

Derivations Tagging for Lexical Tools Updates

This task is done yearly in July for the Lexicon/Lexical Tools release for the following calendar year. It entails judgments about whether pairs of lexical items given to you on a list are derivational pairs, or dPairs. The members of a dPair are derivationally related by **exactly** one derivational step, often the addition of a suffix or prefix; sometimes via “zero derivation” (or functional shift). Examples:

kind (adj)/kindly

approve/preapprove

arrange/arrangement

Arctic (adj)/Arctic (noun)

Each example above exemplifies a rule or abstraction about the derivational relationship between the pair members:

\$|adj||ly\$|adv|

\$|verb|\$pre|verb|

\$|verb|\$ment|noun|

\$|adj|\$|noun|

There are many such dPair rules, and they are assembled and distributed with the Lexical Tools. Because we are concerned with relatedness rather than etymology, directionality is irrelevant here. The rules above could just as validly have been written with the pair’s items in the opposite order (i.e. \$|ly\$|adv| \$|adj|\$; \$|noun|\$|adj| etc.). The focus here is on whether there is or is not a one-step derivational relationship. If there is, the pair is tagged “yes;” if not, it is tagged “no.” With rare exceptions, all spelling variants should be tagged alike. If you tag a new spelling variant in a way that conflicts with previous tagging for that pair of records, the issue will be bounced back to you to resolve. Resolve any conflicts however you see the evidence pointing you.

Dictionary meanings for affixes should be consulted in dPair tagging. The first question should be whether the pair in question actually exemplify the use of the affix, or if it is rather an accidental resemblance. For example, the suffix –y can be added to certain nouns to form related nouns with more abstract meanings, e.g. microscope/microscopy, but in the following examples, the –y does not represent the suffix:

Gu|noun|E0729310|Guy|noun|E0761390|no

Go|noun|E0766472|Goy|noun|E0570695|no

Below, the ‘prefix’ + acronym does not get a yes-tag, because the prefix is actually an abbreviation; *apo* is not the prefix *apo-* (meaning ‘formed from; related to’ in chemical names), but an abbreviation of *apolipoprotein*:

apo|apoL1|noun|E0761135|L1|noun|E0003688|no

apo|apoL1|noun|E0761135|L1|noun|E0618251|no

apo|apoL1|noun|E0761135|L1|adj|E0618257|no

Not every related pair of words will be a dPair. For example, kind/kindness are related by 2 derivational steps (kind>kindly>kindness) and thus are not a dPair. The same can be said for approve/preapproval (approve/preapprove/preapproval) and arrange/rearrangement (arrange/rearrange/rearrangement). Other pairs are related etymologically, but are not dPairs, e.g. medicine/medical.

Prefix or suffix tagging is often straightforward, when the affix appends to a single word that is not a nominalization:

mini|miniresidency|noun|E0761515|residency|noun|E0205002|yes
mini-|mini-residency|noun|E0761515|residency|noun|E0205002|yes

When there are homographs in multiple parts of speech, pay close attention to the derivational steps involved. The adjective *antihypertensive* (/anti-hypertensive/anti hypertensive) most plausibly derives from the adjective *hypertensive* and not from the noun *hypertensive*, which itself is a zero-derivation from the adjective. The noun *antihypertensive* (/anti-hypertensive/anti hypertensive) most plausibly derives from the adjective *antihypertensive*, analogous to the derivation of the noun *hypertensive* from the adjective *hypertensive*. This is reflected in the tagging here:

anti|anti hypertensive|adj|E0009558|hypertensive|adj|E0032819|yes
anti|anti hypertensive|adj|E0009558|hypertensive|noun|E0032820|no
anti|anti hypertensive|noun|E0009559|hypertensive|adj|E0032819|no
anti|anti hypertensive|noun|E0009559|hypertensive|noun|E0032820|no

It will often be the case that verbs will derive from verbs & nouns from nouns:

co|cobind|verb|E0764160|bind|verb|E0012939|yes
co|cobind|verb|E0764160|bind|noun|E0012940|no
co-|co-bind|verb|E0764160|bind|verb|E0012939|yes
co-|co-bind|verb|E0764160|bind|noun|E0012940|no

Related zero derivations are taken case by case. If there were such as noun as *cobind*, you'd have to decide if it derived from the verb *cobind* via zero derivation, or if it was co- + (N) bind.

Though verbs will often derive from verbs, nouns from nouns, adjectives from adjectives, "often," however, is not "always," even within a given affix. In the examples below, the *anti* prefix derives adjectives from nouns. The noun *Parkinson's* may be derived from the possessive form, but it is easily used as an uncount noun, e.g. "He has Parkinson's, not essential tremor." Thus, if the adjective *anti-Parkinson's* is used to describe, say, a treatment regimen, this adjective is readily analyzed as *anti-* + (N) *Parkinson's*. Likewise, if a treatment is described as *anti-Treponema*, it is *anti-* + (N) *Treponema*:

anti-|anti-Parkinson's|adj|E0732511|Parkinson's|noun|E0763255|yes
anti-|anti-Treponema|adj|E0009811|Treponema|noun|E0421913|yes

A specific exception relating to the use of the prefix *anti* must be mentioned. The several meanings of *anti* all have to do with opposition or an opposite effect. Within biomedicine, expressions of the type 'anti-X antibody' are common, and mean 'antibody against X'. It is also commonly accepted that 'X antibody' also means 'antibody against X'. Thus, 'anti-X antibody' and 'X antibody' can have the same referent and are thus not a derivational pair (dPair) where the one with the initial prefix *anti-* refers to something with the opposite effect or value of the expression without it:

anti-|anti-IgE antibody|noun|E0736964|IgE antibody|noun|E0765392|no

In general, nominalization pairs (adj/noun or verb/noun) will be dPairs, and vice versa. One exception is with nouns ending in –ism, which are dPairs, but only rarely nominalizations. Nominalizations are semantically very general. The nominalization of an adjective means ‘the state or quality of being ADJ,’ e.g. kindly/kindliness. The nominalization of a verb means ‘the act or action of VERBing,’ e.g. arrange/arrangement. With –ism, the meaning of the noun often relates to a philosophy or religion, clearly going beyond more neutral nominalization glosses, e.g. imperial/imperialism, minimal/minimalism. Pairs like these do not involve nominalization, but are dPairs. There are a few true nominalizations ending in –ism, e.g. lyric/lyricism, synclitic/syncliticism.

Prefixes and suffixes often apply only to the word to which they are affixed. Thus, prefixes to LMWs (LexMultiWords) will generally not get yes-tags, because the prefix relates not to the LMW as a whole, but to its first element. *Autoimmune disease* is not immune disease that is auto- (‘automatic’) but rather disease that can be described as autoimmune; the steps in word derivation are: *auto-* + *immune*; *autoimmune* + *disease*:

```
auto|autoimmune disease|noun|E0011307|immune disease|noun|E0767305|no
auto-|auto-immune disease|noun|E0011307|immune disease|noun|E0767305|no
auto |auto immune disease|noun|E0011307|immune disease|noun|E0767305|no
```

The order of derivation is critical. *De-pollution* is not *de-* + *pollution*, but [*de-pollute*] + *-ion*, as confirmed by the LB record:

```
de-|de-pollution|noun|E0692017|pollution|noun|E0048669|no
```

```
{base=depollution
spelling_variant=de-pollution
entry=E0692017
    cat=noun
    variants=uncount
    compl=pphr(of,np)
    nominalization_of=depollute|verb|E0692021
}
```

Chemical prefixes are a more complicated thing. If the prefix occurs at the beginning of a LMW, and the prefix’s meaning does not plausibly apply to the entire LMW, it gets a no-tag. Di- means ‘containing two atoms, radicals or groups (of a specified kind)’ and prefixed to a LMW, it cannot plausibly apply to the whole LMW:

```
di|dihydroxyphenyl acetic acid|noun|E0766767|hydroxyphenyl acetic acid|noun|E0751008|no
di|dipeptide boronic acid|noun|E0766501|peptide boronic acid|noun|E0766503|no
di|dimethylphenyl carbinol|noun|E0308994|methylphenyl carbinol|noun|E0507571|no
```

Some tagging involves both prefixes and suffixes. In some of these cases you will have to decide what the order of affix derivation is. Sometimes, more than one order of affix derivation is OK. Take, for example, suffix tagging of this line:

```
pre-chondroblast|noun|E0610906|pre-chondroblastic|adj|E0610789|
```

This line brings up the question of two possible orders of derivation: *pre-chondroblastic* could be either *pre- + chondroblastic* or [*pre-chondroblast*] + *-ic*. Because we aim to be synchronically descriptive, rather than diachronically accurate, the slightly differing meanings of the two derivations are relevant here. Is something that is *pre-chondroblastic* characteristic of a pre-chondroblast ([*pre-chondroblast*] + *-ic*) or is it before a chondroblastic stage (*pre- + chondroblastic*)? Since both meanings are possible, both derivations are possible. In the tagging process, both would get yes-tags.

Similarly, there may arise cases of single lexical items with multiple possible morphological decompositions depending on the semantic interpretation of the term. For example, the same word *singer* could be analyzed as either one who sings (*sing + er*) or one who *singes* (*singe + er*). Since both analyses are plausible without additional context, both of these represent valid dPairs and would be given yes-tags.

We use the terms affix, prefix and suffix rather more loosely than Merriam-Webster's dictionary does. In addition to forms M-W calls affixes, we also include most things M-W calls combining forms (e.g. *auto-*), because of their prolific use in biomedicine. Our more inclusive approach stems from our practical aim to make the Lexicon and Lexical Tools of as much use and interest to NLP researchers as possible.

Revised July 27, 2020 by Amanda Payne