TENDENCIES IN THE FORMATION OF ALPHABETICAL COMBINATIONS: ACRONYMS AND ALPHABETISMS

Alvin L. Gregg

Wichita State University

A special form of abbreviation, by which letters from the beginnings of other words form new words, has become increasingly prominent as a source of new words in modern English. Often called by the generic term "acronyms," words formed in this way are pronounced in two different ways: either orthoepically, like "NATO" or "scuba," or alphabetically, like "VCR" or "LSD." There is a hybrid class of words, alphabetic compounds made of a single letter and one or more unabbreviated words, like "D notice" for "Defence Notice" or "the big C" for "cancer", but I shall not treat these in this study. Here the term "alphabetism" will refer to a word consisting of nothing but letters extracted from larger terms and pronounced alphabetically, and the term "acronym" will refer to such a word pronounced orthoepically (for discussion of terminology, see Quirk and Greenbaum, 1973: 449; Algeo, 1974). The cover term to stand for either of these categories will be the compound "alphabetical combination," first coined in this study (often abbreviated here by the alphabetism, "ABC").

Although such ABCs are not a major source of the new words acknowledged by the lexicographers of English, they certainly occur much more frequently than in previous centuries. According to Algeo's figures (1980) based on recent new word dictionaries, only about four per cent of newly adopted words have come from ABCs. In order to sample the characteristic ABCs in the ordinary vocabulary of English I have chosen for this study to analyze the alphabetical combinations found in five dictionaries of new words: Polyglot's Lexicon (Versand, 1973) , which collects the new word entries published in the Britannica Book of the Year from 1943 to 1963, 6000 Words (Kay et al, 1976) and 9000 Words (Mish, 1983a), separately published supplements to Webster's Third New International Dictionary, The Barnhart Dictionary of New English Since 1963 (Barnhart et al, 1973), and The Second Barnhart Dictionary of New English (Barnhart et al, 1980). In addition, almost all of the ABCs in the main body of Webster's Ninth New Collegiate Dictionary (Mish, 1983b) have also been examined for more information about the ABCs currently accepted as part of the basic vocabulary of the language.

ABCs primarily belong to the peripheral vocabulary of the language. Not usually found in formal writing, they may drop out

of the language when their temporary usefulness is finished. Furthermore, most alphabetisms are only abbreviated alternatives for full-fledged larger lexical expressions that are still in use. For instance, the full forms "television" and "video cassette recorder" still lurk behind the conversational substitutes: "TV" and "VCR." However, many lay people who glibly mention RNA and DNA may not know the names: "Ribonucleic Acid" or "Deoxyribonucleic Acid." Only a few acronyms like "radar" and "laser" have been fully accepted into the lexicon. And only a few alphabetisms like "ABC" (for the alphabet) and "OK," though still informal, now seem permanently embedded in the lexicon. Only the future will show what kind of permanent change in general principles of word formation the current additions may be causing.

Because the principles for the formation of ABCs seem to be both unpredictable and outside the normal operations of lexical production, there is a tendency for linguists to discuss them in a cursory and ad hoc manner. Laurie Bauer points out some of the difficulties (1983: 237-8):

> The lack of predictability in acronyms stems from at least two sources. Firstly, the phrase from which the acronym is taken is treated with a certain amount of freedom to permit the acronym to arise. For example, in BASIC only the first part of a compound adjective [all-purpose] provides a letter for the acronym, while in WASP both parts of Anglo-Saxon provide a letter for the acronym; in GRAS [Generally Recognized As Safe] the particle as provides the A in the acronym, but in FIST [for Federation of Inter-State Truckers] the particle of is not permitted to provide a letter (otherwise the acronym would be FOIST, which is presumably far less effective as the name of a trade union. It seems that the interests of the acronym are the deciding factor in what the "initial letters" of the phrase will be taken to include.
>
> The second main reason for the lack of predictability in acronyms is that not every abbreviation which could be an acronym is treated as one, and there seems to be no particular reason why some abbreviations should be ignored. Clearly, BBC cannot be pronounced as a word, since it violates constraints on the phonological structure of English words, but the same is not true of GOM (Grand Old Man) or OD (Over-Dose). ... Usage alone would seem to make the difference, and it is not clear what factors influence the variant usages.

I agree that the prediction of the exact form of an ABC is not as reliable as our predictions, for example, of the plurals of new nouns. After all, human beings make new words, and it is difficult to predict what human beings will do. This is illustrated by William Safire's (1986) story of the care that went into the name of the Committee to Reelect the President so that no inappropriate acronym might be manufactured. The predictable alphabetism "CRP" was formed (Barnhart, 1980:125), but the unexpected acronym "CREEP" was also formed using the vowel "ee" of "reelect." It will turn out, however, that if Safire's group had had a chance to read this paper, they still might have avoided that embarrassing acronym, since the use of the vowel following an initial consonant is a precedented method of creating an acronym. Perhaps a better example of the unpredictability of the formation of acronyms is the word "KREEP", formed as an acronym for a rock formation found on the moon from the phrase, "Potassium, Rare-Earth Elements, Phosporus," which begins with a K, because that is chemical symbol for potassium. Nevertheless, there are a series of inter-related general tendencies of word formation as well as specific definable tendencies in the formation of ABCs which will make predictions about the probable forms of alphabetisms and acronyms.

The world of ABCs is a microcosm of the larger world of unshortened vocabulary. For example, the lexicon of of regular English has both hills and valleys. On the one hand, there are many homonymous and polysemous lexical items. Thus, "plane" as a noun may be a carpenter's tool, a flying machine, a level surface, or a kind of tree, and it also it has several other related senses as a verb or an adjective. On the other hand, there are numerous potential word forms that have never become real words, such as "splane, lape," or "snape." Human beings seem better able to remember and recognize a smaller number of sound symbols to which multiple meanings can be attached than to remember a larger number of symbols with a different meaning for each one of them. Of course, no one speaker knows all the lexicon of the language, so that some of the homonymy and polysemy of the total vocabulary is reduced for any given speaker. For example, when I hear the word "plane" I never think of it as referring to a tree, since plane trees are not now part of my everyday universe. Similarly, an alphabeticism or acronym may have several meanings; "AI," for example, means "artificial intelligence" or "Amnesty International" or "artificial insemination," not to mention "active ingredient" or "aircraft interception." Yet, when I hear the word "AI," I first think of artificial intelligence. At about the same time a few years ago the acronymic word "WIN" meant "Whip Inflation Now" and also "Work Incentive." In this case two acronyms are homonymous both with each other and also with an unshortened word whose meaning they were probably intended to lean on. In fact, new acronyms are often identical in form to unshortened

words already in the lexicon with totally unrelated meanings: so "Program Evaluation and Review Technique" shortens into "PERT."

The principle called the conflict of homonyms was identified by Gillieron (see Menner, 1936): if two homonyms are intolerably ambiguous, one of them is dropped from the lexicon. When "quean," with its derogatory meaning of loose woman, became pronounced exactly like "queen," in the days when royalty was esteemed, then "quean" dropped from the lexicon. Similarly, the alphabetism "IBM," for "intercontinental ballistic missile," was altered to "ICBM," making it unambiguously different from the name of Big Blue. A more general form of the principle of conflict of homonyms could be called the Principle of semantic Appropriateness. Undoubtedly, this principle explains both the truckers union's preferring to be called FIST rather than FOIST and the creating of CREEP by Nixon's critics as an for the Committee to Reelect the President.

Two general principles of word formation are especially important in the formation of ABCs. Both alphabetisms and acronyms shorten longer lexical units in accordance with Zipf's Law, which posits that the length of lexical forms tends to be inversely proportional to their frequency of use (see Brown,1968). Other traditional ways of shortening still operate, like clipping "optical art" into "op art," or dropping parts of compounds as when "male chauvinist pig" becomes "chauvinist." But ABCs are now frequently the means by which Zipf's Law is realized, so that for a while, at least, "male chauvinist pig" also shortened into the alphabetism "MCP" (Barnhart, 1980: 298). However, an opposing tendency, which I call the Principle of Salient Length, keeps words from becoming too short. There is an optimal length for both the regular words and the ABCs of English. Very few English words, for example, contain only two phonemes, and most of those that are that short are function words, not content words. "Eye" and "ear" are two exceptions to that rule. (But it should be noted that each of these short words is spelled with three letters, increasing its apparent length in the written language.) In the same way, there are optimal lengths for both acronyms and alphabetisms.

I want to look at in detail at the optimal length and structure of ABCs, or, reapplying Nida's term of morphological analysis (1949: 65-66), to look at the canonical shape of alphabetisms and acronyms. On the next page, Table 1 shows the total number of alphabetisms and the percentages of one, two, three, four, and five letter alphabetisms, as reported in my corpus. It is clearly the case that three letter alphabetisms are the norm, the canonical shape, although both two-letter and four-letter alphabetisms are acceptable. It is now appropriate to explore the various reasons why this is so.

DICTIONARY NAME	Total Alpha-betisms	% 1 letter	% 2 letter	% 3 letter	% 4 letter	% 5 letter
Barnhart (1973)	114	.9	20	67.5	11.4	0
6000 Words	145	3.5	20	62	13.8	.7
Additions in 9000 Words to 6000 Words	89	5.6	19.1	61.8	12.4	1.1
Barnhart (1980)	233	1.3	12.9	69.5	15.9	.4
Polyglot's Lexicon	138	0	14.5	52.2	10.1	0
9th Collegiate	78	0	38.5	43.6	16.7	1.3

TABLE 1

PERCENTAGES OF ALPHABETISMS OF VARIOUS LENGTHS

A major factor that has established the triliteral alphabetism as the norm is the typical pattern of shortening a longer form into a triliteral. The usual longer form abbreviated into an alphabetism is a three word compound noun or noun phrase. Shorter expressions have been abbreviated less often because of their very shortness: they had to be used very frequently before the principle of Zipf's Law would cause their shortening. The New Deal's alphabet soup came from words like "National Recovery Administration," which became "NRA," or "Works Progress Administration," which became "WPA." More recently the "Environmental Protection Agency" became the "EPA." In fact, in the data from the two Barnhart dictionaries, 71% of all triliterals come from unhyphenated three word phrases. As I shall show later, however, the triliteral shape, being a canonical shape, influences full expressions of various shapes that are longer than three words to become alphabetisms in the triliteral mold. First, I should like to suggest several other factors that militate in favor of the triliteral form.

If we think about the number of alphabetic combinations available from one letter long to five letters long, we can see a mathematical reason why triliteral alphabetisms are the norm. For monoliterals there are only 26 available symbols; for biliterals there are 676; for triliterals there are 17576; for quadriliterals

there are over 450,000; and for five-letter alphabetisms there are over 11 million possible symbols. There are not enough monoliterals and biliterals together to provide enough symbols to keep from overloading the homonymy and bring about intolerable ambiguity. There are ample symbols available for triliterals to allow both tolerable homonymy and unused symbols. There are, of course, more than enough of either four- and five-letter symbols for a self-sufficiently complete lexicon. I might add that the dictionary source with the greatest percentage of biliterals is the Ninth Collegiate which mainly represents the existing common vocabulary, in which the number of two-letter alphabetisms has reached some level of saturation.

If we look at principles of pronunciation, we discover another reason why the triliteral shape is more normal than a longer one. Since every letter of the alphabet is one syllable long except the trisyllabic "w," a triliteral alphabetism is nearly always also trisyllabic in pronunciation. Originally, alphabetisms must have been pronounced with the accent pattern of lists, in which each successive item received a strong stress and the last item received a nuclear stress. This can be seen in the accent pattern of a short counting sequence, for example: "One, two, three." Later, the prevailing English tendency toward an alternating stress pattern must often have varied this list accent pattern by weakening the stress on the middle syllable, typically without reducing the vowel. The Ninth Collegiate corroborates this pronunciation for triliteralisms, for example: "IUD," which is represented in two ways: first, with a secondary stress on both "I" and "U" and a primary stress on "D," or else with a secondary stress on "I," no stress on "U" and primary stress on "D." A trisyllabic word is longer than the average word in running discourse, but it is similar in length and stress to many English words (see Fudge, 1984: 34-38): "palisade, refugee, debonair, buccaneer, commandeer, masquerade, Tennessee" etc. It should be noted that the middle vowel, even when pronounced with weakened stress, does not reduce in the pronunciations of alphabetisms as it does in the words just cited, so that the trisyllabic pronunciation only approximates that of many of these trisyllabic parallel words with final primary stresses. Because of their length and their stress patterns, which are even less like those of ordinary words, alphabetisms longer than triliterals become rarer. Ten to fifteen percent of the new alphabetisms are four syllables long. But "NAACP" and "FSLIC" nothwithstanding, very few new alphabetisms are five syllables long.

The fact that the graphic appearance of an alphabetism also has an influence on its being perceived as an appropriate canonical shape is a powerful reason why the triliteral shape is preferred to a shorter alphabetism. To be sure, in pronunciation biliterals have a salient stress pattern: either a secondary and a primary or two primaries: as in the recorded pronunciations in the

Ninth Collegiate Dictionary of such a term as "PI" for "principal investigator: either two equally stressed syllables, or a secondary stress followed by a primary stress. A number of English words (Fudge, 1984) have a similar pattern (including unreduced vowels in the initial syllables): "robust, amen, antique, arcade, baboon, ballet, crusade, dilate, unique." However, the graphic length of the biliteral is not as salient as that of the triliteral, so that in written discourse a biliteral tends to appear to the eyes to be less than a word.

Not only have the mathematical odds, the stress pattern of English, and the Principle of Salient Length militated in favor of triliteral alphabetisms, but the sheer number of triliterals formed from three word expressions has also been a self-propagating factor, providing a pattern for future shortenings from expressions shorter or longer than the prevalent three word compound. When function words like prepositions and conjunctions occur in a longer expression, they can provide a symbol or be ignored (most examples from Barnhart, 1980): providing a symbol in "DOT" from "Department of Transportation" or in "MIA" from "Missing in Action," providing no symbol in "ASC" from "altered state of consciousness," or doing both in "MOR" from "Middle of the Road." A two word expression can become triliteral by providing letters from two roots in a single word as in "BFT" (rather than "BT") from "Biofeedback Training," or by providing letters from a prefix and a root as in "DMZ" (rather than "DZ") from "Demilitarized Zone." Even a content word may passed over as in "ACP" from "African, Caribbean, and Pacific Associables." On the other hand, in chemical compounds very long etyma may be shortened, with entire lexical units being ignored as in "DBA" from "dihydro-dimethyl-benzobutyric acid," where the unit "dimethyl" contributes nothing to the final alphabetism. Finally, there is a set of alphabetisms that maintain the trisyllabic pattern of triliterals by keeping a form of "and" between two alphabetic letters, as in "R and R" from "Rest and Recuperation," often written triliterally with an ampersand: "R&R."

As we now turn to acronyms, we will discover that there is less predictability than for alphabetisms. Nevertheless, there seem to be several identifiable tendencies at work. For this analysis I have made a composite list of 131 acronyms I found in several of the dictionaries furnishing my corpus. There are basically two different kinds of orthoepically pronounced constructions that are often called acronyms, one of which I shall call acrostic acronyms and the other I shall call blended acronyms. In addition, an acronym may take the form of an existing word in the language, an homonymous acronym, like AIDS, or it may create a completely new form, an autonomous acronym, like MOSFET, or "Metallic Oxide Semiconductor Field Effect Transistor". Either kind of acronym realizes a way of putting Zipf's Law into action.

BLENDED ACRONYMS	EXAMPLES	%AGE HOMO-NYMS	%AGE NEW WORDS	%AGE TOTAL
1 syllable long	WIN (Work INcentive)	.8%		.8%
2 syllables long	LINAC (LINear ACcelerator)		17%	17%
3 syllables long	INTELSAT (INternational TELecommunications SATellite)		3%	3%
4 syllables long	COINTELPRO (COunterINTELligence PROgram)		2%	2%
				22.5

TABLE 2

BLENDED ACRONYMS

In the blended acronym a syllable is taken from the beginning of each contributing word in a multiword expression, whereas in acrostic acronyms, the initial letters of the expression are taken individually to form a word. Table 2 shows the distribution in the corpus of blended acronyms. Twelve of 18 two word expressions here reduced to autonymous two-syllable blended acronyms, like LINAC for "Linear Accelerator," or MASCON for "Mass Concentration." Blended acronyms are almost always autonomous, not homonymous. Two word expressions nearly always form blended acronyms because of the Principle of Salient Length, since an acrostic acronym would be formed with only two letters. In fact, the blended acronym is virtually the only way to shorten a two word expression by means of an ABC that is also an acronym. In this corpus there were no two-letter acronyms from two word expressions at all: only alphabetisms or blended acronyms. Furthermore, two word expressions formed the majority of the blended acronyms in the corpus: 70% of them.

As with the other ABCs, there are several principles influencing the formation of acrostic acronyms. First, they demontrate the principle of Zipf's Law. Later in Table 4, I will summarize the distribution of the acrostic acronyms. 74% of all the acronyms in the corpus are acrostic acronyms for expressions containing three or more content words. In many cases, what could have been a trisyllabic triliteral acronym has been reduced even further to one or two syllables. In fact, the canonical shape is indeterminate: either one or two syllables is permissible, but it is interesting that most of the acronyms that are homonyms of English words are one syllable long, whereas for autonomous acro-

nyms the longer the original expression, the more likely the acronym is to be two syllables long, with a primary stress on the first syllable and often a secondary on the next one.

FROM NO. OF CONTENT WORDS	TO HOMONYMS OF X NO. OF SYLLABLES		TO NEW ENGLISH WORDS OF X NO. OF SYLLABLES				GRAND TOTALS
	1	2	1	2	3	4	
2			.8%	.8%			1.5%
3	15%		9%	7%	1.5%		32.5%
4	4.5%	1.5%	3%	17%			26%
5	.8%	2%	.8%	7.5%			11%
6				2%	.8%		3%
7 OR MORE		.8%			.8%		1.5%
TOTALS OF ACROSTIC ACRONYMS	20.5%	4%	13.5%	34.5%	3%		72.5%
TOTALS OF BLENDED AND ANOMALOUS ACRONYMS							27.5%

TABLE 3

DISTRIBUTION OF ACROSTIC ACRONYMS
ACCORDING TO LENGTH OF ORIGINAL EXPRESSION

(Percentages are those of the entire corpus of 131 acronyms: acrostic, blended, and anomalous)

Since acronyms, unlike alphabetisms, must be pronounced orthoepically, a major problem in the formation of acrostic acronyms is getting letters, especially vowels, in appropriate places to make pronounceable words. In fact, the chief factor in the formation of an acrostic acronym, besides the need for brevity, is the pronounceability of the acronym. Sometimes in the initial letters of the phrase or term being shortened, there is an obvious orthographic or phonetic resemblance to an existing homonymous word. For example, GHST, for "Global Horizontal Sounding Technique," must look a lot like the word "ghost." How easy to insert the vowel "o" following the "h" in "horizontal," to turn the quadrisyllabic alphabetism into the monosyllabic "ghost." Another way of getting a pronounceable acronym must be fiddling with the original expression in order to get a pronounceable set of initial

letters. For example, the 1947 abbreviation of CARE was "Cooperative for American Remittances to Europe, Inc. (Versand, 1973: 36), whereas now it is "Cooperative for American Relief to Everywhere (Mish, 1983b: 1376). This principle undoubtedly also explains many acronyms like "PERT," for "Program Evaluation and Review Technique," with less semantic appropriateness than in the case of "CARE." It is easier to remember a real word with a new meaning than to remember what is otherwise a nonsense syllable. According to Table 3 on the previous page, of all acronyms in the corpus, 20.5% are monosyllabic homonyms of existing words, whereas only 13.5% are monosyllabic autonomous words. These figures cause one to suspect strongly that there was some manipulation of the original expression to produce a monosyllabic acrostic acronym that was homonymous with some preselected target word. This suspicion is strengthened by some calculations from the data in Table 4 on the next page, which shows the various strategies for making pronounceable acronyms. Two-thirds of the homonymous acrostic acronyms were derived from larger expressions consisting solely of appropriate content words, in which every initial letter was used. In contrast, only one-third of the autonomous acrostic acronyms were derived in this way.

ACROSTIC ACRONYMS	EXAMPLE	%AGE HOMO-NYMS	%AGE NEW WORDS	%AGE TOT-AL
Perfect (all content words)	SALT (Strategic Arms Limitations Talks)	21%	22%	43%
Using a function word	AWACS (Airborne Warning And Control System)	4%	14%	18%
Skipping an available function word	GASP (Group Against Smoke and Pollution)	3%	12%	15%
Using an interior morpheme	AIDS (Acquired Immunodeficiency Disease)	2%	4%	6%
Using an syllabic alphabetic consonant	VTOL (Vertical TakeOff and Landing)		9%	9%
Using an inserted alien vowel	HUFF-DUFF (High Frequency Direction Finder)	1%	3%	4%
Ignoring a content word	MIRV (Multiple Independently Targeted Reentry Vehicle	1%	2%	3%

TABLE 4

MAKE-UP OF ACROSTIC ACRONYMS

There are other ways of getting letters, most often vowels, into pronounceable places. As Table 4 on the previous pages demonstrates with examples, the acrostic acronyms can be perfectly derived from larger expressions made only of content words. Or in their derivation, function words like "of" or "and" beginning with vowels, can either be used or else be ignored. Furthermore, the initial letter of an internal morpheme can either be used or be ignored. Or an alien vowel can be inserted, as in "Huff-Duff" from HF DF, for "High Frequency Direction Finder" (Versand, 1973: 30). Similarly, FNMA, for "Federal National Mortgage Association," became the famous "Fannie Mae" with two inserted vowels. Another solution is to allow a consonant to be pronounced alphabetically, as in VTOL, for "Vertical Takeoff and Landing." Not shown on Table 4 are two other methods of creating pronounceable words out of ABCs. A vowel following a consonant can become the vowel of the acronym, for example, "CREEP," from "Committee to REElect the President." Finally, a resonant consonant can pronounced as a separate syllable as in "TESL" for "Teaching English as a Second Language." It is also clear that in some acronyms, more than one of these methods can apply at the same time.

Since human beings create lexical items out of alphabetic combinations, it is impossible to make completely predictive rules about them. However, it is clear that there are a set of describable tendencies or principles in the creation of ABCs, many of which are associated with the formation of new words in general. Of course, conscious human ingenuity is often at work in ways linguists cannot codify as when it became necessary to refer frequently to "Patient Operated Selector Mechanism," a British term. It could have become the quadriliteral, POSM, but the vowel in that word suggested an acronym. Now a POSM is a machine enabling a severely handicapped person to do many things he couldn't previously do, like turn machines on and off remotely. A vowel was inserted and the word was respelled: POSSUM (Barnhart, 1973), for the Latin verb, which meant "I can."

REFERENCES

Algeo, John. 1975. The acronym and its congeners. The first Lacus forum 1974, ed. by Adam Makkai and Valerie Becker Makkai, 217-34. Columbia, S. C.: Hornbeam.

_____. 1980. Where do all the new words come from? American Speech 55: 265-277.

Bauer, Laurie. 1983. English word-formation. Cambridge: Cambridge University Press.

Barnhart, Clarence L.; Steinmetz, Sol; and Barnhart, Robert K. 1973 (ed.) The Barnhart dictionary of new English since 1963. Bronxville, N. Y.: Barnhart; New York: Harper and Row.

_____. 1980 (ed.) The second Barnhart dictionary of new English. Bronxville, N. Y.: Barnhart; New York: Harper and Row.

Brown, Roger. 1968. Words and things. New York: The Free Press; London: Collier-MacMillan, Ltd.

Fudge, Erik. 1984. English word stress. London: George Allen & Unwin.

Menner, Robert J. 1936. The conflict of homonyms in English. Language 12:229-44.

Mish, Frederick C. 1983a (ed.) 9,000 words: a supplement to Webster's Third New International Dictionary. Springfield, Mass.: G. & C. Merriam Co.

_____. 1983b (ed.) Webster's ninth new collegiate dictionary. Springfield, Mass.: G. & C. Merriam Co.

Nida, Eugene. 1949. Morphology: the descriptive analysis of words. 2nd ed. Ann Arbor: The University of Michigan Press.

Quirk, Randolph, and Greenbaum, Sidney. 1973. A concise grammar of contemporary English. New York: Harcourt, Brace and Jovanovich, Inc.

Safire, William. 1986. Acronyms sought. Exploring language, ed. by Gary Goshgarian, 107-110. Boston; Toronto: Little, Brown and Company.

Kay, Maire Weir; Mish, Frederick C.; and Woolf, Henry Bosley. 1976 (ed.) 6,000 words: a supplement to Webster's Third New International Dictionary. Springfield, Mass.: G. & C. Merriam Co.

Versand, Kenneth. 1973. Polyglot's Lexicon 1943-1966. New York; London: Links Books.