2-3. Antonyms – Source of PrefixD (PD)

Antonym candidates – from Prefix Derivation with negative tags

Negative tags in the Lexicon prefix derivations are used to generate antonym candidates. These candidates are tagged by linguists to generate antonyms and negation detection cue words. The general processes include:

- Generate antonym candidates from lexicon prefix derivations with negative tags
- Manually tagging
- Validate tags
- Update tags to annual release tag file

They are described in the following sections.

1. Generate Antonym Candidates

Antonym candidates are retrieved from the Lexicon derivation file (derivation.data.${YEAR}). These antonym candidates are retrieved from the following prefixes: a-, an-, anti-, contra-, counter-, de-, dis-,dys-, il-, im-, in-, ir, mis-, non-, un-, under- with five POS (noun, verb, adj, adv and prep).
It is possible that prefixes in a prefix derivation pair (dPair) are not included in the above list, yet, it is a valid antonym. For example, extra- is the prefix of ordinary|extraordinary, which is a legit prefix and aPair. Also, antonym pairs might include multiple prefix derivations, such as overdose|underdose which involves prefix derivations of dose|underdose and dose|overdose. The above two cases of aPairs from prefix derivations can be found in the model of collocates in a corpus [CC] and thus are not discussed here.

1.1 Algorithm

- derivation.data include 9 fields as shown in table 1:

Derivation1	POS1	EUI1	Derivation2	POS2	EUI2	Negation	type	prefix

Table 1. Format in the Lexicon derivation file

- Retrieve antonym candidates as follows:
  - Prefix: type=P
  - Negative: negation=N
  - Same POS: POS1=POS2
  - Derivation1 is a legal single word
  - Derivation2 is a legal single word
  - Sort the candidates by EUI2, then Ant2
    => This sorting order makes dPairs with spVars appear next to each other. It is easier for tagging because the same tags can be used for aPairs with spVars.

1.2 Program - Derivation.GetAntCandFromPrefixD

- Input:

- Link ./3.SuffixD/${YEAR}/input/derivation.data
- Link ./1.Lexicon/${YEAR}/input/inflVars.data
- Update ./0.antonym/input/antCand.data.${YEAR}.tag (the latest annual release antonym candidates tagged file)
- Update ./0.antonym/input/domain.data

```
Shell> cd  ${ANTONYM}/bin
Shell> GetAntonyms ${YEAR}
30
```

- output files:
  - ./output/antCandPrefixD.data
  - ./output/antCandPrefixD.data.tag
  - ./output/antCandPrefixD.data.tbd (candidate file)
    - ⇨ If this file is not 0, use this file as new candidates and send it to linguists to tag

1.3 Output – Candidate Format and Examples

The output candidate file has 10 fields, as shown in Table 2. The blank cells are automatically tagged with [XXX_TBD], where XXX are CANON, TYPE, NEG and DOMAIN.

Ant1	**EUI1**	**Ant2**	**EUI2**	**POS**	**Canon**	**Type**	**Negation**	**Domain**	**Source**
dendritic	E0021551	adendritic	E0007276	adj					PD
icteric	E0033238	anicteric	E0009085	adj					PD
bacterially	E0011689	antibacterially	E0597113	adv			O		PD
versive	E0226445	contraversive	E0333343	adj					PD
irritate	E0035928	counterirritate	E0731853	verb					PD
methylate	E0039953	demethylate	E0021473	verb					PD
repute	E0052910	disrepute	E0023440	noun					PD
thyroidal	E0321006	dysthyroidal	E0356056	adj					PD
limited	E0220623	illimited	E0568899	adj					PD

potentia	E0564117	impotentia	E0435443	noun					PD
valid	E0063937	invalid	E0035671	adj					PD
responsive	E0053052	irresponsive	E0229837	adj					PD
type	E0062731	mistype	E0605948	verb					PD
phonetic	E0724841	nonphonetic	E0540832	adj					PD
like	E0419447	unlike		prep	Y	AB	BN	quality	PD
supply	E0059139	undersupply	E0222792	noun					PD
…	…	…	…	…					PD

Table 2. Examples of antonym candidates from Lexicon prefix derivations with negative tags

2. Tag Candidates

Manual tagging is needed for the (new) antonym candidates generated from the above process. The tagged information of pre-existing candidates from previous years is saved and used as the baseline for future releases. Please refer to document 1.2.Antonym-Tags for details.

3. Validate and auto-fix Antonym Candidate Tags

Manual tags are verified by computer programs to:

- ensure all tags are valid
- automatically assign type to [NA] and domain to [DOMAIN_NONE] if Canon is [N]
- check for new domains.

Please refer to document 1-2.LexAntonym-Tag for details.

4. Update to Annual Release Antonym Tag file

- add tag result from source of SD to .${0.Antonym}/${YEAR}/input/antCand.data.tag.${YEAR}
- rerun the processes 1-3 until all candidates have valid tags (antCandPreffixD.data.tbd = 0)

Please refer to document 1-2.LexAntonym-Tag for details.