

## 2-1. Antonyms – Source of Lexicon (Lexical Records)

### Antonym candidates – from Lexical Records with negative tags

Use the negative and broad\_negative tags in the Lexical Records 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 lexical records with negative tags
- Manual tagging
- Validate tags
- Update tags to annual release tag file

They are described in the following sections.

### 1. Generate Antonym Candidates

Use the negative and broad negative tags in the Lexical Records to generate antonym candidates.

Antonym candidates are retrieved from the 7 POSs of [adv|pron|aux|modals|prep|det|conj] that have negative tags as described below. POSs of [noun|adj|verb|compl] do not have negative tags in Lexicon.

#### 1.1 Algorithm

- Adverbs (adv):
  - true negative/strict negation (negative): never, no, not, nowise
  - broadly negative (broad\_negative): hardly, seldom, rarely, even, either, little, scarcely, slightly, barely, seldomly.
- Pronoun (pron):
  - type=indef(neg): none, nobody, nothing, noone, neither, naught
- Auxiliary (aux) - negative:
  - variant=isn't;pres(thr\_sing):negative
  - variant=aren't;pres(fst\_plur,second,thr\_plur):negative
  - variant=don't;pres(fst\_sing,fst\_plur,second,thr\_plur):negative
  - variant=haven't;pres(fst\_sing,fst\_plur,second,thr\_plur):negative
  - ...
- Modal (modal) - negative:
  - variant=mayn't;pres:negative
  - variant=mightn't;past:negative
  - variant=mustn't;pres:negative
  - variant=couldn't;past:negative
  - variant=cannot;pres:negative
  - variant=can't;pres:negative
- Preposition (prep):
  - true negative/strict negation (negative): without
  - broadly negative (broad\_negative): unlikely (not used as negation cue word)
- Determiner (det):
  - true negative/strict negation (negative): no, neither, nary a, nary an
- Conjunction (conj):

- true negative/strict negation (negative): neither, nor

### 1.2 Program – GenAntCandFromLexicon.java

- Inputs:
  - LEXICON: 1.Lexicon/{YEAR}/input/LEXICON
  - Tag file: 0.Antonym/{YEAR}/input/antCand.data.tag.{YEAR}
    - Use the previous annual release tag file as the baseline, then update new tags to it.
  - Domain file: 0.Antonym/{YEAR}/input/domain.data

Shell> cd {ANTONYM}/bin

Shell> GetAntonyms {YEAR}

10

- if ./output/antCandLexicon.data.tbd not 0, use it as candidates and send to linguists to tag

### 1.3 Output - Candidate Format and Examples

Table 1 shows 10 fields of the output candidate file.

Ant1	EUI1	Ant2	EUI2	POS	Canon	Type	Negation	Domain	Source
like	E0419447	unlike	E0063275	prep	Y	AB	BN	quality	LEX
with	E0065516	without	E0065526	prep	Y	B	N	existence	LEX
is	E0012152	isn't	E0012152	aux			N		LEX
could	E0014877	couldn't	E0014877	modal			N		LEX
		little	E0037839	adv			BN		LEX
		neither	E0042155	det			N		LEX
		neither	E0042156	pron			N		LEX
		neither	E0783690	conj			N		LEX
		hardly	E0030831	adv			BN		LEX
		neither	E0783690	conj			N		LEX
		...	...	...			...		LEX

**Table 1 Examples of antonym candidates from Lexicon (lexical records with negative tag)**

Please note that blank cells are automatically filled up with [XX-TBD], such as [ANTONYM\_TBD], [EUI\_TBD], [POS\_TBD], [CANON\_TBD], [TYPE\_TBD], [NEG\_TBD] and [DOMAIN\_TBD].

## 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. The process is detailed as follows.

### 2.1 Action

- Prep: tags are generated completely,
  - to verify all fields
- Aux and modal:
  - to verify Ant1, EUI1, negation
  - to tag Canonical, Type, Domain
- Adv, det, pron and conj:
  - to verify negation
  - to fill/tag Ant1, EUI1, Canonical, Type, Domain
- The candidate file uses [XXX-TBD] for those blank cells in table 1. Please replace [XXX\_TBD] with correct content. If no content is applicable please change [XXX\_TBD] to [XXX\_NONE] for the fields of **Ant1, EUI1 and Domain**.
- New candidates from LEX are expected to be few in number after the first release.

### 2.2 Fields & Tags

Please refer to document 1-2.LexAntonyms-Tag for definitions of fields and tags. Table 2 shows examples of tagged antonym candidates from the lexical records with negative tags.

Ant1	EUI1	Ant2	EUI2	POS	Canonical	Type	Negation	Domain	Source
like	E0419447	unlike	E0063275	prep	Y	AB	BN	quality	LEX
with	E0065516	without	E0065526	prep	Y	B	N	existence	LEX
is	E0012152	isn't	E0012152	aux	Y	B	N	existence	LEX
could	E0014877	couldn't	E0014877	modal	Y	B	N	possibility	LEX
much	E0587115	little	E0037839	adv	Y	UB	BN	quantity	LEX

either	E0024626	neither	E0042155	det	Y	B	N	existence	LEX
either	E0024627	neither	E0042156	pron	Y	B	N	existence	LEX
easily	E0024341	hardly	E0030831	adv	N	UB	BN	quality	LEX
NONE	NONE	neither	E0783690	conj	N	NA	N	NONE	LEX
...	...	...	...	...	...	...	...	...	LEX

**Table 2 Examples of tagged antonym candidates from Lexicon (lexical records with negative tag)**

### 3. Validate and Auto-fix Antonym Candidates

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 LEX to `.$0.Antonym}/${YEAR}/input/antCand.data.tag.${YEAR}`
- rerun the processes 1-3 until all candidates have valid tags (antCandLexicon.data.tbd = 0)

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