Tables of letter sequences varying in order of approximation to English*

KENICHI HIRATA and M. P. BRYDEN
University of Waterloo, Waterloo, Ont., Canada

To obtain more complete tables of letter sequences varying in order of approximation to English than those generally available, sequences of zero through fourth-order approximation were computer-generated using tables of single-letter, digram, trigram, and tetragram frequencies. Two sets of tables are presented. One consists of 100 randomly selected 10-letter sequences of each of zero to fourth-order material. The other consists of 40 8-letter sequences of each type, selected with the restriction that no letter appear more than once in the sequence.

In attempts to manipulate familiarity while controlling meaningfulness, many studies of short- and long-term memory and of tachistoscopic word recognition have employed stimuli differing in order of approximation to English (e.g., Baddeley, 1964; Mewhort, 1967; Mewhort, Merikle, & Bryden, 1969; Merikle, 1969; Miller, Bruner, & Postman, 1954). Different orders of approximation to English reflect various degrees of sequential dependencies of English with higher-order approximations more closely resembling ordinary English.

The concept of orders of approximation to English dates back to Shannon's (1948) original work in the field of information theory. A zero-order approximation to Enligsh is simply a string of letters selected at random, while a first-order approximation takes into account the relative frequency of the single letters of the alphabet. In a second-order approximation, each pair of letters is selected in such a way as to reflect the digram frequencies of English, and so on. In general, it is possible to construct sequences of nth order approximation to English, given the relative frequencies of the n-grams in English.

In practice, several different methods have been used to obtain orders of approximation to English. Miller & Selfridge (1950) and Shannon (1951) employed a guessing technique that takes advantage of the normal adult's familiarity with English letter sequences. Here, a person would be shown a sequence of letters and asked to construct a word using the sequence of letters. The first letter he added to the string would be incorporated into it, the original first letter dropped and the new string passed on to another individual. While this is an interesting and often entertaining classroom demonstration, college students misjudge the relative frequencies of individual letters (Attneave, 1953), not to mention digram or polygram frequencies.

Miller, Bruner, & Postman (1954) followed a basically similar procedure, but used text material rather than people to generate their material. Their procedure required a rather laborious search of text material for the recurrence of particular strings of letters. An alternative method would be to make use of a complete table of the frequencies of n-grams of a given order.

The most complete available tables of different orders of approximation are those of Miller, Bruner, & Postman (1954). They provide 15 eight-letter sequences of zero-, first-, second-, and fourth-order approximation. Since many experiments require longer sequences or more than 15 examples of a given order of approximation, it seemed worthwhile to attempt to generate such material.

In constructing our orders of approximation, we made use of the tables of single-letter, digram, trigram, and tetragram frequencies published by Mayzner & Tresselt (1965) and Mayzner, Tresselt, & Wolin (1965a, b). Our procedure is best illustrated by an example of how a third-order approximation was constructed. Given an initial digram, let us say ST, all trigrams commencing with ST were weighted by assigning them a set of numbers proportional to their relative frequency. A random number was then drawn to indicate which letter should be added to our sequence. Suppose this letter were E: we would then add E to our sequence to make STE and repeat the procedure with trigrams commencing with TE. This procedure was carried out iteratively until it reached a point where all trigrams commencing with the desired digram had observed frequencies of zero. The resulting strings of letters, making use as they do of observed polygram frequencies, should be limited only by the accuracy of the original frequency counts.

To obtain the desired letter sequences, we generated 10,000 letters at each order of approximation. For zero, first, and second orders, this produced a single long string that could be sampled at random. With higher orders of approximation, the fact that particular sets of trigrams or tetragrams had observed frequencies of zero resulted in periodic termination of the string. In such cases, we simply started at a different point and began another string.

For experimental use, we have presented two sets of tables. In the first set, Tables 1-5, there are 100 10-letter sequences of each of zero- through fourth-order material. These have been selected randomly from our computer printout, subject to the limitation that the sequence had to be at least 10 letters in length, and very obvious words or phrases (e.g., AVOICECAME) have been eliminated. Shorter sequences may be constructed by deleting letters from either or both the end and the beginning.

Many Es are also interested in serial position effects, and these are difficult to assess when individual letters are repeated within a sequence. For this reason, we have also selected sequences varying in order of approximation that have no repeated letters. Table 6 gives 40 examples of eight-letter sequences of each order of approximation that have no repeated letters.

*This work was supported by Grant No. A-95 from the National Research Council of Canada to the second author. The authors are indebted to Douglas Mewhort and Philip Merikle for their encouragement and advice.

Table 1
Ten-Letter Sequences of Zero-Order Approximation to English

JRGAVTDZSO	INBOTSHLAD
ODTGUIWHAO	EMIQTUBKUC
UBLAHOQSRP	GMSARURWYC
ZKORFEBYXB	XDPJYLZBLJ
VYJQNMHKZC	ZWNDEBQKPD
BIMZSKBWPS	JQOTOKNOPV
CXPMOYUFCT	TMBZLUWFKM
HBRPMQORVZ	ETZCWFSERP
RCMXZTVDNR	LEAOGGITOS
JCIAFYZJDQ	JSENOQSNMK
MJXAZORWCZ	NPTGRVZIDV
LDYHGARAUR	OLKQAPFZCM
ICTNJNRHUE	CNGRWDZUWF
OIBMEUHPUM	ZFLVUJZQBQ
WXKMEUAEQI	JKFOLTKVFV
CNHZGRQWMA	UOKMCIYTZD
BSUEBTXUOG	UKKQICVBUD
OEAKMDNSQO	PNUCKOSDIW
SPHHFVANWS	RIXOKHCBKT
EVOXGKIEFR	JFZAECZWKR
VXAYNLEYGX	AEWUXMRQTV
KXXHNMFTVA	BYCDYSQLXI
JQNHGWXMUO	ASFQAIZWGS
VJTCVDNUCN	ILUVRQSOBK
WMGBESMNYW	DUNAGKFQAG
PUMWTPYHCA	PQHDCOKIJG
IJCJMOUXIS	PLEBQIGLPO
QOPQNHCYTC	GDRUPLBPWW
CMEPYJOAFB	SAVVPBZNDH
QEIPKCAHPB	BVKFNGWTUA
HWTOSHGZUW	ACFLWBQDFD
KUGKYCEBYZ	MZRFGMQLXP
IXJZSOYNMK	XLOEHAMSBZ
GWVGERPISJ	IUWWMHFSKH
XNZDVZGSEP	DTLVEMFZFK
UEVMJWXVPJ	LJHJOXKMFV
UJIMATMXHM	CFLSNREFPU
HBRPHKYDCO	SCHUGFDYJW
PNTSQMYLFE	XUOCOQSNUX
RHROCTPTEK	AFYUFZAJXB
UOBJUHVGAV	LEHXZVYHIF
WBOPRQWLZV	PTAGLIZCTR
WBTVERNIBM	ZOGSUNWQZL
LHZDBATCXS	RNZECYZIDV
PNUFCNLWFK	GITRKFSHJU
RTLRDTNELD	FTTRKCECZX
AYLHPCZEGX	EVGETYARSH
DYQGVRTJGI	NJNSRPJUMZ
CFHUCJITRJ	NQJKAMULKL
FKHGTIFLSJ	SZREYGWALM

Table 2
Ten-Letter Sequences of First-Order Approximation to English

NMTIOYDSLD	DSERTIGACL
RGHGNONRHL	EAUDNCNIMD
ENLTOGAIDS	OLDNAFEITE
SAEENEHASD	DHPTENYUPP
HGLBAOFAEK	ARETDNTATI
SIHROEWLMO	TEOPBHEINV
SATDTWUSIN	SACLANEBVU
HSNGVSCSOT	ISWMNONTAE
HSLEUERTTS	RKGOAHYFWD
EWTDMYWVAS	DAYNRTTIUI
YTHEFMLAAH	OANELKHTAL
LRTLEYTGBO	TNNMAEAMTI
HSREWEOWAI	TMSVOMDAAU
AIUSININMT	HFDBETLIWA
IUPPAWODEA	NSDAAHENTE
NRETSAIEMT	ETYEIOSWOI
HAARVHVANI	SREOERFITH
DBENFOEDHV	ALRTUCSREA
YHETOAHDMU	RICOADSEHW
LTOAEYAEXA	CFTDNDLSEG
CYTGBNRARO	EIGWYEILAN
ULPNEHBGBO	GEHGMESHIT
WIXVTNUHUD	CNKREHSAHT
NIBUSGCNPS	NDUETLRCWV
OESNOPEBRI	ENNEDSLAPO
KHYOTERTHN	EANAUCHETF
KPTEAOLESO	OEAMCTBTAY
DNUMTERLHW	EYHIWCTETH
GTENRATNJS	HTTIENENYU
RRQNEANYTE	HNSNEHTCIE
RESDONMEFH	PEHYLARUCE
TESYDYOTLE	WTFLEVOENT
WOBSODAUAA	TIRHNAEWSA
RLFHLRILTT	RWHTMHNVGL
IHAETCUNZA	EDYOAEESBL
BNPTYEENLO	COTLEABDYS
TNRHIILFEL	RNGOURDEEF
BCSRFNRATE	CUNXSTASHA
IEHEYTFCTG	MEHMSCIRCH
IECTASDWGH	BOWAILELIT
OAHRTTELSL	TSBODUEETN
SAEOATIBST	LEWSWOCDSO
AOMPHAWFEH	NAEYOTOBJA
EGWTJBGDAT	LOHSDSEEOT
MNRWHOFEGI	DGNLYIREWT
IATEOENHTR	RPHYSTDUDL
AUDIHEAVMS	CAFKDMAHCC
VOITDTFRIE	ETWPIMDATR
EMOWAILTNT	LUWHTHSAEA
TESOTYNLWE	OUCEGBOAKN

Table 3
Ten-Letter Sequences of Second-Order Approximation to English

STHICONGAN	ANDSEFORAI
ENTYODEAYP	ERANGOBEOD
RITHOURLYE	SEACKIFOFE
DEXANGELRI	ONGEDYERST
BANGENDEDY	THERUGOIEV
PLTHOUNTSE	NCORELINCA
HAVEOTHEAN	WOUSTARENO
ANKEDEXPUC	TECASKNDIG
REWOMEVEAT	SKERURKEAY
KITWATHIPR	ICESURUTHA
COLYITHALA	NGEYERONEN
DIKENOUNDU	MOWIDSLARS
ENURYEREID	ITEDSUSEYO
TYSSUGECHA	ICLANOUCOE
GHEADYOUSH	MPOTLLESTH
BUANADIVET	AWINKERSIT
MADINDRORE	EOUSKEROGH
HISOENEDEL	ESOPORRALT
ESOFEDANOU	LDUTHOURBU
RDINESPLOU	MPELYERACL
ENGRINTOSH	HEASACESER
NDELATHIKE	VESLEACAYO
OEEYOWANER	TALEDEREIN
ONJOUTEROL	HALETHAYIT
REANGYSISQ	SUTHERYOMO
HOADANEANT	DGHERKEVEL
ACLONYOUMO	ULATESITLD
ILALERONDT	GIREVIERAI
ALENDETHEV	HOONEDBEIN
NGHERYERIN	SITEDETHOP
ISHALYEATL	GORETHANOW
YONLURBUTE	OLINGOMATH
EADRHEABON	GHERYOWHEI
ONYSENGSTH	TOVEANTEDG
EREDFRSERI	LEFAWNORIT
ANORAVEREA	GHATEDINDE
ORANENGICL	ORERSULDET
GOUNDEREPA	ANDETHRKNO
ORGEDOUYOM	HFFIPICOEN
GOUMEYONEA	UGUTHALOIN
OVAIDARGED	DGESAIKICR
UGELDYOORA	NSOWNGLACR
RAVOIVEITO	ELETHAYOUG
OBOWLAINTH	THUCARAURO
IOVEANGIME	TESESOUCHO
DECKIROPEX	ABUDULLOUZ
LOOAYSTHTI	HITEWAFOUN
OTHIMORSKE	LANOINONEA
HENCKITYOL	FONLLIGATH
ILASCHERAS	ADFRENTALI

Table 4
Ten-Letter Sequences of Third-Order Approximation to English

SENTEAUGHT	ONEWANDLYI
IGHTSELDER	YINGERANYH
BACHOULDSI	SECTUREADE
BEHARELION	VERVEDOWNS
IESTONGESS	CAUSCHESET
GENJOBLIPE	EANCENTOCA
TSEMEARRAN	KEDSINTEDS
GOICEDSITH	FACCENTSEV
VERESERYIE	SESSITHELI
IFELFROUCH	IONEDENTOO
TOODUPEOPE	JECAMEADAY
KINIGHESSA	LABOUGHTSE
ACKPINTSER	MANDECONEE
ASKIDERESS	AMESEDSIRS
ONGETWIFTE	DERYSERMIL
CERSODYINT	GUSTAUSECT
DOWEINGERE	AXESTALKIL
TRALKENTER	ESTEMACHAT
DINGUNGETW	WRIESHEYES
SIGHTERYIN	HOPLAUTHES
WOUSTALLYI	INDENTSEAC
QUIRAIRDSI	ERYSTALMOS
EXACHATEAN	INGRAWAING
POUTSESTON	TYLEDSIONC
SHOOBLUNGE	MICHENDERA
GRANTICTIN	EDINGTHERS
KIDINAMEND	ASTOPENATH
HERSHIMARE	EARLDINHAT
ERSTOPLUCH	SETWERHATH
DESTAKEDOW	LIONECOURN
EGIVEDESSU	LARSOMENDA
ENINGLOSTU	ECOMINSTER
MEREASSAND	TOLSOUNTSE
EATENTODYI	ONCENDSISO
GESOMEDEGI	EREPAROMED
SICHORANDS	PIDEARLDEC
CESSUCTORA	ENTYPEDGES
PLICANDEAR	SIMEREARDE
AROVERFANT	INCERSUCHI
LEDSICHERY	FROESTUNKS
SPIERSTRAC	PEDSIRSTAL
BITHEMEDGE	LACHAVENTE
GERSTONLYI	VOINGLEALL
YARLSONEES	ROADEDSILE
ERENDESTUD	NEIRSTAREA
STRUCHOTTL	RIBEAGETSI
OTHIMENTUA	CHATEDSIME
PEONGERNIN	UNCEPTHEND
FULEATENST	RIEFELVERN
STRIDEACKE	ENYOULBOUR

Table 5
Ten-Letter Sequences of Fourth-Order Approximation to English

ABANDEDSID	IDELECTIME
AMATTENDER	ALMOSTILLA
ENACEDARKE	OVIDEADING
ERSINCHESE	BENCERNALL
CHOOLICECA	SILENDERSE
COLLAGESTY	EDITICENTS
DREACHECAU	EGANGERSES
HEMINDOLLA	EXPECIALSO
FLUIDEALSO	GOVERIENCE
HOPENDLOTH	INTOMBERAL
JECTUALLED	KULLIEVENT
ENTRONGERS	ESTILLIONS
OBSERIOUSE	NICEDARTIC
LOSTILLARG	SYSTERRORS
PILESSAGER	TIMEDITORY
RAIDENTION	RAWEREDICA
DEARSHINGS	RIMESSONST
SLANCEPTHS	SPOKERSONS
TISTATICLE	TREATIONST
TARENTLEDS	URCEDARKED
WORDEREDIC	AMPLEARSHI
EXPRESULTS	MOWERECONT
RISERVOUSE	SINCOMENDL
UNLIVESTUP	TREMOREDIT
LAGENCERTY	SPIRIENDER
CAPABLEEDO	AKINDINGED
AMMEDICABI	ASICALLEDS
STINGLECTS	AWNINESSAG
BLEDSIDERN	BUSBANAGES
DEALICENEA	COOKENDLES
SECOMENTUR	CUTBACKIND
EMBERSONGE	ESKMENDLOT
FASTEACHAI	FOREDIUMPE
SOLDESIRED	GREAMEDITO
TERSONSULT	TEENANALLY
EDUCKETCHE	LINIMAGEST
TAINSTEAMS	MIXTEELERS
TICLEASTIL	OMENTURNED
PALACESSES	PATCHERELY
PLODDENTAC	QUEERIENDE
REAMERALLO	REGAINESTE
ROUBLEMEDY	SHRIENDINI
STEMPETECT	SWANTENSES
TURNINGERS	ENERAGERSO
INFEREIGHE	VERALLETTI
YONDEEDSID	HEEKENERAL
ESTAFFAIRY	PLUSHEEPIS
VINSTEAUTY	LECTIONSET
WHILDERNLY	ATHEMINDOW
ESPITCHELP	OUTINGINST

Table 6
Eight-Letter Sequences Without Repeated Letters

0-order	1st-order	2nd-order	3rd-order	4th-order
OKEBQIHU	NUGLDEFI	KINLESCA	LEPICART	DINGLECT
CVDHMLIY	AYOWUSGC	REAMPICO	DRACKETW	YOREDIUM
HTCRDUNB	LHSNDTYU	ECLOSATH	CHISTAUG	LIMBERSO
LRWDGFYV	AIOPHELK	KNDEYOSA	EPOICHAT	ARSHINKL
PZQVKUGF	OKTAIUPS	EDSTHOUL	INYBETAL	DESPIRAL
OMSBVPZQ	LEUNCOYA	WHEPLOBR	ENCHIMPL	MEDITORY
YLECVGWX	OAHRTEID	NTAROULI	AVERSOMD	WHILDERN
ANXRYOZW	ETCORSBH	ESHARKIC	DERSTANY	CULTINGS
VALRZTQC	DTUOHNAE	NECODSPR	MAGEIWOR	FAILEDSI
NJKEMLDX	YUMADTRP	SITHORAN	EDSITHAL	SPEALTHY
UCQXSAWV	OBNMELDH	EMICHANS	MOUTSELY	POLICKET
HDFBKVJS	MLEIDSHU	OUNGRSTH	NAMICHER	RINTOMBE
YUGKENSJ	NDTCKUGM	ASOTHERM	CHOURNIS	PRAINSTE
KWRFHVLJ	DPRAUBEO	IRSENYOC	HANITSEL	SURFINDO
JMNROEBV	LHUAEOTG	DSEROUTH	STEMPORL	OMENTACK
FKMITPSC	NEGMFCUH	XASTYINE	ANSTRIGH	ONGESTIL
AMOYSVQF	DSERWIYV	ATHIBUNG	HEIRSTAN	APONSTER
PRVSYCHW	UETDHCAM	ONSILDAB	REASPICH	SLANCEPT
ZDXFCRKE	EDKRAQTV	THASEWNG	ASYMBOUP	ALICENTU
EPXCMZRF	IFAODCKT	TAYORGUC	ACKENTOM	HEMINDOL
MNUWHVLE	WAGTDBHE	HEDICKNG	LOBIGNER	REATIONS
KZBQEMFV	BWAENYRI	TUGHAPOF	GANDEOPL	MERIOUST
NDZRISMK	IHENATUV	BUECADFT	BOASKENL	CLIMPANY
WCRDTIBM	ETFHRMUO	INTHAMED	STABOURN	UMPERALS
TQGPMLND	HEPGTILA	OWAGUNER	OUPLYIND	LAPERSIN
DUOGNAIV	AUHTBROE	UNDECHIV	ALKINGER	DICTORES
HWPTFNCA	SNHELACY	LCAINDST	CALFROVE	HAMBEDRO
GFZCSOTQ	SOYWAGTF	DSPRAVEN	TRACHINK	FAMILEDS
RXLPOAEV	EOKABUSL	AIKNEDST	WORTAING	GLOBSERI
FKMERJGC	GESBOAIN	SWHALTED	TREADISB	ONCERTAI
HPFVXCKR	TSPEALNU	DOWACLTS	ORTSEACH	PENDLOTH
URTMEVFA	NRTSGOWA	ELFRISAG	PLADINGE	EDICALMO
CBINGSYJ	ICPEDTAB	WATHIPRE	DESTRUNG	BUILDENS
HCZYSWUO	PNFROELT	MEBUITHA	ESPORNIC	PECIALSO
EBUIWKMH	LRTNADEG	AIEROUNG	PERABOSI	IMPROBLE
OMWVHCBQ	KISDWFEN	HELARSWN	HATUDIES	CRYSTEMP
IDVYENMK	RNYOTLHA	WACKNSPE	HERYINST	TOMINERS
QTPZIFMV	DOIELMST	PSTHORID	STRIDEAC	BIOUSTYL
JNUWHVGB	VNUIHETO	CINDASTH	INGETUCH	OMINGERS
EMFVDOXN	CTUDOLHR	OUGRLASH	SEDOCKAG	REDUCKIN

REFERENCES

ATTNEAVE, F. Psychological probability as a function of experienced frequency. Journal of Experimental Psychology, 1953, 46, 81-86.

BADDELEY, A. D. Immediate memory and the "perception" of letter sequences. Quarterly Journal of Experimental Psychology, 1964, 16, 364-367.

MAYZNER, M. S., & TRESSELT, M. E. Tables of single-letter and diagram frequency counts for various word-length and letter-position combinations. Psychonomic Monograph Supplements, 1965, 1(2, Whole No. 2), 13-32.

MAYZNER, M. S., TRESSELT, M. E., & WOLIN, B. R. Tables of trigram frequency counts for various word-length and letter-position combinations. Psychonomic Monograph Supplements, 1965a, 1(3, Whole No. 3), 33-78.

MAYZNER, M. S., TRESSELT, M. E., & WOLIN, B. R. Tables of tetragram frequency counts for various word-length and letter-position combinations. Psychonomic Monograph Supplements, 1965b, 1(4, Whole No. 4), 79-143.

MERIKLE, P. M. Presentation rate and order of approximation to English as determinants of short-term memory. Canadian Journal of Psychology, 1969, 23, 196-202.

MEWHORT, D. J. K. Familiarity of letter sequences, response uncertainty, and the tachistoscopic recognition experiment. Canadian Journal of Psychology, 1967, 21, 309-321.

MEWHORT, D. J. K., MERIKLE, P. M., & BRYDEN, M. P. On the transfer from iconic to short-term memory. Journal of Experimental Psychology, 1969, 81, 89-94.

MILLER, G. A., BRUNER, J. S., & POSTMAN, L. Familiarity of letter sequences and tachistoscopic identification. Journal of General Psychology, 1954, 50, 129-139.

MILLER, G. A., & SELFRIDGE, J. A. Verbal context and the recall of meaningful material. American Journal of Psychology, 1950, 63, 176-185.

SHANNON, C. E. A mathematical theory of communication. Bell System Technical Journal, 1948, 27, 379-423, 623-656.

SHANNON, C. E. Prediction and entropy of printed English. Bell System Technical Journal, 1951, 30, 50-64.