

Introduction to the CommonWords Database 2016

CommonWords is the main table in the database and consists of 20 fields, which are described in more detail below. In addition to the CommonWords table, the database also contains the following nine tables:

- (1) CORRESPONDENCES: SOUND-TO-SPELLING, which lists the 356 sound-to-spelling correspondences found in the 8,591 CommonWords;
- (2) CORRESPONDENCES: SPELLING-TO-SOUND, which lists the 356 spelling-to-sound correspondences in CommonWords;
- (3) THEMES, which lists 169 themes, or topics, for which more than 7,200 words have been tagged;
- (4) SOUNDS COUNT, which lists the total instances of each of the 68 sounds in the 8,591 words in CommonWords,
- (5) LETTERS COUNT, which lists the total instances of the 26 letters,
- (6) SPELLINGS COUNT, which lists the total instances of the 155 different spellings recognized in CommonWords.
- (7-9) COMMONPREFIXES, COMMONSUFFIXES, and COMMONBASES, which list the total instances for each prefix, suffix, and bound base in the 129,000+ word Lexis database. The figure is taken from Lexis because there are so few instances in the smaller CommonWords.

Each of the ten tables is discussed below. In these discussions letters and spellings are enclosed in arrowhead brackets: <a> indicates the first letter of the alphabet, <sh> indicates the consonant digraph in *show*. Sounds are enclosed in square brackets: [b] equals the first sound in the word *bat*. Explications – that is, analyses of written words into their prefixes, bases, and suffixes – are underlined. In an explication the forward slant, /, indicates that the following letter is to be deleted, as in *decided* [de+cid/e1+ed]¹. Search strings to be used in filters are in green.

COMMONWORDS. This table, the main table in the database, lists 8,591 words, more than 6,500 of which are high-frequency. It consists of 20 fields:

1. Word. The Word field lists the 8,591 words and can be filtered to words

with various letter strings – for instance, filtering the Word field on **sh** returns all 215 words with the consonant digraph <sh> anywhere in the word (plus *grasshopper*, grass+hop+p+er01, in which the <sh> is not a digraph, but a product of concatenation). Filtering Word on **ends with sh** returns only those 66 words with final <sh>.

You can filter, among other things, for these four different kinds of consonant strings: blends, doublets, consonant digraphs or trigraphs, and simplifications:

Blends. The following fifty consonant blends are strings of two or three consonant letters that spell two or more consonant sounds in the same syllable. Some blends are word-initial, more are word-final, a few are both. For instance, filtering Word on **rm** returns 93 words, some of which are false hits, the <r> and <m> being divided by a syllable or element boundary as in *chairman*. However, all of the blends listed below are also tagged **cb** (consonant blend) in the Analysis field. So to reduce false hits, you can filter the Word field on **rm** and the Analysis field on **cb**, to return 42 words with <rm>, with very few false hits. The remaining false hits are words like *garment* in which the <rm> is due to concatenation gar5+ment] but the <nt> is a blend. Other blends:

 = [bl] as in *blue*
<chr> = [kr] as in *chronicle*
<cl> = [kl] as in *clue*
<cr> = [kr] as in *crew*
<ct> = [kt] as in *act*
<dr> = [dr] as in *draw*
<fl> = [fl] as in *flaw*
<fr> = [fr] as in *from*
<ft> = [ft] as in *soft*
<gl> = [gl] as in *gloom*
<gr> = [gr] as in *groom*
<ld> = [ld] as in *sold*
<lf> = [lf] as in *shelf*
<lt> = [lt] as in *belt*
<mp> = [mp] as in *camp*
<nch> = [nch] as in *branch*
<nd> = [nd] as in *brand*

<nk> = [ŋk] as in *sank*
<nt> = [nt] as in *sent*
<nth> = [nth] as in *tenth*
<pl> = [pl] as in *place*
<pr> = [pr] as in *price*
<pt> = [pt] as in *slept*
<qu> = [kw] as in *quarter*
<rch> = [rch] as in *march*
<rd> = [rd] as in *hard*
<rk> = [rk] as in *mark*
<rl> = [rl] as in *girl*
<rm> = [rm] as in *arm*
<rn> = [rn] as in *barn*
<rt> = [rt] as in *short*
<rth> = [rth] as in *birth*
<sc> = [sk] as in *scale*
<sch> = [sk] as in *school*

<scr> = [skr] as in *scrape*
<shr> = [shr] as in *shrill*
<sk> = [sk] as in *ask*
<sl> = [sl] as in *sled*
<sm> = [sm] as in *small*
<sn> = [sn] as in *sneak*
<sp> = [sp] as in *spoke*
<sph> = [sf] as in *sphere*
<spl> = [spl] as in *splash*

<spr> = [spr] as in *spring*
<squ> = [skw] as in *squeak*
<st> = [st] as in *last*
<str> = [str] as in *straw*
<sw> = [sw] as in *swell*
<thr> = [thr] as in *throw*
<tr> = [tr] as in *true*
<tw> = [tw] as in *twin*
<tz> = [ts] as in *quartz*

Doublets and Doublet Equivalents. You can also filter the Word field to the following consonant doublets and doublet equivalents, which spell a single consonant sound, usually after a short vowel, nearly always in word-medial position, and often due to the twinning of a final consonant when adding a suffix (as in *twinning* tw+n+ing¹) or the assimilation of final consonants in prefixes (as in *announce* a/d+n+nounce and *acquire* a/d+c+quire). These doublets and doublet equivalents are tagged **db** in the Analysis field:

<bb> = [b] as in *robber*
<cc> = [k] as in *accurate*
<ck> = [k] as in *rock*
<cq> = [k] as in *acquire*
<dd> = [d] as in *reddest* or *odd*
<dg> = [j] as in *bridge*
<dj> = [j] as in *adjourn*
<ff> = [f] as in *offer* or *off*
<gg> = [g] as in *bigger* or *egg*

<ll> = [l] as in *allow* and *tell*
<mm> = [m] as in *hammer*
<nn> = [n] as in *dinner* or *inn*
<pp> = [p] as in *happy*
<rr> = [r] as in *carry*
<ss> = [s] as in *missing* or *kiss*
<tch> = [ch] as in *catch*
<tt> = [t] as in *cotton*
<zz> = [z] as in *dizzy* or *fuzz*

Consonant Digraphs and Trigraphs. You can filter to consonant digraphs or trigraphs – two or three consonant letters that spell a single consonant sound, which are tagged **c2** and **c3** respectively in the Analysis field:

<ch> = [ch] as in *church* or [k] as in *echo*
<gh> = [f] as in *laugh* or [g] as in *ghost*
<ght> = [t] as in *night*
<ph> = [f] as in *phone*
<rh> = [r] as in *rhyme*

<rrh> = [r] as in *myrrh*
<sh> = [sh] as in *shirt*
<tch> = [ch] as in *witch*
<th> = [th1] and [th2] as in *thin* and *this*
<wh> = [h] as in *whole*, or [w] (or [ʰw]) as in *while*
<wr> = [r] as in *write*

Simplifications. And you can filter to the following simplifications, which retain the original longer spellings of one-time blends that have simplified over time to single consonant sounds. These consonant simplifications are tagged **cs** in the Analysis field:

<cht> = [t] as in <i>yacht</i>	<mn> = [m] as in <i>column</i>
<ft> = [f] as in <i>often</i>	<pb> = [b] as in <i>cupboard</i>
<ght> = [t] as in <i>light</i>	<ph> = [p] as in <i>shepherd</i>
<gn> = [n] as in <i>sign</i>	<ps> = [s] as in <i>psychology</i>
<kn> = [n] as in <i>knight</i>	<qu> = [k] as in <i>conquer</i>
<ld> = [d] as in <i>could</i>	<sc> = [s] as in <i>muscle</i>
<lf> = [f] as in <i>half</i>	<sl> = [l] as in <i>island</i>
<lk> = [k] as in <i>talk</i>	<st> = [s] as in <i>listen</i>
<lm> = [m] as in <i>calm</i>	<sth> = [s] as in <i>isthmus</i>
<ln> = [n] as in <i>Lincoln</i>	<sw> = [s] as in <i>sword</i>
<mb> = [m] as in <i>bomb</i>	<tg> = [g] as in <i>mortgage</i>

The analysis in CommonWords seldom speaks in terms of silent consonant letters, but if you want to work with silent letters, filtering the Analysis field on **cs** will return all of the words containing the simplifications listed above and thus provide words for work with what many would call silent consonants. For more on the treatment of silent letters in CommonWords see my *American English Spelling* (Baltimore: Johns Hopkins UP, 1988) (hereafter *AES*), pp. 54-55.

2. The Sound-to-Spelling Correspondences field gives all sound-to-spelling correspondences found in each word, in order. This field is primarily for teachers of spelling and writing – the idea being that spellers usually know the sound from the spoken language and are trying to find its spelling in the written language..

3. The Spelling-to-Sound Correspondences field gives all spelling-to-sound correspondences and is primarily for teachers of reading since readers have the spelling and are trying to find its sound to help identify it in the spoken language.

In the correspondences fields the equal sign translates to “is spelled with” or “spells”. Thus, in the Sound-to-Spelling Correspondences field “[k]=<c>” translates to “the sound [k] is spelled with the letter <c>”, and in the Spelling-to-Sound Correspondences field “<c>=[k]” translates to “the letter <c> spells the sound [k]”. Curly braces mark silent vowels, usually <e>: {D} marks silent letters that serve some diacritical function; {ND} marks silent letters with no diacritical function. Thus “{D}=<e>” indicates a diacritical silent <e>, as in *time* where it marks a long vowel, or *clothe* where it marks a voiced <th>, *ounce* where it marks a soft <c>, or *bronze*, *clause*, *league*, *active*, where it insulates a letter that normally doesn’t occur at the end of word-final bases. On the other hand, {ND}=<e> indicates a non-diacritical silent <e>, as in *fixed* [fikst]. and with the final <e> in words like *feature*. For more on the diacritical functions of silent final <e>, see *AES*, pp. 145-48 or “Spelling for Learning,” pp. 34-48 in the Short Articles venue of dwcummings.com.

Both correspondences fields allow various kinds of searches and filters. For instance, if you are dealing with phonics, you can filter to certain correspondences or to certain sounds or to certain spellings: At *meadow* the Sound-to-Spelling Correspondences field contains the following: “[m]=<m> [e1]=<ea> [d]=<d> [o2]=<ow>”. (You can find more about the sounds with numbers like [e1] and [o2] in the Sounds Count table.) Thus, you can filter to all the words in which [e1] – that is, short <e> – is spelled <ea> ([e1]=<ea>, 73 words), or to all words that contain short <e> ([e1]), however it’s spelled (1,363), or to the <ea> spelling (<ea>) – sometimes spelling short <e> (as in *meadow*), sometimes long <e> (as in *streak*), sometimes long <a> (as in *steak*), and sometimes schwa (as in *ocean*) (252 words).

4. The **Explication** field analyzes – or explicates – written words into their elements, or smallest meaning-bearing parts – that is, their prefixes, bases, and suffixes. It also shows any deletions, insertions, or replacements that occur when the elements combine – for instance, final <e> deletion in *hast/e+y* at *hasty*; final consonant insertion in *twin+n+ing*

at *twinning*, and replacement in [a/d+p+pear at appear and in the <y> to <i> replacement in acivities act1+iv/e]+it/y]+i2+es]2. For more information on the senses, functions, and relationships of elements given in Explication, you can consult the appropriate tables from CommonPrefixes, CommonSuffixes, and CommonBases.

This Explication field can be used to filter to words with various prefixes, bases, suffixes, and procedures. The following suggests some possible filter strings:

To find words that contain the prefix *de-* : **de+** (126 words)

To find words that contain the base *fect*: **fect** (12)

To find words that contain the verbal suffix *-ing*] : **+ing]1** (202)

To find words that contain the blend <sh> within a single element: **sh**, thus avoiding pesky false hits (215).

To find explications that contain the following procedures:

For instances of assimilation, filter the Analysis field on **ASSIM** (414);

for silent final <e> deletion filter Analysis on **DELE** (417);

for words that contain twinning or require twinning, filter Analysis on **TR** (368).

Since you can search on two or more fields at once, you can make your search quite specific to your grade level. For instance, If you were working with 5th and 6th grade students on silent final <e> deletion, you could filter the Analysis field on **DELE** and the Rank field on **C** to return 190 words that include instances of <e> deletion – *athletic*, *confident*, etc. The Explication field for each word shows the <e> deletion. For more on explication see *AES*, chapter 2, “The Explication of Written Words” 32-66 and “On Explication” in the Short Articles venue of dwcummings.com.

5. Analysis. The Analysis field lists several of the orthographically significant features in a word, each of which can be filtered to. For more details on these features see the references to *AES* given in parentheses below:

Tactical Strings and Rules:

CV# = Consonant + long vowel at the end of the word, stressed or unstressed, as in *by* and *many*

Cr# = Holdout to the CV# rule, with a reduced vowel at end of word, as in *larva*

CVC# = Consonant + stressed short vowel + consonant at end of the word, as in *bat* (AES, 93-94)

CLC# = Holdout to the CVC# rule, with a final long vowel, as in *control*

VCC = A stressed short vowel+consonant+consonant, as in *lettuce* (AES, 96-107). When not syllable-initial, <x> is treated here as two consonants; thus *tax* and *taxi* are tagged as containing a VCC string.

LCC = Holdout to the VCC pattern, with a long head vowel, as in *blind* (AES, 101-11)

VCV = A stressed long vowel+consonant+vowel, as in *vapor* and *rate* (AES, 96-100, 107-11)

SCV = A holdout to VCV pattern, with a short stressed vowel, as in *done* and *love* (AES, 107-11)

VCCle = A stressed short vowel in vowel+consonant+consonant+<le> string, as in *little* and *candle* (AES, 105-06)

VCle = A stressed long vowel in vowel+consonant+<le> string, as in *title* (AES, 105-06)

SCle = A holdout to VCle pattern, with short stressed vowel, as in *butler* (AES, 105-07)

VCr = A stressed long vowel in vowel+consonant+<r> string, as in *secret* (AES, 106)

SCr = A holdout to VCr pattern, with a stressed short vowel, as in *fabric*

VrV and Vrr = A version of the VCV and VCC patterns that involve the consonants <r> and [r]. For details on this complicated issue, see “<Vre> Spellings” in “Notes on the Vowel Analysis in CommonWords” in the Short Articles venue of dwcummings.com

V.V = A long vowel+vowel with a syllable boundary between them, as in *lion* and *create* (AES, 91-93)

FLR = Instance of the French Lemon Rule with a short head vowel in a VCV string, as in *lemon* and *consider* (AES, 127-28, where it is called the Stress Frontshift Rule).

3VR = Instance of Third Vowel Rule with a short head vowel in a VCV string three vowel sounds from the end of the word, as in *national*, as compared with *nation*. Sometimes the vowel in question is more than the third vowel from the end. (AES, 131-142, where it is called the Third Syllable Rule)

Suffix Rules:

IC = An instance of the Suffix *-ic* Rule, with a short head vowel in a VCV string preceding the suffix *-ic*, as in *athletic*, *critic*, *panic* (AES, 115-18)

LC = A holdout to the Suffix *-ic* Rule, with a long head vowel in a VCV string, as in *aerobic* (AES, 116-18)

ION = An instance of the Suffix *-ion* Rule (AES, 118-19), with short <i> or long <a, e, o, u> as head of VCV string preceding the suffix *-ion* as in *addition*, and *formation*, *completion*, *emotion*, and *conclusion*, (AES, 118-19)

SIO = A holdout to the Suffix *-ion* Rule, with a short vowel preceding the suffix *-ion*], as in *companion* and *discretion*

IT = An instance of the Suffix *-it* Rule with a short vowel in a VCV string preceding the suffix *-it* (AES, 120), as in *credit*, *limit*, and *visit*

LT = A holdout to the Suffix *-it* Rule, with a long head vowel, as in *unit*

TY = An instance of the Suffix *-ity* Rule, with short head vowel in a VCV preceding the suffix *-ity* as in *sanity*, as compared with *sane* (AES, 112-15)

Letter Strings:

cb = Contains a consonant blend, like the <nt> in *agent*

c2 = Contains a consonant digraph, two consonant letters spelling a single consonant sound, as in *with* (AES, 71-72)

cs = Contains a consonant simplification – two or more consonant letters that spell a single consonant sound due to a simplifying sound change, as <mb> at the end of *bomb*

c3 = Contains a consonant trigraph, as in *witch*

db = Contains a consonant doublet or doublet equivalent, as the <ss> in *kiss* and the <dg> in *grudge*

vd = Contains a vowel digraph, as in *head*. Spellings of diphthongs are tagged as vowel digraphs.

vt = Contains a vowel trigraph, as the <iou> in *ambitious*

Procedures:

ASSIM = Contains an assimilation of the final consonant in a prefix, as in *concert* [co/m+n+cert] (AES, 177-98)

CMP = A compound word, as with *baseball*

DELE = Contains an instance of silent final <e> deletion, as in *devotion* [de+vot~~e~~+ion]. (AES, 145-60)

RDEL = A word that requires <e> deletion when adding a suffix beginning with a vowel, as with *bake*.

DL! = Contains a nonregular deletion of final <e>, as in *argument* [argu~~e~~+ment] (AES, 158-59) or other unusual deletions.

EXS = Contains a deletion of <s> after prefix [ex-, as in *expect*,

[ex+\$pect

i>y and y>i = Contains instances of the <i> to <y> change, as in *lying* from *lie*, or of the <y> to <i> change, as in *tries* from *try*. Also includes words with derived forms that would involve these changes and instances of <y> and <i> deletion (AES, 84-87,157)

PELE = Contains an instance of penultimate <e> deletion, as in *angry*, angér+y]

SWR = An instance of the Short Word Rule, as in *egg* and *pie*, with double final consonant or silent final <e> added to avoid words of less than three letters (AES, 87-89)

TR = Contains an instance of twinning or has derived forms with twinning, as in *batter* or *bat*.

CTR = A word that contains an instance of twinning. (AES, 161-76)

Some Uses of the ANALYSIS Field. If you are working with vowel digraphs, you can filter the Analysis field on **V.V** – to find words that contain two adjacent vowel sounds that are separated by a syllable boundary, as in *diet*. Filter the Syllables field on **2** to keep things simple. This search returns 37 words like *client*, *create*, *lion*. One activity could be to get the youngsters to identify the V.V, which should be quite easy. Then ask them to give you a word in which those two vowel letters are a vowel digraph – as in, *piece*, *bread*, *nation*, which can be much harder. If you need help, you can consult the Correspondences: Spellings to Sounds table. Or go the other direction, from vowel digraphs to instances of V.V. Here are some digraphs that can also be V.V strings: <ea>, <ei>, <ie>, <oe>, <oi>, <ue>, and <ui>.

If you are working with older students on some of the processes involved when elements are combined to form words: To get a list of words containing doublets or doublet equivalents appropriate for older students, filter the Analysis field on **db** and the Rank field on **D**. This filter returns 165 words, some of which raise more complications than you need to tackle. If you add Syllables contains **2** to the filter, you get 66 words with fewer complications. A first step in the discussion could be to ask what the doublet or doublet equivalent is in a given word – which in words like *kiss* is pretty obvious, but a bit more challenging in, say, *scene* or *acquittal* (where

there are actually two, <cq> being the equivalent of a double <q>). A second, and more difficult, question could be why that doublet or doublet equivalent is there instead of a single consonant. It could be because of the VCC pattern as in, say, *funnel*, or because of twinning, as in *acquittal* [a/d+c+quit+t+al]2,, as shown in the Explication field, which also explains the <cq> because of assimilation. Some doublets are due to a little-known rule in English called the Short Word Rule, according to which only function words – such as prepositions and conjunctions – and extremely common verbs, like *is* and *go*, can be two letters long. One way of obeying the Short Word Rule is by adding a second consonant, as in *add*, *egg*, *inn*, and *odd*. (Sometimes a final <e> is added, as in *awe*, *die*, *eye*, and *wee*.) A list of words whose spelling is affected by the Short Word Rule is returned if you filter the Analysis field on **SWR**.

6. Themes. In this field more than 7,560 words are tagged for 169 themes, or topics, with which they can be associated. It is intended to be useful for generating word lists dealing with a common theme, such as “Colors” or “Sports”. There is nothing very authoritative or exhaustive about these taggings. Subjective judgements abound, and occasional violence is done to some formal, scientific categories. All I can say is that on at least one day, one retired English teacher saw each word plausibly belonging to the various themes for which it was tagged.

Due to homography, as a given form moves from one theme to another, it often becomes a different word. For instance, the form <molar> “chemical measure” in the Science4 theme is a homograph of the form <molar> “tooth” in the Anatomy1 theme – that is, an entirely different word with the same spelling, a homograph.

The following is a full list of the themes, many of which are organized into groups, tagged with a group name and a numerical index: **Anatomy1**, **Anatomy2**, etc. The description of each theme concludes with a parenthesis containing two example words and the number of words tagged for that theme. A more compact and orderly presentation of the following information can be found in the data table titled **Themes**, with separate fields for the eight grade levels, and for substantives (S), adjectives (J), verbs (V), adverbs (B), and prepositions (E).

Anatomy1 lists words dealing the skeletal and muscular systems of the body

(*ankle, vertebrate*; 111).

Anatomy2 lists words dealing with the organs, genes, and glands (*skin, gastrointestinal*; 99).

Anatomy3 lists words dealing with fluids and other substances within the body (*blood, insulin*; 33).

Animals1 lists birds and things associated with birds (*crow, hatch*, 52);

Animals2 does the same for insects (*beetle, hive*, 34),

Animals3 for warm-blooded animals, (but not birds) (*kangaroo, hominid*, 106),

Animals4 for cold-blooded reptiles and fish, and a few others that actually do not contain blood, like mollusks and sponges (*oyster, invertebrate*, 34).

Animals5 lists miscellaneous words dealing with animals in general (*hibernate, zoo*, 31).

Archaic lists words that were common in earlier English but are now encountered mostly in early texts, such as the King James Bible (*couldst, spake*, 21).

Art1 lists words dealing with print and literature (*haiku, manuscript*, 132);

Art2 words referring to stage and film art (*actress, playwright*, 63);

Art3 words referring to musical instruments and voices (*baritone, guitar*, 50);

Art4, words dealing with musical types and qualities (*jazz, allegro*, 68);

Art5, words dealing with the visual arts: painting, sculpture, architecture, etc. (*impressionism, architecture*, 94);

Art6, words dealing with miscellaneous aspects of the world of music (*Beethoven, octave*, 68).

The Business group deals with the world of business and commerce:

Business1, words for grades one and two (*market, shop*, 74);

Business2, words for grade three (*capital, credit*, 76);

Business3, grade four (*career, employer*, 85);

Business4, grades five and six (*international, promotion*, 128);

Business5, grades seven and eight, including several words from Hirsch et al's *Dictionary of Cultural Literacy* (*bonus, inventory*, 79).

Business6, more business and commerce words from Hirsch (*bureaucrat*, 45).

Calendar lists the names of months, weekdays, and holidays, as well as periods in the day (*Friday, Halloween, 54*).

Cities lists the names of cities of the world (*Venice, Philadelphia, 46*).

Clothing1 lists specific articles of clothing (*hat, shirt, 44*).

Clothing2 lists accessories and parts of clothing (*collar, lace, 50*).

Colors lists the names of colors and their qualities (*green, bright, 53*).

Communication1 lists verbs that deal with the various functions or uses of communication acts (*announce, persuade, 92*).

Communication2 lists nouns that refer to various products or end results of communication (*argument, understanding, 69*).

Communication3 deals with miscellaneous methods, aspects, and qualities of communication (*media, conciseness, 93*).

Containers lists various kinds and attributes of containers (*box, enclose, 54*).

Countries lists proper nouns that name countries (*Japan, Italy, 39*).

Crime1 lists types of crime and criminal (*embezzlement, murderer, 69*).

Crime2 lists words about various people and things involved in the law and justice system (*police, court, 101*).

Crime3 a miscellaneous group of things and qualities involved with crime in general (*contraband, offense, 75*).

Entertainment1 lists types of entertainment and entertainers (*magician, roulette, 69*).

Entertainment2 lists actions of people who are being entertained and the effects entertainment has on them (*mirth, excitement, 65*).

Entertainment3 lists miscellaneous words that refer to entertainment in one way or another (*costume, audience, 101*).

Family1 lists members of a family (*parent, sister, 67*).

Family2 lists actions, events, qualities, and things relating to families (*birthday, inheritance, 85*).

Farming1 lists things raised on farms (*crops, mutton, 64*).

Farming2 lists equipment and workers found on farms and the things they do (*tractor, pesticide, 53*).

Farming3 lists words that refer to miscellaneous things related to farming (*meadow, graze, 59*).

Feeling1 lists nouns that refer to positive feelings (*confidence, vigor, 79*).

Feeling2 lists nouns that refer to negative feelings (*fright, woe, 88*).

Feeling3 lists positive adjectives (*beautiful, fearless, 106*).

Feeling4 lists negative adjectives (*greedy, treacherous, 95*).

Feeling5 lists positive verbs (*enjoy, trust, 43*).

Feeling6 lists negative verbs (*distress, vex, 65*).

Feeling7 lists miscellaneous words related to various aspects of feelings, including a number of adverbs (*happily, wish, 97*).

Food1 lists words dealing with fruits and nuts (*strawberry, walnut, 44*).

Food2, lists grains and bread (*wheat, biscuit, 32*).

Food3 lists meat, fish, poultry, and dairy products (*beef, egg, 60*).

Food4 lists sweets (*cookies, pudding, 32*).

Food5 lists vegetables (*turnip, potato, 32*).

Food6 lists drinks (*juice, pop, 50*).

Food7 lists miscellaneous words dealing with food (*eating, buffet, 90*).

The Gender group lists words dealing with gender, sex, and sex difference:

Gender1 lists words that mark the distinction between male and female for people and other creatures (*son, daughter, 83*).

Gender2 lists words dealing with sex and reproduction (*conception, penis, 52*).

Gender3 lists words dealing with miscellaneous aspects of gender and sexuality (*herpes, sexism, 39*).

Geography1 lists common and proper nouns referring to geographical places, excluding countries and cities (*Europe, planet, 28*).

Geography2 lists nouns appropriate for grades one through four that refer to natural geographical features (*desert, ocean, 53*).

Geography3 lists nouns for grades five and up that refer to natural geographical features (*bayou, isthmus, 63*).

Geography4 lists miscellaneous words dealing with geography (*environment, geology, 73*).

Government1 lists nouns that refer to people and groups involved in the

governing process (*king, congress*, 111).

Government2 lists mostly abstract nouns (and a few modifiers) that are appropriate for grades one through six and refer to types of government, their aspects and qualities (*democracy, rights*, 52).

Government3 lists more abstract nouns and modifiers that are appropriate for grades seven and eight (*communism, impeachment*, 92).

Government4 lists words that deal with the process of governing (*campaign, filibuster*, 90).

Government5 lists verbs that refer to actions of governments (*appoint, install*, 38).

Government6 lists miscellaneous words dealing with government and governing (*bandwagon, gerrymander*, 102).

Groups1 lists words that refer to groups that always, or at least usually, contain people (*junta, panel*, 82).

Groups2 lists words that refer to all other kinds of groups (*bunch, litter*, 97).

The Health group includes words dealing with health, sickness, and death:

Health1 lists words dealing with medications and drugs (*antibiotic, cortisone*, 36).

Health2 lists words dealing with care and treatment (*dentist, hospital*, 87).

Health3 lists nouns that refer to strictly or mostly mental conditions (*hysteria, phobia*, 50).

Health4 lists words dealing with physical illness, diseases, and death (*asthma, cardiac*, 113).

Health5 lists adjectives dealing with health (*mortal, tender*, 67).

Health6 lists verbs (*relieve, suffer*, 63).

Health7 lists nouns (*calorie, injury*, 81).

History1 lists nouns and adjectives dealing with American history (*colonial, Lincoln*, 72).

History2 lists nouns and adjectives dealing with ancient history (*classical, Troy*, 30).

History3 lists nouns and adjectives dealing with European history (*knight, Napoleon*, 67).

Home1 lists movable furniture and furnishings found in the home (*couch, blanket*, 68).

Home2 lists fixtures, rooms, and spaces (*ceiling, den*, 68).

Home3 lists miscellaneous words associated with house and home (*address, deed, 105*).

Language1 lists words dealing with: grammar, spelling, word structure, parts of speech, and punctuation (*alphabet, noun, 102*).

Language2 lists words dealing with semantics and meaning (*dictionary, meaning, 34*).

Language3 lists words dealing with the spoken language and pronunciation (*homophone, pronounce, 37*);.

Language4 lists words dealing with rhetoric, or the uses of language and its effects (*argument, slang, 57*).

Language5 lists miscellaneous words dealing with language, including the names of various languages (*Japanese, printing, 42*).

Light1 lists verbs about light and its qualities (*gleam, reflect, 39*).

Light2 lists nouns (*moonlight, sheen, 34*).

Light3 lists adjectives and adverbs (*brilliant, intense, 24*).

The Location group includes locations, positions, and directions:

Location1 lists words that are either prepositions or modifiers or both and are appropriate for grades one and two (*beyond, under, 61*).

Location2 lists prepositions or modifiers appropriate for grades three and four (*opposite, wherever, 35*);

Location3 lists prepositions or modifiers appropriate for grades five through eight (*offshore, underground, 34*).

Location4 lists other location words appropriate for grades one through three (*corner, middle; 53*);

Location5, appropriate for grades four through six (*latitude, suburb, 44*);

Location6, appropriate for grades seven and eight (*perigee, longitude, 15*).

Materials1 lists metals and metallic materials (*alloy, wire, 24*).

Materials2 lists minerals and mineral-like materials (*coal, pearl, 52*);

Materials3 lists materials from vegetable matter or from animals, including from petroleum (*charcoal, leather, 62*).

Materials4 lists miscellaneous words that deal with materials and are hard to fit into any of the preceding three (*stuff, plastic, 30*).

Math1 lists number names (*digital, sixteen, 74*)

Math2 lists mathematical concepts and calculations (*equal, multiply, 59*).

Math3 lists miscellaneous math words appropriate for grades one through four (*pair, problem, 19*).

Math4 lists miscellaneous words for grades five and six (*plus, subset, 31*).

Math5 lists miscellaneous words for grades seven and eight, including several from Hirsch *et al* (*axiom, linear, 65*).

Measure1 lists adverbs dealing with amounts, degree, and sizes (*often, loudly, 64*).

Measure2 lists adjectives (*enormous, abundant, 82*).

Measure3 lists nouns (*amount, handful, 75*).

Measure4 lists words dealing with calculated measurements (*breadth, frequency, 79*).

Measure5 lists units of measurement, including monetary units (*dollar, inning, 96*).

Military1 lists words about military paraphernalia and equipment (*helmet, rocket, 53*).

Military2 lists military personnel (*captain, regiment, 80*).

Military3, lists military actions and operations (*raid, maneuver, 80*).

Military4 lists miscellaneous military words (*honorable, strategic, 72*).

Mind1 lists nouns (mostly rather advanced) that refer to types and schools of intellection or thought (*philosophy, science, 51*).

Mind2 lists verbs referring to various mental acts (*believe, calculate, 76*).

Mind3 lists nouns referring to the results of mental acts (*discovery, certainty, 142*).

Mind4 lists words dealing with miscellaneous aspects of the mind and mental acts, including psychological constructs (*meanings, irrational, 57*).

Myth lists names and qualities of myths and mythological figures, ancient and modern (*phoenix, werewolf, 56*).

Names lists common names, both first and surnames (*Dorothy, Franklin, 88*).

Occupation1 lists nouns with the agent suffixes *-ar]2, -er]01, -or]2, or -ess]1* (*advisor, farmer, 79*).

Occupation2 lists other occupational nouns (*housewife, politician, 99*).

People1 lists nouns and pronouns appropriate for grades one and two that refer to individual people – their roles, jobs, demeanors (*group, officer, 91*).

People2 lists such words for grade three (*chum, nurse, 101*).

People3 lists such words for grade four (*bride, follower, 181*).

People4 lists such words for grades five and six (*fugitive, magician, 258*).

People5 lists words for grades seven and above, including words from Hirsch *et al.* (*baritone, extrovert, 150*).

Plants1 lists nouns and adjectives appropriate for grades one through four that refer to or describe plants and their parts (*cotton, leaves, 64*).

Plants2 lists nouns and adjectives for grades five through eight (*deciduous, lavender, 57*).

The Religion group lists words dealing with various aspects of religion and religions, including Christianity, Judaism, Islam, Buddhism, Hinduism, and related areas.

Religion1 lists words appropriate for grades one through four that deal with various aspects of religions and religiousness (*blessing, faith, 103*).

Religion2 lists such words for grades five and six (*sermon, eternity, 88*).

Religion3 lists words from for grades seven and above, including words from Hirsch *et al.* (*godliness, heretic, 124*).

School1 lists verbs dealing with school and schooling (*read, subtract, 64*).

School2 lists nouns, and a few adjectives, appropriate for grades one through four (*class, excellent, 83*).

School3 lists nouns, and a few adjectives, appropriate for grades five through eight, including words from Hirsch *et al.* (*fraction, calculator, 70*).

Science1 lists nouns referring to kinds of science, and a few quasi-sciences, (*chemistry, alchemy, 36*).

Science2 lists nouns and adjectives dealing with biology and appropriate for grades one through four (*feather, gene, 59*).

Science3 lists biological nouns and adjectives for grades five through eight (*bacteria, evolution, 150*).

Science4 lists nouns and adjectives for chemistry (*element, oxygen, 91*).

Science5 lists nouns and adjectives for physics and astronomy (*comet, particle, 152*).

Science6 lists verbs, and a few nouns, referring to scientific actions and processes (*analyze, experiment, 63*).

Science7 lists miscellaneous nouns and adjectives dealing with science and technology (*laboratory, formula, 81*).

Senses1 lists words dealing with speech and hearing (*ear, loud, 63*).

Senses2 lists words dealing with sight (*look, visible, 30*).

Senses3 lists words dealing with smell and taste (*nostril, sweet, 19*).

Senses4 lists miscellaneous words dealing with other senses and senses in general (*extrasensory, pain, 46*).

Sports1 lists sports equipment (*diamond, ball, 64*).

Sports2 lists other sports words appropriate for first and second grades (*game, race, 70*).

Sports3 lists sports words for third grade (*league, underdog, 70*).

Sports4 lists words for fourth grade (*eagle, surf, 73*).

Sports5 lists words fifth and sixth grades (*handicap, skiing, 75*).

States lists the names of states (*Michigan, Oregon, 51*).

Time1 lists time words for grades one and two (*hour, soon, 59*).

Time2 lists words for grades three and four (*clock, sunset, 61*).

Time3 lists words for older students (*eternity, prompt, 64*).

Tools1 lists words about tools (defined rather broadly) for grades one through three (*jack, phone, 42*).

Tools2 lists tool words for grades four through eight (*computer, hydraulic, 77*).

Transportation1 lists words dealing with vehicles and other means of transportation (*bicycle, shuttle, 52*).

Transportation2 lists words dealing with routes, roads, times, and places (*interstate, arrival, 51*).

Transportation3 lists other transportation words appropriate for grades one through four (*freight, passenger, 66*).

Transportation4 lists words for grades five through eight (*gasohol, supersonic, 43*).

Trees lists the names and other features of trees (*birch, forest, 54*).

Value1 lists nouns that refer to qualities to which we ascribe subjective values, good or bad. (*curse, friendship*, 169).

Value2 lists value-laden adjectives (*excellent, ugly*, 173).

Value3 lists value-laden verbs (*forgive, pollute*), 109).

Weather1 lists nouns that refer to weather and climate (*cloud, meltdown*, 82).

Weather2 lists weather and climate adjectives (*dusty, tropical*, 27).

Some Uses of the Themes Field. Assume you want a list of nouns that deal with plants and are appropriate for fourth graders. Filter the Themes field on **Plants1**; filter the Rank field on **B** for 4th grade. Filter the Parts of Speech field on **rs**, for regular nouns. This search returns 37 regular nouns dealing with plants and appropriate for fourth graders, including *ash, blossom, cherry, daisy, limb, moss, needle, orchard, pitch, reed, stump, timber, violet, weed*.

If you would like some activities to heighten students' semantic awareness for vocabulary study: Filter the Themes field on **Feeling1**, positive feelings, and if you're working with 5-6 graders, filter the Rank field on **C**, to return 33 words like *certainty, vigor, zeal*. One activity could be to ask at what point would, say, certainty, become a negative thing? And what would you call it? Though it's an open question, you would be looking for words like *bull-headedness, close-mindedness*, etc. And then you could have the students go on to discuss exactly what the tipping point might be, giving examples of when certainty can be said to become plain old bullheadedness.

You could work the other way, filtering the Themes field on **Feeling4**, negative feelings, and the Rank field on **C**, which returns 46 words like *doubtful, furious, guilty*, etc. The idea here would be to get students to discuss when such negative feelings could become a good thing – for instance, being doubtful could be a good thing if it makes you cautious and inquisitive.

7. Homophones. Homophones are words that sound the same but mean different things and are spelled differently, as with *pear* and *pare*. They can pose special problems for spellers and can benefit from some special attention.

Some Uses of the Homophones Field: Filter the Homophones field on **is NOT null** (to return words tagged as homophones), and filter the Rank field on **B** (appropriate for fourth graders). This filter returns 150 homophones such as *alter*, *baron*, *brake*. One activity could be to give students one of the words and ask them for another word that sounds the same as the given word: What is it? How is it spelled? What does it mean? Use it in a sentence.

8. Homographs. Homographs are words that are spelled the same but that mean different things and are usually have different pronunciations. They pose no particular problems for spellers, but they can for readers. Many of them contrast in pronunciation simply by shifts of stress and contrast in meaning simply by shifts in part of speech – for instance, *convict*, a noun with stress on the first syllable vs. *convict*, a verb with no stress on the first syllable.

Some Uses of the Homographs Field. Filter the Homographs field on **is NOT null**. Filter the Syllables field to **1** (to avoid getting a lot of two-syllable noun-verb pairs like the two *convict*'s). This filter returns 46 homographs like *bear*, *bow*, and *does*. Discussion questions could be “What other word is spelled <bear>? What does it mean? What other word is spelled <does> and what does it mean?”

9. Other Problem Spellings. This field is a companion to Homophones and Homographs, covering a variety of problems that could benefit from some special attention. It lists near homophones and non-homophonic look-alike words – such as *accept* vs. *except*, *latter* vs. *later*, and *angle* vs. *angel*. Common misspellings are tagged with an asterisk. Words tagged with an exclamation point in this field appear on at least one list of spelling demons.

10. Spelling Difficulty. This field lists the level of difficulty for nearly half of the words in CommonWords, based on the percentages of fourth graders who spelled the given word correctly in Harry Andrew Greene, *The New Iowa Spelling Scale* (Iowa City: State University of Iowa, 1954, 1977). A suggested categorization would be:

- 1-4 = Very hard (361 words)
- 5-13 = Hard (796)

14-47 = Medium (1,507)
48-71 = Easy (756)
72-99 = Very Easy (382)

11. Rank. This field is meant to help in deciding when to introduce certain words to students. It is generally based on Edward Thorndike and Irving Lorge's *Teacher's Word Book of 30,000 Words* (New York: Teachers College Press, 1944, 1972) (hereafter T-L), which is primarily aimed at readers rather than spellers. The T-L score used here is that given in the "G" column in their list, which gives the number of occurrences per one million running words. T-L suggests appropriate grade levels:

- AA** = Appropriate for grades 1-2.
- A** = Appropriate for grade 3
- B** = Appropriate for grade 4 (A T-L score of 49-20)
- C** = Appropriate for grades 5-6 (A T-L score of 19-10)
- D** = Appropriate for grades 7-8 (A T-L score of 9-1)

For more on T-L's rankings see T-L, pp. x-xii. Words with a T-L score between 1 and 6 or that have no T-L score are assigned to a grade level based on my informed best guess, supported with the rankings in the *The American Heritage Word Frequency Book* (John B. Carroll et al, eds., Boston: Houghton Mifflin, 1971). The original T-L scores are based strictly on frequency; my assignments try to balance frequency with difficulty. Obviously these assignments are quite approximate.

T-L normally does not list inflected forms separately. The score it lists for the base form sums up all inflected and non-inflected forms. Since CommonWords does list many inflected forms separately, I've chosen usually to give the inflected forms the same ranking as that of the base form listed in T-L. Exceptions to this procedure are cases where there is a complication in the spelling of the inflected form (that is, a deletion, a twinning, or a change of <y> to <i> or of <i> to <y>), in which cases I've adjusted the ranking of the inflected form up one level so that the inflected form of **AA** words becomes **A**, and those of **A** words become **B**. I did not make this adjustment on words ranked **B** or higher.

The 2000+ words tagged **H** in the Rank field are a special group. Those tagged simply **H** do not appear in T-L's main word list, but are drawn from

E. D. Hirsch *et al's Dictionary of Cultural Literacy* (Boston: Houghton, Mifflin, 1988) (hereafter Hirsch). Words from Hirsch that occur in T-L are also tagged with their normal T-L score; all of those with a T-L score between 9 and 1 are tagged **DH**. The words from Hirsch play a particularly important part in what Hirsch and his team call cultural literacy, the “common knowledge or collective memory [that] allows people to communicate, to work together, and to live together” (p. ix). Obviously, the same could be said for all of the words in CommonWords (and many, many others), but the words tagged **H** have a special importance. They are only a sample, for Hirsch includes thousands of other words and phrases, including many proper names of people, places, events, and things that are for the most part excluded from CommonWords. I have included these **H** words because I believe it is important for students to be exposed to such words as soon as possible, even though they are often quite technical and advanced.

12 and 13. Range and Subrange. The Range field indicates into which of five ranges each of 5,680 tagged words falls. Ranges are intended to provide help in finding words appropriate to the students' level of mastery. For instance, the 1,070 words tagged **1** in Range are all completely regular and completely analyzable if the students have had work with the Range 1 sound-to-spelling correspondences listed below. The ranges are organized so that each of the first four ranges contains only one spelling for each sound and only one sound for each spelling. This regularity is not true of the correspondences in range 5, due to the existence of several sounds that have more than five different spellings.

Subranges are subsets of ranges. In the Subrange field, words tagged **1A** contain only the consonant and short vowel correspondences from range 1, words with the regular patterns for short vowels – namely, VCC and VC#. Words tagged **1B** contain only the consonant and long vowel correspondences from range 1, and the regular patterns for long vowels – VCe#, VCV, and several digraphs. Words tagged **2a** are range 2 words that contain only the ranges 1 and 2 consonant and short vowel correspondences. Words tagged **2b** contain only ranges 1 and 2 consonant and long vowel correspondences.

Range 1:

The Short Vowels:

[a1] = <a> as in *pat*

[e1] = <e> as in *pet*

[i1] = <i> as in *pit*

[o1] = <o> as in *pot*

[u1] = [u] as in *but*

The Long Vowels and Diphthongs:

[a2] = <a...e> as in *mate*

[e2] = <ee> as in *meet*

[i2] = <ie> and <i...e> as in *pie* and *pile*

[o2] = <oe> and <o...e> as in *woe* and *quote*

[u2] = <oo> as in *boot*

[yu2] = <ue> and <u...e> as in *hue* and *huge*

[oi] = <oi> as in *foil*

[ou] = <ou> as in *foul*

The Consonants:

[b] = as in *bob*

[d] = <d> as in *dad*

[f] = <f> as in *fluff*

[g] = <g> as in *gag*

[h] = <h> as in *hot*

[j] = <j> as in *jot*

[k] = <c> as in *cat*

[l1] = <l> as in *lot*

[m] = <m> as in *mom*

[n1] = <n> as in *nun*

[ng] = <ng> as in *bring*

[p] = <p> as in *pop*

[r] = <r> as in *roar*

[s] = <s> as in *sit*

[t] = <t> as in *tot*

[v] = <v> as in *vine*

[w] = <w> as in *wine*

[y] = <y> as in *yet*

[z] = <z> as in *zip*

[ch] = <ch> as in *chin*

[sh] = <sh> as in *shin*

[th1] = <th> as in *thin*

This may seem like a lot of correspondences, but notice that in nearly every case the spelling uses the same letter as we normally use to symbolize the sound. The symbol "...e>" indicates that the long vowel letter is followed by a single consonant letter and a silent final <e>, which is marking the long vowel sound, as in *mate*. Most of these correspondences are very high frequency. Vowels that precede [r] often vary considerably in their pronunciation from that when they precede some other consonant. Consider, for instance, the different pronunciations of <a> in *mare* and *mate*.

Range 2. The 916 range 2 words are completely regular and analyzable if the students have had work with the range 1 correspondences and the following 33:

The Short and Reduced Vowels:

[e1] = <ea> as in *bread*

[i1] = <e> as in *basket*

[o1] = <a> as in *ball*

[u1] = <o> as in *from*

[u3] = <oo> as in *wood*

[u4] (schwa) = <a> as in *allow*

The Long Vowels and Diphthongs:

[a2] = <ai> as in *rain*

[e2] = <e...e> as in *theme*

[i2] = <y...e> as in *type*

[o2] = <oa> as in *boat*

[u2] = <ue> and <u...e> as in *due* and *dune*

[yu2] = <ew> as in *few*

[oi] = <oy> as in *coy*

[ou] = <ow> as in *cowl*

[a3r] = <air> as in *hair*

[o3r] = <or> as in *cord*

The Consonants:

[b] = <bb> as in *ribbon*

[d] = <dd> as in *ridden*

[f] = <ff> as in *stuff*

[g] = <gg> as in *rugged*

[j] = <g> as in *large*
[k] = <k> as in *lake*
[l1] = <ll> as in *tall*
[m] = <mm> as in *summer*
[n1] = <nn> as in *runner*
[ng] = <n> as in *brink*
[p] = <pp> as in *happy*
[r] = <rr> as in *marry*
[s] = <c> as in *cent*
[t] = <tt> as in *attic*
[w] = <u> as in *quit*
[y] = <i> as in *onion*
[z] = <s> as in *dogs*
[ch] = <tch> as in *catch*
[sh] = <s> as in *sure*
[th2] = <th> as in *then*

It would be good, though not necessary, for the students to have worked with the reasons for double consonant letters: twinning, the assimilation of consonants at the end of prefixes, simple addition, and the VCC tactical pattern.

Range 3. The 1,114 range 3 words are completely regular and analyzable if the students have had work with ranges 1 and 2 and with the following correspondences and tactical patterns:

The Vowels:

[a1] = <au> as in *laugh*
[i1] = <y> as in *system*
[o4] = <aw> as in *law*
[u3] = <u> as in *put*
[a2] = <ay> as in *day*
[e2] = <ea> as in *speak*
[o2] = <ow> as in *low*
[u2] = <o...(e)> as in *move*
[yu2] = <eu> as in *feud*
[u4] = <e> as in *children*
[u4r] = <er> as in *batter*

The Consonants.

[f] = <gh> as in *laugh*
[h] = <wh> as in *whole*
[j] = <d> as in *graduate*
[k] = <ck> as in *pick*
[r] = <wr> as in *write*
[s] = <ss> as in *miss*
[z] = <zz> as in *buzz*

In addition to these sixteen correspondences range 3 words assume that the students have had work with two tactical patterns for long vowels: (i) the stressed head vowels of VCV strings are normally long – for instance, the <a> in *bacon* spells [a2], long <a>, and (ii) vowels at the end of syllables are also regularly long – for instance, the <i> in *lion* spells [i2], long <i>. The first of these two, which is essentially an extension of the range 1 and 2 correspondences with “...e>”, is discussed in chapter 4 of *AES* as the VCV pattern, the second as the V.V pattern.

Range 4. The 1,258 range 4 words are completely regular and analyzable if the students have had work with ranges 1, 2 and 3 and with the following correspondences and tactical patterns:

The Vowels:

[i1] = <a> as in *chocolate*
[o5r] = <ar> as in *hard*
[o4] = <au> as in *sauce*
[u] = <oo> as in *blood*
[a2] = <ea> as in *break*
Unstressed [e2] = <y> as in *funny*
Stressed [e2] = <ie, ei> as in *piece, receive*
[u2] = <ew> as in *drew*
[u4] = <io> as in *region*
[u4l] = <le> as in *jungle*
[u4r] = <or> as in *doctor*
[yu4] = <u> as in *deputy*
[yu3r] = <ur...(e)> as in *cure*

The Consonants:

[f] = <ph> as in *telephone*
[j] = <dg> as in *judge*
[ks] = <x> as in *fix*

[k] = <q> as in *quit*
[n1] = <kn> as in *know*
[r] = <rh> as in *rhythm*
[s] = <sc> as in *scene*
[sh] = <t> as in *nation*

In addition to these eighteen correspondences range 4 words assume that the students have worked with silent final <e>'s that serve various diacritical functions other than marking long vowels and with silent final <e>'s that serve no diacritical function at all. It also assumes familiarity with the <i>-before-<e> pattern. Holdouts to this pattern with <ei> are included in range 5.

Range 5. The 1,320 Range 5 words are completely regular and analyzable if the students have had work with ranges 1, 2, 3, and 4 and with the following correspondences and tactical patterns:

The Vowels.

[a3r] = <are> as in *rare*
[a1r] = <ar> as in *tariff*
[a1r] = <arr> as in *carriage*
[e2] = <ei> not after <c> as in *neither*
[e2] = <i> as in *machine*
[u4] = <i> as in *horrible*
[u4] = <o> as in *million*
[u4] = <u> as in *awful*
[u4] = <ou> as in *courteous*
[u4r] = <ar> as in *coward*
[u4r] = <ur> as in *injury*

The Consonants.

[gz] = <x> as in *exact*
[k] = <cc> as in *account*
[k] = <ch> as in *school*
Syllabic [l] = <l> as in *battle*
[u1r] = <ear> as in *earth*
[u1r] = <er> as in *term*
[u1r] = <ir> as in *firm*
[u1r] = <our> as in *courage*
[u1r] = <ure> as in *sure*

[t] = <ght> as in *night*
[hw] = <wh> as in *why*
[ch] = <t> as in *feature*
[sh] = <c> as in *social*
[sh] = <ss> as in *mission*
[zh] = <s> as in *casual*

Range 5 words also assume some work with the VCle# long vowel pattern, with the apostrophe, and with non-diacritical, non-final silent <e>'s.

14. The **CHARACTERS** field lists the number of characters (letters, punctuation marks, and blank spaces) in each word.

15. The **Syllables** field lists the number of syllables in each word. Some words have variant pronunciations with different numbers of syllables – for instance, one pronunciation of *average* has three syllables, another has only two, so the Syllables field shows both: **3 2**. A few final syllables are quite weak, consisting of only a syllabic consonant, as in *button* and *little*.

Some Uses of the Syllables Field. If you are working with primary students on silent final <e> and its various functions, you could filter the Syllables field on **1**, the Rank field on **A**, and the Word field on **ends with e**. This three-dimensional filter returns eight false-positives with non-silent final <e> (for example, *be, me, she, the, we*) and 245 words like *bake, breathe, choice, urge, false, love, league, seize, owe*, illustrating the various diacritic functions of silent final <e>: respectively marking a long vowel, marking a voiced <th>, marking a soft <c> or <g>, insulating an otherwise final <s>, <v>, <u> or <z>, complying with the Short Word rule. For more on the functions of silent final <e>, see *AES*, pp. 145-54 or “Spelling for Learning” (pp. 34-48) in the Short Articles venue of dwcummings.com.

16. Syllable Structure. This field is for teachers who work with the notion of closed vs. open syllables or with word stress – as when teaching meter and rhythm in poetry. Closed syllables, ending with a consonant sound, are tagged **C**; open syllables, ending with a vowel sound, are tagged **O**. Additionally, unstressed syllables are tagged **u**; stressed syllables are tagged **s**. The vowel in each syllable is tagged (i) **t** if it is tense – that is, in general, orthographically long, (ii) **l** if it is lax, or orthographically short, or **r** if it is reduced to schwa or is spelled with a syllabic consonant. Thus, the tagging for the two-syllable word *alone* is **OurCst**, which means that the

first syllable, [u4], is open, unstressed, with a reduced vowel, while the second syllable, [lo2n] is closed, stressed, with a tense, or long, vowel. The tagging of *sequence*, **OstCul**, means that the first syllable is open, stressed, with a tense (or long) vowel, while the second syllable is closed, unstressed, with a lax (or short) vowel.

Primary and secondary stress are not distinguished here, both being represented with a simple **s**. Several words have more than one stress pattern, depending usually on the part of speech they are filling – for instance, the verb *convict* with stress on the second syllable vs. the noun *convict* with stress on both syllables. Also, to avoid getting false hits, it's a good idea when searching this field to indicate in the SYLLABLES field the length of the words in which you are interested.

Unstressed short <i> I treat as lax though it could legitimately be treated as reduced. For one thing, it often occurs in open syllables like schwa and unlike stressed lax vowels.

Some Uses of the SYLLABLE STRUCTURE Field:

If your primary class is working with long and short vowels:

Filter the Syllables field on **1**, the Characters field on **3**, the Rank field on **A**, and the Syllable Structure field on **Csl** – that is, closed, stressed, with a lax, or short, vowel – to return more than a hundred three-letter monosyllables with short vowels: *act, cat, end, jog, six*, etc.

Filter the Word field on **ends with e**, the Syllables field on **1**, the Rank field on **A**, and the Syllable Structure field on **Cst** – that is, closed, stressed, with a tense, or long, vowel – to return 200 words in which silent final <e> is marking a long vowel: *age, base, grave, price*, etc.

17 and 18. Prefixes and Suffixes. The Prefixes and Suffixes fields provide a very large sample of the prefixes and suffixes that can be affixed to each word in CommonWords following the normal procedures of combination – such as simple addition, final <e> deletion, twinning final consonants, <y> to <i> replacements, and occasional vowel deletions to avoid unwanted double vowels – as with the <a> deletion when *-an1* is affixed to *utopia*. These fields offer just a sample, with no attempt at exhaustivity. Actually, exhaustivity would not be possible, since prefixes and suffixes constitute a

huge source of potential new words, only some of which have been put to use and recorded in dictionaries – and many more of which have been put to use as nonce words but never recorded in dictionaries. Combined, the lists in Prefixes and Suffixes bring the total number of words in CommonWords to well over 100,000 – some of which, of course, are not all that common. In fact, some of the combinations returned are so uncommon that it takes a pretty big unabridged dictionary to track them down.

I have tried to exclude affixations that produce archaic, obsolete, or dialectal words. Several of the affixes are numbered to discriminate homographs. To find to which prefix or suffix each number refers, see the CommonPrefixes and CommonSuffixes tables.

Prefixes. Prefixes are listed in three places: (i) the Explication field lists any prefixes contained within the listed word itself, tagged with a leading left square bracket and a following + sign; (ii) the CommonPrefixes table, and (iii) the Prefixes, which field lists prefixes that can be added to the listed word.

Actually, the question of what a prefix is remains quite undecided. Dictionaries do not agree on the distinction between prefixes and bases, especially those usually bound bases often called "combining forms." For instance, although the *AHD* uses *electr+* as an example of a combining form (at "combining form"), in the main word list it is labeled "prefix," as are all other combining forms. On the other hand, the *RHUD* and *W3* both distinguish carefully between affixes and combining forms. At a different extreme in *Prefixes: and Other Word-Initial Elements of English* (Old Lyme, CT: Verbatim, 1998) Laurence Urdang and Alexander Humez collapse the distinction completely, speaking only of "word-initial elements" in their list of nearly 3,000 forms.

Elements signifying numerical values can illustrate the indecision: In *W3 bi-* "two" is labeled a prefix, but *tri-* "three" is a combining form. *RHUD* labels both as combining forms; *AHD* labels both as prefixes. I treat all numerical elements as combining forms – that is, bases, usually bound – and restrict prefixes essentially to prepositions (*in2-* "in", *ad-* "to, towards"), negatives (*in1-*, *non-*, *un1-*), adverbs (*se-*, *per1-*), and a few derivationals, (*en1-* and *be-*). For more on this, see the Prefixes table and the "Introduction to the Prefixes Table" in the "Introduction to the Lexis Database" at dwcummings.com.

Ellipses, Parentheses, Braces, and Diagonals. In the Prefixes field ellipses represent the word, separating the prefix from any optional or required suffixes. A suffix is optional if the stem formed by the prefix-plus-word is free; a suffix is required if the stem formed by the prefix-plus-word is not free and requires the addition of a suffix to make a recognized word. One or more suffixes enclosed in parentheses are optional and all create recognized words. For example, at *addict*, “non...(ing1 ive ed1)” represents *nonaddict*, *nonaddicting*, *nonaddictive*, *nonaddicted*. However, a single required suffix is not enclosed in parentheses – for instance, at *audit* “un1...ed1” represents only *unaudited*, since we have no recognized word *unaudit. Two or more suffixes, any one of which are required to make a recognized word, are enclosed in curly braces. Thus, at *admire* “un1...{able ing1}” represents *unadmirable* and *unadmiring* since we do not have the word *unadmire.

Each of the prefixes enclosed in square brackets can take the suffixes following the ellipses. Thus, at *condense* “[inter ultra]...er01” represents *intercondenser* and *ultracondenser*. Because of space limitations, in a very few cases – at *verse*, for instance – prefixes enclosed in diagonals – as “/con in2 ob re tra un1/” – do not have their acceptable suffixes listed. A prefix enclosed in parentheses and immediately preceding another prefix can be added to the word formed with the second prefix – thus, at *vow* “(dis) a3...(al2)” represents *avow*, *avowal*, *disavow*, *disavowal*.

Suffixes. Suffixes are listed in four places: (i) the Explication field, which lists any suffixes contained within the listed word itself, tagged with a following right square bracket; (ii) the CommonSuffixes table, (iii) the Suffixes field, which lists suffixes that can be added to the listed word, and (iv) in the Prefixes field some suffixes are listed that occur in conjunction with certain prefixes.

In the Suffixes field, suffixes in parentheses can be added only after the immediately preceding suffix has been added. For instance, the word *flesh* can take the comparative suffix -er]02 “more” only after it has taken the adjective suffix -y]1: *fleshier* but not *flesher *”more flesh” – thus, at *flesh*, y1 (er02). I’ve tried regularly to mark with a double right parenthesis the ends of strings containing embedded suffixes. For instance, with the regular <y> to <i> replacements, the word *irony* can take the following suffixes: ic1 (al1) (ly1) ness)) ist1 – in the words *ironic*, *ironical*, *ironically*, *ironicness*, *ironist*. (N.B. I’ve tried to be consistent and clear in showing the

embeddings, but I'm sure I have not always succeeded.) .

The main concern here is with derivational suffixes, not inflections. Although regular nouns, verbs, adverbs, and adjectives obviously can add the normal inflectional suffixes, inflectional suffixes are not normally listed in the Suffixes field. I list only those inflections that must be added before adding subsequent derivational suffixes – so at *haunt* “ed1 (ness) er01 ing1 (ly1)” equals *haunt, haunted, hauntedness, haunter, haunting, hauntingly*.

In some cases different suffixes can be affixed to two different senses of a homographic stem. For instance, at the word *camp* the suffixes *-aign* and *-er01* suffixes can only be added to *camp1* with the sense “field, temporary dwelling” while *-y1* can only be added to *camp2* with the sense “humorous banality”. I've tried neither to include nor exclude all suffixes for all the homographic senses.

Several words can take many different suffixes, so in order to make searching the strings of suffixes easier, I've tried consistently to alphabetize the lists.

Some Uses of the SUFFIXES Field: Say you are working with the suffix *-ness*, which turns adjectives into abstract nouns – as in *sad* and *sadness*: Filtering the Explication field on **ness**, returns the 37 CommonWords that contain the suffix *-ness*: *bitterness, goodness, etc.* But if you want more such words, filter the Suffixes field on **ness**, which returns 2,135 words, some of which have quite rare and even odd-sounding *ness* derivatives, but many of which can be useful for teaching vocabulary and word-building: *attentiveness, awkwardness, childishness, etc.*

19. Parts of Speech. This field lists the parts of speech that a word can play. It uses the following tags:

ry = Regular adjectives – that is, those that can take the comparative and superlative inflectional suffixes *-er, -est* – as in *dark, darker, darkest* (though in many, or all, cases the comparative and superlative can also be shown periphrastically with *more* and *most*).

nj = Nonregular adjectives, which includes (i) those that show comparative and superlative only periphrastically, as in *admirable, more admirable, most admirable*; (ii) those that have comparative and superlative forms with bases different from the positive form, as in *good, better,*

best; and (iii) those that only rarely or never have comparative or superlative forms – for instance, ordinals like *eighteenth*; possessive adjectives like *her, his, your, my, our*; and certain absolutes like *every, subsequent, prior*.

rb and nb = Regular and nonregular adverbs, similar to the above distinction between regular and nonregular adjectives

rs = Regular substantives – that is, nouns or noun equivalents – that can form plurals with -s or -es – like *cat/cats* or *kiss/kisses*. I use <s>, for *substantive*, to represent nouns and noun equivalents – like the past participle *forbidden*, as in “The forbidden is always tempting.”

ns = Nonregular substantives – that is, those that can form plurals in other ways, including sets like *goose/geese* and *woman/women*. *nova/novas/novae*. Several nouns are both regular and nonregular – for instance, *nova*, which has the regular plural *novas* and the more technical *novae*. Also nouns that have the same form for singular and plural are tagged “ns”: *fish, deer*.

rv = Regular verbs—that is, verbs that form the past tense with -ed, like *dress/dressed*

nv = Nonregular verbs – that is, so-called “strong” verbs like *swim/swam* and verbs that have the same form for present and past, like *put*

c = Conjunctions.

p = Pronouns.

e = Prepositions.

a = Articles

tl = Past participles, both regular and nonregular

tn = Present participles

in = Interjections

With several words there is not a perfect match between the analyses in the Sound-to-spelling Correspondences and Spelling-to-Sound Correspondences fields and the parts of speech in the Parts of Speech field. For instance, in the Sound-to-spelling Correspondences and the Spelling-to-Sound Correspondences fields the word *alternate* is analyzed phonetically with a long <a> in the final syllable, which is its pronunciation as a verb. But when *alternate* is used as a noun or adjective, that vowel is de-stressed to a short <i>. Nevertheless, in the Parts of Speech field *alternate* is tagged as verb, noun, and adjective. One way of thinking about it is that in the phonetics fields we have to settle on one pronunciation, but in the Parts of Speech field we can take an inclusive view, including heterophonic uses of the written word.

20. Sources. The Sources field gives the lineage of each word. A lineage is the language or languages from or through which a word came into English. The immediate source is the last item in the lineage, so complex lineages, which are presented chronologically left-to-right, are most easily read in reverse. Thus the lineage **Greek > Latin > French** means “English got the word from French, which got it from Latin, which got it from Greek”. Lineages can get quite lengthy – for instance, that for *sugar* is **Sanskrit > Prakrit > Persian > Arabic > Italian > Latin > French**. (Sanskrit and Prakrit were ancient languages of India.)

CommonWords lineages often simplify the more detailed treatment given in dictionaries. For instance, the etymology given in the *American Heritage Dictionary* for the word *meddle* would lead to the lineage “Latin > Vulgar Latin > Old French > Anglo-Norman”, which is simplified in CommonWords to **Latin > French**. (Vulgar Latin was the nonliterary, common speech of the Romans. Old French was French as it was spoken from the 9th to the 16th century. Anglo-Norman was the dialect of Old French spoken by the Normans of Normandy who conquered England in 1066.) Also, CommonWords lineages do not distinguish between chronological periods of a language – for instance, scholars distinguish five ages of Latin: Old Latin (9th century B.C. to 3rd century B.C.), Latin (3rd century B.C. to 2nd century A.D.), Late Latin (3rd century to 7th century), Medieval Latin (8th century to 16th century), and New Latin (from 16th century to the present). In the CommonWords lineages all five ages are collapsed into one, tagged simply **Latin**. The one exception is Old English, which is distinguished from later English. **Norse** refers to any one of the three Scandinavian languages – Swedish, Norwegian, Danish, old and modern. The lineages do not distinguish between High and Low German. The lineages do not always show the etymology of affixes. For instance, *atonement* is tagged **Old English** for the *at* and *one* that form the word, *atone*. But the suffix *-ment*], which is from Latin via French, is not included in the lineage.

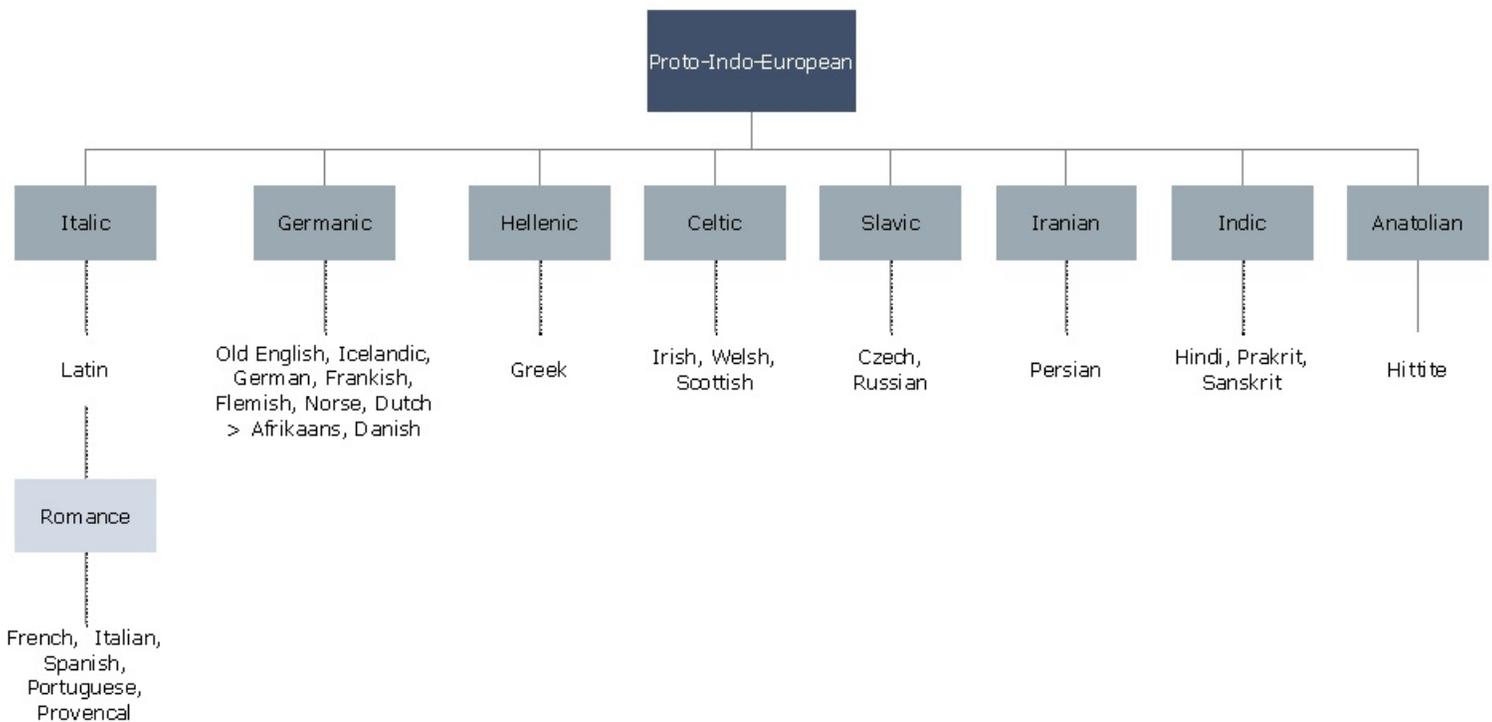
“**OOO**” means “of obscure origin”. A question mark in a lineage usually means “probably” – sometimes “maybe”. Words tagged **Imitative** were usually formed in English, though sometimes it is not clear exactly what is being imitated. Words that come from proper names are tagged **Eponym**. Those few from trademarks are tagged **Trademark**. Those tagged with an exclamation point have etymologies that are surprising or otherwise interesting.

English is one of several languages in the Indo-European super-family, which includes languages in the Slavic, Germanic, Celtic, Italic, Hellenic, Anatolian, and Indic sub-families, and some others not represented in CommonWords. (See the chart below, in which languages appear unboxed, the Romance sub-family in a very light gray box, families in darker grey, the Indo-European super-family in darkest gray.) Proto-Indo-European, the mother tongue of the Indo-European super-family, is thought to have been spoken around 5000 B.C. in the area north of the Black and Caspian Seas. Over the millennia it spread east to India and central Asia, west to modern Greece, Italy, Spain, south to Iran, Pakistan and Afghanistan, and north to Germany, Britain, and Scandinavia.

The non-Indo-European Semitic languages descend from the separate super-family, Afro-Asiatic. Semitic languages represented in CommonWords are Canaanite, Akkadian, Arabic, and Hebrew, out of which Yiddish developed.

The tag **Amerindian** includes a number of non-Indo-European languages from North, Central, and South America. Tamil is a member of the Dravidian language family, spoken in southern India. Sami includes any of the Finnic languages spoken by the Lapps.

The Indo-European languages, their families, and a subfamily that are represented in CommonWords:



Some Uses of the SOURCES Field. Suppose your class is studying American Indians, and you would like to have them work with a list of words that English has adopted from Native American languages. Filter the Sources field on **Amerindian**. The search returns 61 English words derived from American Indian languages: *Alabama*, *barbecue*, *canoe*, *caucus*, *moose*, *Nebraska*, etc. An opening discussion question might be “Why do you think there are so many Amerindian placenames – cities, states, etc.?”

Filtering the Sources field on **!** returns over 340 words with interesting etymologies for further study. Useful dictionaries for such study include the following: Particularly useful because of its appendix of Indo-European roots, the *American Heritage Dictionary* (5th edn., J.P. Pickett et al, eds. [Boston: Houghton Mifflin, 2011]). Fuller treatments (though of fewer words) are available in the *Barnhart Dictionary of Etymology* (R. K. Barnhart, ed. [NY: H. W. Wilson, 1988], which also includes references to the Indo-European roots, including some not mentioned in the *American Heritage*.

Eric Partridge's *Origins: A Short Etymological Dictionary of Modern English* (NY: Macmillan, 1958) is perhaps my favorite: It clusters words containing

the same element; Partridge is willing to make some inspired guesses about words of uncertain origin; and he provides useful information on Latin and Greek morphology for those of us with not much by way of a Classical education. Ernest Weekley in his *An Etymological Dictionary of Modern English* (NY: Dover, 1967), though more concise, is also willing to offer guesses in uncertain cases. Also useful is Joseph Shipley's *Dictionary of Word Origins* (Totowa, NJ: Littlefield, Adams, 1967). Much more detail on fewer words can be found in books like Charles Funk's *Thereby Hangs a Tale* (NY: Harper & Row, 1955).

And the best source of information on the changing spelling and senses of words once they came into English is the *Oxford English Dictionary* now available on CD-ROM and on-line

Other Data Tables:

The CORRESPONDENCES: SOUND-TO-SPELLING table contains six fields: (i) Sound-to-spelling, which lists the sound-to-spelling correspondences; (ii) Examples, which gives an example word containing each correspondence; (iii) Instances, which gives the number of words in CommonWords that contain at least one instance of the correspondence; (iv) AES, which cross-references to sections of my *American English Spelling* dealing with the correspondences; (v) Percentages for this Sound, which gives the percentage that this correspondence constitutes for this sound, and (vi) Sort, which is a number used to set the sort order for the table. The Sort field can also be used to select subsets of sounds in the correspondences, which are listed in the following order, with the following beginning and ending Sort numbers:

Short (Lax) Vowels: 1-23, 39-44
Long (Tense) Vowels: 49-93
Tense but Not long Vowels: 26-38*
Diphthongs: 94-97

Schwa: 98-117
[r]-Colored Vowels: 117.1-184
Vowels with initial [y]: 185-194
Consonants: 195-302
Silent Letters: 303-313
Punctuation: 314-317

* For an explanation of Tense but Not Long Vowels, see “Notes on the Vowel Analysis in CommonWords” in the Short Articles venue in dwcummings.com.

The **CORRESPONDENCES: SPELLING-TO-SOUND** table contains five fields: (i) Spelling-to-Sound, which lists the spelling-to-sound correspondences; (ii) Examples, which gives an example word for each correspondence; (iii) Instances, which gives the number of words in CommonWords that contain at least one instance of the correspondence; (iv) AES, which cross-references to sections of AES dealing with the correspondences; and (v) Percentages for this Spelling, which gives the percentage that this correspondence constitutes for this spelling. For more information, see the **SOUNDS COUNT** and **SPELLINGS COUNT** tables.

The **SOUNDS COUNT** table lists the frequency of occurrence of each of the 68 sounds found in the 8,591 words in CommonWords. It contains four fields: (i) Sound; (ii) Sample Word; (iii) Instances, which gives the number of times that sound occurs in CommonWords; and (iv) Rank, which shows each sound’s ranking in number of instances.

The **LETTERS COUNT** table lists the frequency of occurrence of each of the 26 English letters. Notice that unlike the Spellings Count table this one deals only with the 26 individual letters.

SPELLINGS COUNT. This data table lists the frequency of the 154 different spellings from the **CORRESPONDENCES: SPELLING-TO-SOUND** table. The large number of spellings arises primarily from three historical processes: First, the simplification of earlier clusters of vowel and consonant sounds in which

the vowel and consonant letters are still written, as in the <mb> spelling at the end of words like *bomb* or the <io> spelling of schwa in words like *nation*. Second, the formation of doublets via twinning, assimilation, and simple addition. And third, the complicating effects of [r] on preceding vowels. This table is sortable on the Spelling and Instances fields.

COMMONBASES. Since the senses of free bases – that is, simple words – should be fairly evident, this table includes only bound bases. Sometimes the Comment field offers a straightforward gloss of the sense carried by the base – as when *aut1*, the base of *autistic*, is glossed as “Self.” Sometimes the connection is only slightly more complicated – as when *com2*, as in *comet*, is glossed as “Hair, long-haired,” due a comet’s shape as it soars through space. But most of the time the senses of bound bases can be quite elusive. In such cases the Comment field usually gives a chronological list of earlier senses carried by words containing that or a closely related base – as when *colon1*, as in *colony*, is glossed as “Cultivate, farmer, estate, settlement.”

In nearly all cases the Examples field lists all of the words in CommonWords that contain the base in question. Many of the words listed in the Examples field are quite technical because bound bases by their very nature tend to occur in longer, rarer words.

The figure in the Instances field is based on the 129,000+ words in the Lexis database. There are so few instances of each base in CommonWords that the figures drawn from that small list would not be very meaningful. The Instances figures can indicate bases that are productive enough to be worth some extra attention. To find those instance words, you can filter on the desired base in the Explication field in the Lexis database. But for those bases with silent final <e> or final <y>, you’ll have to do two searches and add their results together: the first search showing the silent final <e> or final <y>, the second showing the <e> and <y> being deleted. For instance, to search Lexis for words with the base *venge*, search one would search on Explication contains *venge*, the second on *veng/e*.

Some Uses of the CommonBases table: The string of example words can suggest sometimes subtle links among cognate words such as *especially*, *special*, *species*, *specific*, *specimen*, all of which contain the base *speci* “To observe, look at, a seeing, form.” In such cases it can be useful to ask students how they think the modern senses of those words could have

emerged out of those earlier senses. When bases have the same earliest senses, they are historically related. Discussing these relationships can elicit some good thinking from the students – while impressing on them that words and their bases are not just random lists. Some of the more puzzling ones – as, for instance, what in the world does the modern word *larva* have to do with masks and specters? – can lead to some interesting sleuthing into words' etymologies.

The earlier senses can impress on students the importance of prefixes, especially the negatives like [*in-*, [*un-*, [*dis-*, [*de-*: Often a base's string of past senses contains a sense that is just the opposite of the modern word with its negative prefix. For instance, the senses listed for the base *dain* in *disdain* include "worthy, deserving," which are exactly reversed by the negative prefix [*dis-*. The same is true with, say, the base *spair* is *despair*, whose string of earlier senses is "To thrive, prosper, to hope, hope," which is negated by the prefix [*de-*.

Some bases appear in modern words that have widely different senses – for instance, the base *mon1* in *demonstrate*, *monster*, *monument*, *summon*. A good thinking activity could be to ask the students how can those words be linked so as to account for their having the same base? The senses listed in the Comment field can help by suggesting relationships among the senses.

Since the senses listed in the Comment field are in chronological order, bases that share common earlier senses can be taken to be related – sometimes subtly so. By filtering the Comment field on sense words you can sometimes find larger groups of related bases, from which you can broaden the discussion of relatives and their shared senses.

COMMONPREFIXES and COMMON SUFFIXES. The Comment fields present information about the sense and function of each prefix and suffix and some special filtering instructions for assimilated prefixes. To find all the words in CommonWords with a given prefix or suffix, filter the Explication field: To find the 19 words containing the prefix [*ab1-* filter Explication to **contains [ab1+** and to find the 74 words containing the suffix *-ment*] filter Explication to **contains +ment]**. The figures in the Instances fields are again from the larger Lexis database.



N.B. In something as complex and labor-intensive as CommonWords there are seemingly endless opportunities for error. The Prefixes and Suffixes fields seem to me to be particularly vulnerable. Should you find any errors, I would appreciate your notifying me at donwcummings@charter.net.

D. W. Cummings
Emeritus Professor of English
Central Washington University
Ellensburg, WA