One Sunday in 1953, Cornell students Doug McIlroy and Bob Vanderhoek, weary of solving crossword puzzles, decided to try a different word game -- that of finding examples of words containing each of the 676 bigrams from AA (bazaar) to ZZ (puzzle). More than a decade later, Doug posted the incomplete list on a Bell Telephone Laboratories bulletin board in the mathematics department, asking for help in filling the gaps. Ed Gilbert, another Bell Labs employee, had recently acquired a computer-generated list of all tetragrams appearing in words in Webster's Second Edition; using clues provided by this, he was eventually able to discover examples for the 613 different bigrams this implied were in Webster. The results appeared in a Bell Labs memo and were later published in the November 1969 Word Ways. With the aid of the Webster computer tape used to generate the tetragram list, Doug issued another Bell Labs memo in 1974 listing the shortest dictionary words for each bigram. (The longest such "short" examples were the twelve-letter words arfvedsonite for FV and benzpinacone for ZP.)

Examples for ninety percent of the bigrams had been found; could the remainder be similarly illustrated with words? Obviously, one would have to turn to other English-language sources, but rich troves existed: Webster's Third and other American unabridged dictionaries, the Oxford English Dictionary, the Times Index-Gazetteer. Arrlledan surnames, although not assembled in any single source promised the greatest riches of all -- more than one million different surnames were known to exist in Social Security records alone.

This article presents the successful results of this search: the commonest examples of words containing each of the 676 bigrams. For at least 90 percent of the bigrams in the list, the commonest word was decided by referring to Kucera and Francis's Computational Analysis of Present-Day American English (Brown University Press, 1967). This tabulates the number of occurrences of all words in a million-word sample of American English of 1961, carefully balanced to give representation to different subjects and media (books, magazines, newspapers). In the list below, the word having the greatest frequency among all words containing that bigram is given first; often, the word with the second-greatest frequency is given as well.

The remaining bigrams do not occur in the Kucera and Francis list, or occur so rarely that considerable doubt is raised that the commonest word with that bigram has been identified. For these bigrams, a differ-
ent ranking system had to be devised. I used the following, based in part on the reference work in which the word is found:

1. Words in Webster's Pocket Dictionary or Collegiate Dictionary, plus inferred forms (plurals, gerunds and past tenses). In addition, given names (in the Collegiate appendix) have been allowed, as well as certain extremely common geographical names: names of countries, of US states and major US cities (listed in the main body of Webster's Third).

2. Words in the main section of Webster's Second or Third Editions (or analogous US dictionaries such as Random House or Funk & Wagnalls), excluding all personal and geographical placenames of the type omitted from the main body of the Second.

3. Names of foreign cities (as listed in Webster's Third); US place names (as found, say, in Rand McNally's Commercial Atlas); surnames having 10,000 or more representatives in Social Security files (available in a 1974 Social Security Administration report).

4. Obsolete words or variants not found in Webster's Second or Third, but in Webster's First or the Oxford English Dictionary; words in specialized dictionaries.

5. Foreign place names in atlases such as the Times Index-Gazetteer; surnames not in the Social Security list, but found in Who's Who, telephone directories, etc.

Occasionally, the commonest Kucera and Francis word turned out to be of second rank, or even lower, even though a word of the first rank existed. In these situations, I placed the Kucera and Francis word in parentheses, and added the commonest word of first rank to the entry.

For words of second rank or more, the source is given: 2 and 3 for Webster's Second and Third; O for the Oxford English Dictionary (followed by the dictionary entry to look under); TIG for the Times Index-Gazetteer; RM for the 1981 Rand McNally Commercial Atlas and Marketing Guide; OSN for the Official Standard Names Gazetteers issued for different countries.

A few references, highly specialized, were used only once or twice: CCN for the Century Cyclopedia of Names, 1954; NS for De Nederlandse Geslachtsnamen, 1971; CTL for Current Trends in Linguistics, edited by T. Seboek (Volume 1, Soviet & East European Linguistics, 1963); WB for Webster's Biographical Directory, 1966 addenda; D for Dorland's Illustrated Medical Dictionary, 24th Edition, 1965; and TD for telephone directories.

Eight bigrams -- TX, JO, MX, PX, QG, QJ, QK and QP -- are tagged with stars. These come from sources that many readers would question as English; six are place names associated with long-extinct Indian settlements in North America, one is a nonce-word, and one is a Maltese word. Improvements for these six bigrams would be most welcome.

I am much indebted to Philip Cohen, Dmitri Borgmann and Jeff Grant for filling in various holes in the table.
217

AA Isaac, bazaar
B about, above
C back, each
D had, made
E aesthetic
F after, staff
G against, again
H ahead
I said, against
J major, majority
K make, take
L all, also
M same, came
N and, an
O Laos, extraordinary
P perhaps, happened
Q aqueous, opaque
R are, years
S as, was
T at, that
U because, beautiful
V have, gave
W away, saw
X tax, maximum
Y may, way
Z jazz, magazine

BA back, probably
B Bobby, lobby, grabbed
C subcommittee
D subdivision
E be, been
F clubfoot, obfuscate
G subgroups
H clubhouse
I big, Bill
J subject, objective
K subkingdom
L public, possible
M submarine
N obnoxious
O about, both
P bombproof, subpoena
Q subquality 3
R brought, brown
S observed, jobs
T obtained, doubt
U but, business
V obviously, obvious
W subway
X subxerophilous 3
Y by, baby
Z subzone 3

CA can, because
B Macbeth, ecbolic 3
C according, account
D anecdote
E since, place
F flicflac 3
G eggonine 3
H which, such
I city, social
J arcjet 3 addenda
K back, black
L close, clear
M acme
N picnic
O could, come
P ecphonosis 3
Q acquire, acquired
R across, increase
S politics, characteristics
T fact, action
U cut, particular
V Mechanicville RM New York
W cwm
X acxoyat 2
Y policy
Z Czechoslovakia, czar

DA day, days
B goodbye, handbag
C broadcast
D added, suddenly
E made, under
F grandfather, handful
G knowledge, bridge
H childhood
I did, didn't
J adjustment, adjusted
K handkerchief
L hardly, friendly
M administration
N didn't, couldn't
O do, down
P standpoint
Q headquarters
R children, hundred
S hands, friends
T width, breadth
U during, individual
V advantage, advance
W Edward, dwelling
X Dxr Charf OSN Morocco
Y body, already
Z adz
EA years, each
B baseball, somebody
C because, second
D used, united
E been, see
F before, left
G began, college
H behind, behavior
I their, being
J rejected, prejudice
K week, weeks
L well, himself
M them, system
N been, when
O people, George
P keep, department
Q required, equipment
R were, her
S these, states
T get, between
U Europe, neutral
V even, never
W new, few
X next, example
Y they, eyes
Z squeeze, freezing

FA fact, far
B halfback
C briefcase
D serfdom
E life, few
F of, different
G afghan
H offhand
I first, find
J fjords
K calfkill 3
L floor, influence
M (Hoffman), feoffment
N aloofness, stiffness
O for, before
P offprint
Q Futalaufquen TIG Argentina
R from, free
S beliefs, offset
T after, left
U full, future
V arvedsonite 3
W halfway
X fix
Y identify, justify
Z styfziekte 3

GA against, again
B (Digby), pegboard
C dogcart
D kingdom
E get, general
F meaningful
G suggested
H through, might
I give, given
J logjam
K ginkgo, Bangkok
L English, single
M judgment
N foreign, design
O good, go
P bagpipe, flagpoles
Q Ngqika CCN Xhosa subgroup
R great, program
S things
T Washington, strength
U figure
V dogvane 3
W (Sedgwick), dogwood
X Longxuyen 3 Viet Nam
Y energy
Z zigzagging

HA that, had
B neighborhood
C (Khrushchev), forthcoming
D the, he
E youthful, faithful
F churchgoing
G withheld, withholding
H his, this
J highjacker 2
K (Oshkosh), latchkey
L highly
M establishment
N John, technical
O who, should
P mouthpiece, birthplace
Q earthquake
R through, three
S months
T might, right
U church, thus
V bochvark 3
W highway
X (Hajian O ( ask), meohx O ( mix)
Y why, physical
Z (mahzeer), mahzor 3

IA social, sp
B possible, and
C which, An
D said, did
E experience
F if, life
G might, right
H likelihood
I Hawaii, s
J marijuana
K like, like
L will, still
M him, time
N in, into
O national, P
Q techniques
R their, first
S is, his, that
T it, with
U medium
V give, give
W Taiwan, h
X six, fixed
Y biyearly
Z size, org

JA James, ja
B rajbansi 3
C bijce O (b
D sloj 3
E subject, p
F fijf 0 (five
G fijg 0 (fig
H ihool 3
I Jim, jig
J Hajji
K riksaald
L majlie 3
M. jajman 3
N jnana 3
O John, maj
P Rajput 3
Q fejqien 9
R bajra 3, G
S Wijs meth
T mujtahid 3
U just, Justin
V blijver 3
W bijwomer 2
X Dijkstra
Y yungine 3
Z Gijsegem
A remarkable social, special workbench, checkbook possible, described which, American said, did experience, field if, life might, right Hawaii, skiing mar-Ljuna like, likely will, still him, time in, into national, action equipment, principle techniques, technique their, first is, his, this it, with medium give, given Taiwan, handiwork six, fixed biyearly size, organization

James, jazz rajbansi 3 bijce O(byssus), Cejc TIG Cz. subject, project five, AL Al Jarif TIG Arab. fig O(fig), Bosilgrad TIG Yug. hheel 3 Jim, Jig hajji rijkedaalder 3 majlis 3 jajman 3 jnana 3 John, major rajput 3, rajpramukh 3 feijlen * bajra 3, Gujarati 2 Wijs method 3 mujtahid 3 just, justice blijver 3 bijwoner 2 Dijxhoorn WB jyngine 3 Giżegem TIG Belgium

remarkable workbench, checkbook possible, described which, American said, did experience, field if, life might, right Hawaii, skiing mar-Ljuna like, likely will, still him, time in, into national, action equipment, principle techniques, technique their, first is, his, this it, with medium give, given Taiwan, handiwork six, fixed biyearly size, organization

James, jazz rajbansi 3 bijce O(byssus), Cejc TIG Cz. subject, project five, AL Al Jarif TIG Arab. fig O(fig), Bosilgrad TIG Yug. hheel 3 Jim, Jig hajji rijkedaalder 3 majlis 3 jajman 3 jnana 3 John, major rajput 3, rajpramukh 3 feijlen * bajra 3, Gujarati 2 Wijs method 3 mujtahid 3 just, justice blijver 3 bijwoner 2 Dijxhoorn WB jyngine 3 Giżegem TIG Belgium
Three of the bigrams with asterisks are found in Frederick Webb Hodge’s Handbook of American Indians North of Mexico: AOKIYENIK is a variant of AKIYENIK, the name of an Indian tribe; MAQPIATO is a synonym of ARAPAHO, a major Indian tribe (taken from “The Ghost-Dance Religion and the Sioux Outbreak of 1890” by James Mooney, in the 14th Report of the Bureau of American Ethnology, Part 2, 1896); and SAQUI-GITUNAI is a variant of SAGUI-GITUNAI, the name of a family of the Eagle Clan of the Haida, a major Indian tribe (originally used by Franz Boas in the 12th Report on the Northwestern Tribes of Canada, 1898). TEMXWIGSAN is a place name (meaning ‘gooseberry point’) in C. Hart Merriam’s Studies of California Indians (University of California, 1955). CHUFYTACHYQJ is a Creek ‘province’ in South Carolina, according to John R. Swanton’s “Early History of the Creek Indians and their Neighbors” in the Smithsonian Institution’s Bureau of American Ethnology Bulletin 75 (Government Printing Office, 1922). FXR is a nonce-spelling of the word ‘for’ in Edgar Allan Poe’s short story, “X-ing a Paragrap”. For FJ, the Maltese word FEJQIEN means ‘recovery from illness, cure’. Finally, Hodge’s Handbook of American Indians North of Mexico contains Q’ OA PX, a variant spelling of KOAPK, an Indian village on the coast of British Columbia.