 357 contain representation of a compound by only the first free form's first letter, as in *ULMS* 'underwater long-range missile system'. 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* 'quasi-stellar object'). Only 4 exhibit the third, where the base is unrepresented (*CN* 'chloroacetophenone'). The fourth occurs in 13 items that may be considered the exceptions because they utilize a varying 2—3 letters of a constituent, and so are closest to visual devices like Hungarian *Szt* 'Szent' and English *vb* 'verb', where pluralization creates a problem. These range from *Pa* 'pascal' to *Jpn* 'Japan'. Fifth, 5 items utilize a replacement letter, as in the pronunciation letter found in *BX* 'base exchange'.

The other 54 items lack representation for at least one unaffixed free form that is almost invariably a function word. First, 41 lack representation for one unaffixed free form while representing all other words by their first letter, as in the unrepresented *of* in 23 items like *DA*. However, 3 items lack representation for substantive words (*OTC* 'one-stop inclusive tour charter'). Second, *of* is unrepresented in two names of college degrees where one constituent word is represented by its first two letters (*DEd*). Finally, 11 items, exhibiting 8 more patterns, lack representation of two function words. These range from *MAT* 'Master of Arts in Teaching' to *ESL*.

Semantically, our abbreviations are scattered across the spectrum, with 99 (20%) concerned with chemistry, biology, and/or health. Also, 33 refer to computers, 31 to transportation, 30 to the military, 28 to educational certificates, and 26 to a variety of organizations.

Our 130 acronyms differ so much from the abbreviations that they require their own category. All of the source items are compounds (vs. 84%), and all but 3 are nouns (98% vs. 92%). These are modifying prepositional phrases (*COLA* 'cost of living adjustment'), compounded compounds (*ROM* 'read-only memory'), conjoined compounds (*AWACS*), inflected forms (*ESOL*), and affixed forms (*Imp* 'indeterminate 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'). The source nouns contain numerous names, usually geographical (*Boston*). The 3 nonnoun sources include an adjective compound (*Generally Recognized as Safe*) and 2 assembled sets of adjectives (*YAVIS* 'young, attractive, verbal, intelligent, successful'). While there are no sentence sources, the sources of some of

our acronyms and our abbreviations are still probably too varied to permit generation by a phonological rule of the type $A \Leftrightarrow B$.

Perhaps 70 acronyms (54%) are proper nouns, the subject areas of which are less scattered. Seventeen name organizations and the like (*BOSS* 'Bureau of State Security'), 15 name systems or programs (*MAC* 'Municipal Assistance Corporation'), and 10 refer to the military (*MACV*). There are 3 trademarks (*Taser*). 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 the inflectible verbs *MARV* and *Taser*. The American labeling is higher than that for abbreviations (10.5% as in *FIDO* 'flight dynamics officer'), but the British labeling is a bit lower (2.3% as in *quango*). A striking difference is that few of our acronyms have homonyms. They are also graphemically longer, ranging from 3 to 9 letters, which may be a principled maximum. Only 2 items each have 8 or 9 letters (*Intelsat*, *COMUSMACV*). They average 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.

Eighty-six acronyms, comprising 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', of which 39 sources contained no function word (*SWAPO*), and 11 contained at least one (*GIGO* 'garbage in, garbage out'). Three 'coinages' are the least straightforward, where a sequence of balanced nouns that fits orthoepic practice was chosen for its semantic or phonological impact, before the 'constituent words' were chosen (*MEÖW* 'moral equivalent of war'). Second, 33 items exhibit 5 different patterns, ranging from 16 with representation of a bound form (*FLIR* 'forward-looking infrared') to *POSSUM* 'Patient Operated Selector Mechanisms', where 2 extra letters had to be added for orthoepic purposes.

The 44 items 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 (*BASIC* 'Beginners All-Purpose Symbolic Instruction Card'). Second, 8 items reveal various patterns of multiple omissions of function words (*TESOL*). Third, 4 items constitute a set where the first letter is pronounced as an abbreviation, and the remaining 3—4 letters are the first letters of represented words except for an unrepresented *and* (*CTOL*). Fourth, 5 items have no representation for one word, but represent the rest by 2—3 opening letters (*cumecs* 'cubic meters per second'). *CREEP* 'Committee to Reelect the President', which represents *reelect* by 3 letters but does not represent two function words, is the sole example of the last pattern.

Semantically, our acronyms are distributed in quite different proportions from those of our abbreviations. Though they appear across the semantic spectrum, they are concentrated in 7 areas: 20 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.

Clearly, acronyms exhibit the most unpredictable forms of all of my 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. This description has shown numerous substantive differences between acronyms and abbreviations that require their categorial differentiation,

as in the adding of orthoepic letters or in the pronouncing of the item as a word. For instance, *ARVN* needs 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, have fewer variant forms, and are more likely to be proper nouns.

Now we can refine the relevant definitions so as to fit our large, modern corpus. Abbreviations and acronyms are two of the several categories of shortening. The most writing-based of all categories of English word-formation today, about half function as common nouns (some even taking possessives). An **abbreviation** is an item created from 1—2 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 a few coinages, 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 acronyming or abbreviating took place. Most initialisms evidently came from written full forms, but dozens may have come from a perhaps undatable *oral* full form or even oral initialism, in which case we are witnessing a dynamic intermixing of speech and writing. Chronological and other information reveals that (1) an acronym's source seldom appears in writing frequently enough to gain admission to dictionaries, and (2) the source of a given one of our abbreviations is ten times more likely to appear in *WIII* than is an acronym's source. Thus an acronym may have an easier time in supplanting its full form. Yet 61% of our abbreviations appear in at least one of 7 current desk dictionaries, vs. 52% of the acronyms, as an index to their being a part of general English rather than reference-book items. As only 8% of each are geographically marked, with little or no stylistic tagging, a goodly number of these recent initialisms may already be in international English. Also, most of our initialisms apparently came into existence as common nouns, in contradistinction to earlier ones, which usually had the handicap of first having to shift from being a restricted proper noun. Ours are still too new for us to know if they will be productive, but a few examples like *ufology* and *ufological* suggest that an abbreviation (*UFO*) can become an acronymic stem without affecting its abbreviation status, or its semantic similarity but different morphology from its source *unidentified flying object*. All the evidence indicates that the already vast number of initialisms is expanding at ever-higher rates, and that they are now moving more easily and quickly into general national if not international English.

An interesting finding is that lexicographers seldom disagree as to whether an item is an abbreviation or an acronym. The salient exceptions are *CAD* 'computer-aided design' and *VAT*, which are too few for us to adopt Gramley and Pätzold's composite category of *Abkürzung und Akronym* (though there are a few other variable items such as *ROTC* and *RAF*). An item that looks like a word is likely to be pronounced as a word, though there are numbers of acronyms like *TESL*, *BMEWS*, and *PLSS*, and abbreviations like *NAD*, *DOD*, and *MAT*. So there is still some arbitrariness. Most initialisms evidently come into the lexicon as abbreviations and remain as such, with only two having changed to acronyms—*REM* and

CREEP. Scholars have found only one reverse change—*UN*, which began as an acronym.

One wonders whether, in the abstract, a language might be more likely to borrow the initialism rather than its full form when both exist in any language that uses an alphabetic writing system. English, for example, took the acronym rather than the long source in Persian *SAVAK* and Portuguese *UNITA*. English has also massively expanded the medieval practice of having a single capital letter represent several words, with comparable multiple homonyms. Thus *GALE* records 18 different interpretations for the sequence *AID*. So the old virtue of a private, privileged password may have vanished. If I say, 'There's the *IRA*,' no one needs to run for shelter unless I have forgotten to say that I am discussing the American *Individual Retirement Account*. The same is true for *TV*, which has a very different connotation from that of a transvestite. Yet the given initialism seldom shifts in evaluative associations and scope, as in pejoration and specialization. Since one needs to define the initialism in advance, the economy is gone unless one uses the item several times. And, in a sense, this part of neologizing has become more of a mere reductive process rather than of individuality and imagination, with the metaphorical quality (if it ever existed) being replaced by the economy and utility that first motivated the appearance of initialisms on the linguistic scene.

In conclusion, continuing change in English acronyms over recent decades has widened the parameters of that category. Only 40% are like the typical *jato*. By contrast, 71% of our abbreviations are still of the traditional type like *BBC*, which emulate the patterns found in 150-year-old French and British dictionaries, which in turn emulated 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 German and French reveals that our redefinitions based on recent English initialisms are inadequate for them and would need reworking in order to describe them, much less to work toward a definition that would cover initialisms in all languages today.

My study of English blends (1986) concluded that oral slips and deliberately created blends 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. By contrast, perhaps 800,000 written initialisms, primarily abbreviations, have been checked for purposes of this paper and probably surpass the total number of English blends; and these do not include probably many oral, unwritten forms. The majority of the written initialisms may be restricted to the subject fields in which they are mainly used, but this still leaves 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 creation. The number of pseudo-coinages like *MEOW* is evidently growing, but still constitutes a small percentage of the total 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 my 13,683 data. Instead of relegating them to a peripheral process of word manufacturing, as some scholars did in recent decades, we should study them for possible insights into language change. The fact that many of them are non-rule-governed makes them conceptually different from items like derivations and compounds. 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.

References

Cited Dictionaries

- ACRONYMS, INITIALISMS, & ABBREVIATIONS DICTIONARY. (GALE). 1987. 12th ed. Ellen T. Crowley and Helen E. Sheppard (eds.). Detroit: Gale Company.
- DICTIONNAIRE DES ABBREVIATIONS LATINES ET FRANÇAISES. 1846. Alphonse Chassant (ed.) Évreux: Cornemillot.
- DICTIONARY OF ABBREVIATIONS. 1855. Edward S.C. Courtenay (ed.). London: Groombridge and Sons.
- A DICTIONARY OF ABBREVIATIONS. 1855. Michael Macgregor (ed.). Edinburgh: J.M. Miller.
- 12,000 WORDS. 1986. Springfield, Mass.: Merriam.
- WEBSTER'S THIRD NEW INTERNATIONAL DICTIONARY OF THE ENGLISH LANGUAGE. (WIII). 1961. Springfield, Mass.: Merriam.

Other Literature

- Algeo, John. 1975. 'The Acronym and its Congeners' in (eds.) Adam and Valerie Makkai (eds.) *The First LACUS FORUM, 1974*. Columbia, S.C.: Hornbeam Press. 217—34.
- Algeo, John. 1981. 'Where do all the new words come from?' in *American Speech* 55:264—277.
- Cannon, Garland. 1986. 'Blends in English Word-Formation' in *Linguistics* 24:725—753.
- Cannon, Garland. 1987. *Historical Change and English Word-Formation: Recent Vocabulary*. Bern: Peter Lang.
- Feutry, A.A.J. 1775. *Manuel tironien, ou Recueil d'abréviations facile et intelligibles de la plus grande partie des mots de la langue française*. Paris: Debure frères.
- Gramley, Stephan and Kurt Michael Pätzold. 1985. *Das moderne Englisch*. Paderborn: Ferdinand Schöningh.
- Kreidler, Charles. 1979. 'Creating New Words by Shortening' in *Journal of English Linguistics* 13:24—36.
- Makkai, Adam. 1974. 'Acronymy in English (a stratificational re-examination)' in *Proceedings of the Eleventh International Congress of Linguistics, Bologna — Florence, 1972*. 11:345—363.
- Malkiel, Yakov. 1968. 'Secondary Uses of Letters in Language' in *Essays on Linguistic Themes*. Berkeley: University of California Press. 357—398.
- Menzel, Hans-Bernd. 1983. *Abkürzungen im heutigen Französisch*. Rheinfelden: Schäuble.
- Modus legendi abbreviaturarum*. 1475? Weidenbach? [Copy in British Library.]
- Quirk, Randolph et al. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Stein, Gabriele. 1973. *English Word-Formation over Two Centuries: In honour of Hans Marchand on the occasion of his sixty-fifth birthday*. Tübingen: TBL-Verlag.
- Walther, Johann Ludolf. 1745. *Lexicon Diplomaticum, abbreviationes syllabarum et vocum in diplomatibus et codicibus*. . . Gottingen: Pet and Schimidios.
- Wells, Rulon. 1956. 'Acronymy' in Morris Halle et al. (eds.). *For Roman Jakobson: Essays on the occasion of his sixtieth birthday*. The Hague: Mouton. 662—667.