Lexbuild training materials:

First Things First:

I. Links sent by Chris on procedures for login, etc

II. Allen's manual, AKA The Specialist Lexicon Technical Report, by Allen C. Browne, Alexa T. McCray and Suresh Srinivasan, June 2000. On training documents website as 1st bullet point under "I. Lexicon". https://lsg3.nlm.nih.gov/LexSysGroup/Projects/lexicon/2018/release/LEX/DOCS/techrpt.pdf

III. List of dictionaries & grammars

A. Dictionaries

1. Merriam-webster unabridged (http://unabridged.merriam-webster.com/unabridged/ John Nguyen will set up the acct)
2. Dorland's Illustrated Medical dictionary
3. Stedman's Medical Dictionary
4. Webster's Medical Desk Dictionary
5. Jablonsky's Dictionary of Medical Acronyms and Abbreviations
6. Longman's Dictionary of Contemporary English
7. Oxford Advanced Learner's Dictionary
8. Collins Cobuild English Language Dictionary

B. Grammars

1. Quirk et al. Grammar of Contemporary English
2. Huddleston & Pullum The Cambridge Grammar of the English Language

IV. Websites used

a) The UMLS Metathesaurus
b) Google Scholar [better than regular Google, b/c it's mostly vetted & edited scholarly articles]
c) USPTO for trademark info
d) WIPO for worldwide trademark info

e) Wikipedia [for general background info in order to decide whether to make a record; reminders on Greek alphabet]

V. Essie: the in-house search engine developed & maintained by Russell Loane & Nick Ide

A. jar file attached to email of 2/23/15 sent by Russell, will launch starter windows

B. server 130.14.81.58

1. port 451: Historic Essie
2. port 452: Clinical Essie

C. most-used Essie tools

1. Hit Patterns (drill-down on results to check for MWEs, acronym expansions)
2. Match Patterns (for acronym expansion checks, examples)

How to Make Lexicon Records with Lexbuild

I. Basics applying to all POSs

A. Look to edited usage (especially Medline, Google Scholar) & aim to reflect that.

B. Try to reflect what the accepted (biomedical & general unabridged) dictionaries show, as well as the usage you see. Most dictionaries will have far less than our Lexicon, but guidance can be found by looking up the head of a longer string, adjectives for adverbs, etc.

C. Check for spelling variants (via Essie), which may or may not be common. When in doubt, include spelling variants:

1. British & other non-American variants (heme/haem, cele/coele, color/colour –ize/-ise), whether independent words or morphemes.
2. Spacing, hyphenation variants, e.g. *presurgical, pre surgical, pre-surgical*. Check morpheme boundaries for these.
3. The Metathesaurus sometimes has uncommon spelling variants, but make sure only to include English ones, as the Met gives many languages together in its search results.
4. Spelling variants must be **strictly homophonous** with the base and each other. Some variants of entries from paper or online dictionaries are not homophonous, just very similar, so do not automatically include these as spelling variants. For example, adjectives ending in *–ical* vs *–ic*, *-geneic* vs *–genetic*. But also note that homophony can occur across morpheme boundaries: *Job syndrome* has the (homophonous) spelling variant *Job's syndrome*, for example. *Telecommunication technology* and *telecommunications technology* are not strictly homophonous, so each has its own LB record, but *telecommunication system* and *telecommunications system* are homophonous, and share a single LB record. Homophony need not obtain across every standard variety of English; just make your choice defensible.

D. Include examples in the annotation field, whenever that makes your LB choices clearer, or if the term seems opaque without illistration.

E. What is an error, vs alternate usage/form? This can be a tough call, but Allen Browne's rule of thumb was, "Would the writer take it back, if the expression were questioned?" Frequency is not always a good meteric; "dairy" & "diary" are frequent typos of each other. But any writer mistakenly writing one form for the other, would readily take it back. In rare cases, spelling variants contain demonstrable errors, which are included because they are so common that authors arguably see this usage as correct. We include these spelling variants because when in doubt, we reflect edited usage. Example:

{base=Alzheimers disease
spelling_variant=Alzheimer's disease
spelling_variant=Alzheimers' disease
entry=E0000237
  cat=noun
  variants=uncount
}

Technically, only "Alzheimer's disease" shows the correct form of Dr Alzheimer's name. But because Metathesaurus source vocabularies include punctuation-stripped forms, so does Lexbuild ("Alzheimers disease"). Erroneous "Alzheimers' disease" is called a spelling variant here, because it does find its way into (edited) publications, and Lexbuild's utility as a search tool is enhanced by a looser approach to name spelling in such LexMultiWords (LMWs).

II. How to make a noun record: for basics, see Allen's manual.

A. The count/uncount distinction = major focus. Many nouns will be both count & uncount. Some uncount biomedical usage can be unexpected or odd, to those of us outside that field. In this example, uncount usage is all Essie shows; Google Scholar did yield some count usage:

```
{base=radical abdominal trachelectomy
entry=E0766113
        cat=noun
        variants=reg
        variants=uncount
}
```

i) Just because a count noun is rarely seen in singular (or plural) does not mean it isn't still a count noun. Many LMWs ending in "cells" would be of this type.

ii) Body part nouns in Latin (NA, nomina anatomica) are often found only in singular form, in English, so these are coded as having "sing" variants:

```
{base=aorta descendens
entry=E0438912
        cat=noun
        variants=sing
}
```

B. Nouns with initial capitalization can be:

1. Genus terms (or larger constructs like orders, phyla). These, and genus-species(-subspecies) multiwords (*Escherichia coli*, *Homo sapiens sapiens*) always take invariant & uncount as their variants. Enough usage goes in that direction, that Allen Browne made the call in the mid-1990s, that they should be consistent in their variants.
2. Trademarked medicines, medical devices, organizations (some universities are; some aren't)
3. Proper nouns, as well as organization names that we have **not** called proper nouns (e.g. *National Cancer Institute*), though future policy changes could alter that.
4. Names of questionnaires or similar instruments. Usage of capitalization may vary, but include both unless one or the other appears so infrequently as to seem erroneous. Example:

```
{base=Remote Associates Test
spelling_variant=remote associates test
entry=E0766111
        cat=noun
        variants=reg
}
```

5. Names of dyes. As with questionnaires, usage may vary, but when in doubt, be inclusive.

C. In Essie, look at the Right Contexts for frequently occurring prepositional complements. Prepositional complement information improves the utility of the Lexicon in NLP, though the lack of complement information is not erroneous, per se. If you note specifically worded complements, include those, but it is generally a good idea to then also include similar but more general complements. Example:

```
{base=stool antigen test
entry=E0766106
        cat=noun
```

variants=reg
compl=pphr(for,np|diagnosis|,pphr(of,np|Helicobacter pylori|))
compl=pphr(for,np|diagnosis|,pphr(of,np))
compl=pphr(for,np)

D. In Essie, look at both Left & Right Contexts for indications that your lookup is an acronym or abbreviation: parenthetical or appositional constructions. For more on acronym and abbreviation record building, see section XII below.

1. For acronyms/abbreviations, see if there's already a LB record for that base.

(1) If so, just add the expansion to the exisiting list of expansions, plus the variants for your addition (if they aren't already there). Count acronym/abbreviation Ns will have "metareg" as the variant, & like other Ns, often both count & uncount apply.

(2) If not, look up other expansions for that base, via Essie & Jablonsky (Google Scholar will likely take you too far into other specialty fields)

2. Acronym vs abbreviation can be a tough call; don't dwell too long on that.

a) Acronyms (e.g. *ELISA*) are normally comprised of single letters representing words or morphemes of a longer string, the expansion. They are usually uppercase, though not always, & sometimes the first 2 letters of 1 or more words of the expansion will be in the acronym.

b) Abbreviations are larger chunks of the expansion (e.g. *blvd /boulevard*). Sometimes these end in a period, sometimes not; spelling variants with both options also exist. Single letter bases are considered abbreviations (e.g. *H /hydrogen*). Sometimes, abbreviation expansions don't fit well because they are based on Latin or other FL terms (e.g. *Au/gold*).

c) Acronyms & abbreviations can be part of longer multiword terms. When the rightmost item is a word rather than the acronym or abbreviation (e.g. *HIV infection*), choose the variants that fit that last word (i.e. not metareg).

3. Acronym/abbreviation expansions are candidates to be LB bases themselves.

a) Look up each expansion in new Essie windows, to make sure of getting all spelling variants.

b) Certain types of expansions are not LB bases:

(1) Chemical strings that are more like formulae than words, containing numbers with commas, indexical (not abbreviation) single letters, etc. You will also encounter these in contexts other than as acronym expansions, but they do not merit LB records, in any case.

(a) Example:

```
{base=L-NMMA
spelling_variant=l-NMMA
```

```
entry=E0696977
        cat=noun
        variants=uncount
        acronym_of=NG-monomethyl-L-arginine
}
```

(b) Another example:

```
{base=NT5E
entry=E0766207
        cat=noun
        variants=uncount
        acronym_of=ecto-5'-nucleotidase
}
```

(2) Expansions that are not a single POS, especially prepositional phrases (that are not names of institutions like National Institutes of Health). Example:

```
{base=QOL
spelling_variant=QoL
entry=E0419257
        cat=noun
        variants=uncount
        variants=sing
        acronym_of=quality of life
}
```

(3) Single POSs that are lists of characteristics, drugs in cancer regimens, and the like. Examples: T&T has expansions time and temperatute, and touch and tone, neither of whilch have their own LB base records:

```
{base=T&T
spelling_variant=T & T
entry=E0203063
        cat=noun
        variants=metareg
        abbreviation_of=time and temperature
        abbreviation_of=touch and tone
}
```

Also, CISCA has the expansion cisplatin, cyclophosphamide, and Adriamycin. The expansion is not a single concept, & so has no record of its own:

```
{base=CISCA
entry=E0521756
        cat=noun
        variants=uncount
        acronym_of=cisplatin, cyclophosphamide, and Adriamycin
annotation=part abbrev, part acro
}
```

(4) Studies are considered too ephemeral to have Lexbuild records, though they can be listed among the expansions of an acronym or abbreviation, if usage warrants that.

(5) Expansions that never occur, anywhere, need not be given their own records. Exceptions may be made for truly word-like (or term-like) expansions, especially those with which are part of a set of terms with occurrences for other members of the set.

(6) Otherwise, do strongly consider expansions as Lexbuild base candidates. Our policy has been to include LB records for expansions except in the situations given above. Thus, a rarely occuring string may be given a Lexbuild record. My own standard has been that the number of Google Scholar occurrences should not be less than 7; preferably more than 10. Policy on this is subject to change.

E. Nouns that are heads of chemical string formulae may be given LB records. The following example is based on usage like PMID 15934927: *6-oxocamphor hydrolase cleaves nonenolizable cyclic beta-diketones*

```
{base=oxocamphor hydrolase
entry=E0523179
        cat=noun
        variants=uncount
}
```

F. If your noun has the morphology of a nominalization (ending in *–ness, -tion, -ity* etc.), check to see if there is or should be an associated adjective or verb record. For adjectives ending in *–ed*, this can be a tough call. Check dictionaries, then see if other verbal forms (*-s, -ing*) are used.

1. Example of a nominalization from dictionary adjective entry:

```
{base=globularness
entry=E0552675
        cat=noun
        variants=uncount
        nominalization_of=globular|adj|E0029894
}
```

2. Example of typical nominalization:

```
{base=cauterisation
spelling_variant=cauterization
entry=E0015659
        cat=noun
        variants=reg
        variants=uncount
        compl=pphr(of,np)
        compl=pphr(by,np)
        nominalization_of=cauterise|verb|E0015660
}
```

3. Nominalization of more than 1 verb:

```
{base=respiration
entry=E0053022
        cat=noun
        variants=reg
        variants=uncount
        compl=pphr(of,np)
        nominalization_of=respire|verb|E0053037
        nominalization_of=respirate|verb|E0312498
}
```

4. Nominalization of more than 1 adjective:

```
{base=pennation
entry=E0604752
        cat=noun
        variants=uncount
        compl=pphr(of,np)
        nominalization_of=pennated|adj|E0593967
        nominalization_of=pennate|adj|E0046273
}
```

III. How to make a verb record: for basics, see Allen's manual sections 1.3, 4.1, 5.1, 5.2, 5.3

A. For complements, Longman's dictionary has complement info, but be sure to check Essie usage & expand the complement list as appropriate.

B. Not all possible complements can be coded. This may be an area for future Lexbuild expansion.

C. The basic complement codes (for transitive, instransitive, ditransitive) need to be correct. Beyond that, the more complex complement codes enable you to refect usage, but fall into the category of "nice extras" rather than information that parsers/POS taggers cannot do without. We do strive to make the Lexicon as useful as possible, though, so including as many extras as practical, is one of our goals.

D. Typical verb record:

```
{base=decatenate
entry=E0425553
        cat=verb
        variants=reg
        tran=np
        nominalization=decatenation|noun|E0425552
}
```

E. Verb record with more complements:

```
{base=treat
entry=E0061964
        cat=verb
        variants=reg
        intran
        tran=np
        tran=pphr(with,np)
        tran=pphr(of,np)
        ditran=np,pphr(to,np)
        ditran=np,pphr(with,np)
        ditran=np,pphr(for,np)
        cplxtran=np,advbl
        nominalization=treatment|noun|E0061968
}
```

Note that we call prepositional phrase complements of verbs transitive (or 2nd arguments of ditransitive) complements.

F. Another verb record with many complements:

```
{base=signal
entry=E0055889
        cat=verb
        variants=reg
        variants=regd
        tran=np
        tran=pphr(to,np)
        tran=pphr(for,np)
        tran=fincomp(t):subj
        tran=infcomp:arbc
        ditran=np,fincomp(t):subj
        ditran=pphr(to,np),fincomp(t):subj;datmvt
        ditran=np,pphr(to,np);datmvt
        cplxtran=pphr(to,np),infcomp:objc
        nominalization=signaling|noun|E0234006
}
```

G. Verb record for existing nominalization. "*Preevaluation*" came up in a Daily Review, as an existing record with no nominalization; potentially linked verbs & adjectives should be checked whenever nouns have the morphology of nominalizations. Here is the newly submitted record for *preevaluate*. An EUI will replace the placeholder EUI E000000 after the record is approved :

```
{base=preevaluate
spelling_variant=pre-evaluate
entry=E0000000
        cat=verb
        variants=reg
        tran=np
```

```
        ditran=np,pphr(for,np)
        nominalization=preevaluation|noun|E0611457
annotation=PMID 9081135: platelet aggregation was preevaluated and adjusted
with antiplatelet agents.
annotation=PMID 8222452: Patients are preevaluated for their rehabilitation
potential.
annotation=PMID 16009191: the seminal viral load should be preevaluated
before enrolling an HIV-serodiscordant couple
annotation=PMID 20424600: cells that have been comprehensively pre-evaluated
for genotoxic potential
}
```

IV. How to make an adjective record: for basics, see Allen's manual sections 1.5, 4.3, 5.5, 6.

A. Most new adjective records will be stative (i.e. not under the speaker's control; test S: "Don't be so [nonstative adj]!")

B. Adjectives ending in –ed, -en or –ing may be verb forms used adjectivally or may be bona fide adjectives. Follow published dictionaries whenever possible, and include an annotation about published dictionary status when appropriate.

1. Example of an adjective having a Merriam-Webster's entry:

```
{base=controlled
entry=E0018835
        cat=adj
        variants=inv;periph
        position=pred
        position=attrib(3)
        position=attrib(1)
        stative
        nominalization=controlledness|noun|E0226177
}
```

2. Example of a verb form often used adjectivally: "activated"

```
{base=activate
entry=E0007090
        cat=verb
        variants=reg
        tran=np
        nominalization=activation|noun|E0007100
}
```

3. Add negation, and the result can only be an adjective:

```
{base=unactivated
spelling_variant=un-activated
entry=E0312689
        cat=adj
        variants=inv
        position=attrib(3)
```

```
        position=pred
        stative
}
```

4. Multiword expressions containing adjectives having Lexicon records as adjectives or as verbs may be considered for inclusion in the Lexicon as LMWs, e.g.:

```
{base=mitogen-activated protein
entry=E0764510
        cat=noun
        variants=reg
        variants=uncount
annotation=Google Scholar: Erk2 binds to mitogen-activated proteins
(Morishima-Kawashima and Kosik, 1996).
annotation=Google Scholar: modulator of many mitogen-activated proteins such
as Erks or p38 kinases.
annotation=Also cf Metathesaurus.
}
```

C. An adjective LMW based on a verb form used adjectivally, can have a nominalization. Take for example:

```
{base=CMV infected
spelling_variant=CMV-infected
entry=E0764109
        cat=adj
        variants=inv
        position=pred
        position=attrib(3)
        stative
        nominalization=CMV infection|noun|E0764110
        acronym_of=cytomegalovirus infected|E0764107
}
```

D. Adjective multiwords exist, e.g.

```
{base=case controlled
spelling_variant=case-controlled
entry=E0632179
        cat=adj
        variants=inv
        position=pred
        position=attrib(3)
        stative
annotation=EXPERIMENT-1
annotation=SPVAR: case controlled
annotation=PMID 12139229: The study was case controlled, matching for age,
skin color, and parity.
```

```
}

{base=near infrared fluorescent
spelling_variant=near-infrared fluorescent
spelling_variant=near-infrared-fluorescent
spelling_variant=near infra-red fluorescent
spelling_variant=near-infra-red fluorescent
entry=E0760689
        cat=adj
        variants=inv
        position=pred
        position=attrib(3)
        stative
        nominalization=near infrared fluorescence|noun|E0760688
}
```

E. Adjectives may be adjectivalizations of nouns, though we do not mark this relationship in Lexicon records:

```
{base=cheilorhinoplastic
spelling_variant=cheilo-rhinoplastic
entry=E0438719
        cat=adj
        variants=inv
        position=attrib(3)
        position=pred
        stative
}

{base=cheilorhinoplasty
spelling_variant=cheilo-rhinoplasty
entry=E0603709
        cat=noun
        variants=uncount
}
```

F. Adjectivalizations of noun LMWs also exist, and as these examples show, whether a given term is a single word or a multiword, is sometimes a matter of spelling variation:

```
{base=crosssectional
spelling_variant=cross sectional
spelling_variant=cross-sectional
entry=E0355253
        cat=adj
        variants=inv
        position=attrib(3)
        position=pred
        stative
}
{base=cross section
```

```
spelling_variant=cross-section
entry=E0019875
        cat=noun
        variants=reg
        variants=uncount
        compl=pphr(of,np)
        nominalization_of=cross-section|verb|E0727331
}

base=immunoelectronmicroscopic
spelling_variant=immunoelectron microscopic
spelling_variant=immuno electron microscopic
spelling_variant=immuno-electronmicroscopic
spelling_variant=immunoelectron-microscopic
spelling_variant=immuno-electron microscopic
spelling_variant=immuno-electron-microscopic
entry=E0317238
        cat=adj
        variants=inv
        position=attrib(3)
        position=pred
        stative
}
{base=immunoelectronmicroscopy
spelling_variant=immuno electronmicroscopy
spelling_variant=immunoelectron microscopy
spelling_variant=immuno electron microscopy
spelling_variant=immuno-electronmicroscopy
spelling_variant=immunoelectron-microscopy
spelling_variant=immuno-electron microscopy
entry=E0330459
        cat=noun
        variants=uncount
}
{base=differential scanning calorimetric
entry=E0760773
        cat=adj
        variants=inv
        position=pred
        position=attrib(3)
        stative
annotation=Adjectivalization of differential scanning calorimetry.
}
```

G. As with nouns, adjective acronym and abbreviation expansions are given Lexicon records, unless they do not meet basic word-like requirements. Examples:

```
{base=BC
entry=E0504614
        cat=adj
        variants=inv
        position=attrib(3)
```

```
        stative
        acronym_of=brachiocephalic|E0332947
        acronym_of=board certified|E0504612
        acronym_of=buccocervical|E0014268
}

{base=DC
entry=E0203109
        cat=adj
        variants=inv
        position=attrib(3)
        position=pred
        stative
        acronym_of=diffuse cortical
        acronym_of=direct-coupled|E0589555
        acronym_of=distocervical|E0023534
annotation=also refers to electric defibrillator using DC discharge
}
```

H. If an adjective is the nominalization of an existing noun record, the EUI will be added automatically after the new record is submitted. If it is the nominalization of a newly made noun record, a later post-process will prompt you to check cross-referencing of these records.

V. How to make an adverb record: for basics, see Allen's manual sections 1.6, 4.4.

A. Most new adverb records will be invariant; those with *–ier* and *–iest* tend to be in the general English vocabulary, which was entered in the Lexicon's early phases.

B. The "intensifier" label is not quite accurate, deriving from common so-called intensifiers such as very and highly. Any adverb that can modify an adjective is an intensfier:

```
{base=weakly
entry=E0300031
        cat=adv
        variants=inv;periph
        modification_type=intensifier
        modification_type=verb_modifier;manner
}
```

C. It is possible to have more than 1 code for modification type, as *weakly* shows, above, and *multilocally* shows, below.

D. The "manner" modification type sometimes indicates that neither of the other 2 types fits, or that some extra meaning sense sometimes extends that adverb beyond strictly locative or temporal meaning:

```
{base=multilocally
spelling_variant=multi-locally
entry=E0622959
        cat=adv
        variants=inv
        modification_type=verb_modifier;manner
        modification_type=verb_modifier;locative
        modification_type=intensifier
}
```

E. Biomedical adjectives often have lexically associated adverbs, which may not be in any published dictionary. Add them unless you are unsure of the correct coding.

F. Adverbial multiwords are sometimes encountered, some with forms similar or identical to adjective multiwords:

```
{base=cross sectionally
spelling_variant=cross-sectionally
entry=E0420107
        cat=adv
        variants=inv
        modification_type=intensifier
        modification_type=verb_modifier;manner
}

{base=twentyfold
spelling_variant=twenty fold
spelling_variant=20-fold
spelling_variant=twenty-fold
entry=E0234971
        cat=adv
        variants=inv
        modification_type=verb_modifier;manner
        modification_type=intensifier
}

{base=twentyfold
spelling_variant=twenty fold
spelling_variant=20-fold
spelling_variant=twenty-fold
entry=E0234970
        cat=adj
        variants=inv
```

```
        position=attrib(3)
        stative
}
```

G. Sentential adverbs can sometimes be found in Google Scholar searches, so always check for those. Similarly, when editing or checking an older adverb record, make sure the coding reflects the usage you see in Essie Hit Patterns contexts and Google Scholar text samples. These resources were unavailable earlier, and coding reflected whatever usage was seen at that time.

VI. How to make a preposition record: for basics, see Allen's manual

A. The coding for prepositions could not be simpler:

```
{base=about
entry=E0006548
        cat=prep
}
```

B. Phrasal prepositions should be entered into the Lexicon whenever found; see Quirk et al. sections 6.4 & 6.5 on "simple and complex prepositions". Example:

```
{base=except for
entry=E0228920
        cat=prep
}
```

VII. More on how to make acronym & abbreviation records: For basics, see the subsection on acronyms and abbreviations in the secion above on nouns, 7d.

A. Most of these will be nouns. For count nouns, the variants slot will be marked as "metareg," which will indicate plurals with *–s* or *–'s*. Just as with regular nouns, more than 1 variants type is allowed; both count (metareg) and uncount variants are commonly found in acronym and abbreviation records. Just reflect the edited usage you see. Note the possible error/omission in variants in the record for *qt*.

```
{base=qt
entry=E0319113
        cat=noun
        variants=metareg
        abbreviation_of=quantity|E0051565
        abbreviation_of=quart|E0051579
```

```
}


{base=MPX
entry=E0516874
        cat=noun
        variants=metareg
        variants=uncount
        acronym_of=1-methyl-3-propylxanthine
        abbreviation_of=mastoparan X|E0516873
        abbreviation_of=myeloperoxidase|E0041628
}
```

B. Not all expansions will have their own records, and thus will not have an EUI given at the end of the acronym_of or abbreviation_of line:

```
{base=bdb
entry=E0423403
        cat=noun
        variants=plur
        acronym_of=elongated dense bodies below the nucleus
}


{base=VDB
spelling_variant=vDB
entry=E0720190
        cat=noun
        variants=metareg
        acronym_of=vertical limb of the diagonal band
        acronym_of=vertical limb of the diagonal band of Broca
annotation=PMID 2207699: Neurons in the medial septum (MS) and vertical limb
of the diagonal band (vDB) undergo degenerative changes
annotation=PMID 18804152: from the HDB and the ventral limb of diagonal band
(VDB) and the adjoining areas.
annotation=PMID 12957499: in AD and the vertical limb of the diagonal band of
Broca (VDB)
annotation=PMID 12093089: The vertical limb of the diagonal band of Broca
(VDB or Ch2) and the nucleus basalis of Meynert (NBM or Ch4)
}
```

C. Not all bases that look and act like acronyms and abbreviations will have expansions:

```
{base=QT
spelling_variant=Q-T
entry=E0764046
        cat=noun
        variants=metareg
```

```
        variants=uncount
annotation=electrocardiogram measurement points; QT interval
annotation=PMID 697472: Change in the Q-T (delta QT) in the electrocardiogram
}
```

D. Overlap in spelling variants presents a complication. We are currently handling that by saying in the annotation field, which other base(s) will have that (or those) particular expanion(s), e.g.:

```
{base=Tx
entry=E0418183
        cat=noun
        variants=uncount
        variants=metareg
        abbreviation_of=treatment|E0061968
        abbreviation_of=therapy|E0060549
        abbreviation_of=thromboxan|E0060821
        abbreviation_of=transplantation|E0061846
        abbreviation_of=tamoxifen|E0059858
        abbreviation_of=thyroidectomy|E0060952
        abbreviation_of=transection|E0061861
        abbreviation_of=transplant|E0061842
annotation=All expansions except 'therapy' also have the abbreviation TX; see
record for TX.
}
```

```
{base=TX
entry=E0418184
        cat=noun
        variants=uncount
        variants=metareg
        acronym_of=tendon xanthoma|E0721900
        abbreviation_of=thromboxan|E0060821
        abbreviation_of=transplantation|E0061846
        abbreviation_of=treatment|E0061968
        abbreviation_of=primary tumor cannot be assessed
        abbreviation_of=tamoxifen|E0059858
        abbreviation_of=thyroidectomy|E0060952
        abbreviation_of=transection|E0061861
        abbreviation_of=transplant|E0061842
        abbreviation_of=Texas|E0005684
annotation=part of TNM System for cancer staging
annotation=These expansions also have the abbreviation Tx (see record for
Tx): treatment, thromboxane, transplantation, tamoxifen, thyroidectomy,
transection, transplant.
}
```

E. Because different meanings of the same base are grouped into a single Lexbuild record (for acronyms and abbreviations as well as full forms), it will often be the case that some variants apply to some expansions but not others. We do not index expansions to indicate which variants (metareg, uncount, invariant, plural) apply to each expansion, though users have on occasion requested something like that. The one concession we make is when an expansion can occur with plural reference, but not plural form. We use the annotation field to say which expansions this plural designation applies to:

```
{base=QM
entry=E0701448
        cat=noun
        variants=plur
        variants=metareg
        variants=uncount
        acronym_of=quantum mechanics|E0701438
        acronym_of=quality management|E0701443
        acronym_of=quinacrine mustard|E0701444
        acronym_of=quinone methide|E0701446
annotation=As 'quantum mechanics,' QM can be either plural or uncount.
}
```

F. Sometimes a noun acronym or abbreviation will have count and uncount use, and the expansion will have only count use:

```
{base=GHRHR
spelling_variant=GHRH-R
entry=E0769188
        cat=noun
        variants=metareg
        variants=uncount
        abbreviation_of=GHRH receptor|E0769187
        abbreviation_of=growth hormone releasing hormone receptor|E0769186
annotation=PMID 23199197: HESX1, LHX4, OTX2 and SOX3 polymorphisms were
screened in 33 PSIS patients, and GH1 and GHRHR in 4 NPS patients.
annotation=PMID 15853821: mRNA for wild-type GHRH-R or SVs of GHRH-R were not
observed
}

{base=growth hormone releasing hormone receptor
spelling_variant=growth hormone-releasing hormone receptor
entry=E0769186
        cat=noun
        variants=reg
}
```

G. For adjective abbreviations & acronyms, the expansion will sometimes be the past participle of the verb, and so will not have an adjective EUI associated with that expansion, e.g. the expansion *transected* in the Lexbuild record for Tx and the expansion *predicted* in the record for the base *pred*:

```
{base=Tx
entry=E0721901
        cat=adj
        variants=inv
        position=pred
        position=attrib(3)
        stative
        abbreviation_of=thyroidectomised|E0222330
        abbreviation_of=transected
annotation=thyroidectomized also has the abbreviation TX; see adj record for
TX.
}

{base=pred
entry=E0604548
        cat=adj
        variants=inv
        position=pred
        position=attrib(3)
        stative
        abbreviation_of=predicted
}
```

H. Adjective and adverb abbreviation/acronym bases can sometimes be identical, e.g. *post-IL2* and others with the prefixes *pre-* and *post-*:

```
{base=post-IL2
spelling_variant=post-IL-2
entry=E0696208
        cat=adj
        variants=inv
        position=pred
        position=attrib(3)
        stative
        abbreviation_of=post-interleukin 2
}


{base=post-IL2
spelling_variant=post-IL-2
entry=E0696209
        cat=adv
        variants=inv
        modification_type=verb_modifier;temporal
        abbreviation_of=post-interleukin 2
}
```

I. Adverb acronyms and abbreviations are coded the same way as non-acronym/abbreviation adverbs, with the addition of expansion information, as the above example of *post-IL2* shows. As with other acronyms and abbreviations, if the expansion has a Lexbuild record associated with it, that base's EUI will be appended to the expansion:

```
{base=LO
spelling_variant=lo
spelling_variant=lo.
entry=E0208345
        cat=adv
        variants=inv
        modification_type=verb_modifier;manner
        abbreviation_of=low|E0038105
}

{base=univ
entry=E0604635
        cat=adv
        variants=inv
        modification_type=verb_modifier;manner
        modification_type=intensifier
        abbreviation_of=universally|E0063255
annotation=per merriam-webster.com
}
```

J. Verb acronyms and abbreviations are rare, but do exist. Allen's original training materials do not address how to code them, as those had apparently not been encountered in our work before the mid 1990s. This example may be used as a template:

```
{base=KO
spelling_variant=K/O
entry=E0208336
        cat=verb
        variants=irreg|K/O|K/O's|K/O'd|K/O'd|K/O'ing|
        variants=irreg|KO|KO'd|KO'd|KO'd|KO'ing|
        tran=np
        tran=pphr(of,np)
        acronym_of=keep out
        acronym_of=knock out
annotation=for variants See Websters 9th New Collegiate -acb
annotation="KO'd by R & D speak" -- UI - 95057710
annotation="He has continued to improve KO'ing Oleg and giving Marco Ruas
everything he could handle before losing by ankle lock. He got KO'ed by Igor
Vovchanin at Pride 4." -- www.fightingtalk.com on Gary Goodridge.
annotation="knock out" verb listed under the verb "knock"
}
```