

## 2-4. Antonyms – Source of Training and Test Set (TT)

### Establish Antonym Training and Test Set

In order to find the characteristics and patterns of antonyms, we established a collection of antonym pairs (aPairs) from various sources on the internet. Some sources include duplicated aPairs. For example, aPairs [absence|presence] and [presence|absence] are considered as the same aPair and counted as 1 unique aPair. In addition, antonyms in aPairs are lowercased and single word only. Multi-word aPairs, such as [already|not yet] or [none of|a lot of], are removed from the collection. The source web sites, the number of unique aPairs and URLs of this training and test set are shown in Table 1.

Source	No. of unique aPairs	URL
Sherwood School	449	<a href="http://sherwoodschool.ru/en/lessons1/english/english_antonyms_list">http://sherwoodschool.ru/en/lessons1/english/english_antonyms_list</a>
Proof Reading Services	418	<a href="https://www.proofreadingservices.com/pages/antonym-pairs">https://www.proofreadingservices.com/pages/antonym-pairs</a>
Enchanted Learning	324	<a href="https://www.enchantedlearning.com/wordlist/opposites.shtml">https://www.enchantedlearning.com/wordlist/opposites.shtml</a>
7ESL	339	<a href="https://7esl.com/opposites/">https://7esl.com/opposites/</a>
English Grammar Here	321	<a href="https://englishgrammarhere.com/opposite-words/1000-opposite-words-list">https://englishgrammarhere.com/opposite-words/1000-opposite-words-list</a>
Synonyms Antonyms	301	<a href="http://www.synonyms-antonyms.com/antonyms-list.html">http://www.synonyms-antonyms.com/antonyms-list.html</a>
SLP Lesson Plans	251	<a href="http://www.slplessonplans.com/files/antonymlist.pdf">http://www.slplessonplans.com/files/antonymlist.pdf</a>
ESL Forums	198	<a href="https://eslforums.com/antonyms/">https://eslforums.com/antonyms/</a>
My English Tutors	170	<a href="https://myenglishtutors.org/antonyms-examples/">https://myenglishtutors.org/antonyms-examples/</a>
Love To Know	167	<a href="https://home-school.lovetoknow.com/Antonyms_Word_List">https://home-school.lovetoknow.com/Antonyms_Word_List</a>
Your Dictionary	159	<a href="https://examples.yourdictionary.com/examples-of-antonyms.html">https://examples.yourdictionary.com/examples-of-antonyms.html</a>
Classic Thesaurus	100	<a href="https://www.classicthesaurus.com/100_examples_of_antonyms">https://www.classicthesaurus.com/100_examples_of_antonyms</a>
Power Thesaurus	100	<a href="https://www.powerthesaurus.org/100_examples_of_antonyms">https://www.powerthesaurus.org/100_examples_of_antonyms</a>
Smart Words	9	<a href="https://www.smart-words.org/list-of-synonyms/">https://www.smart-words.org/list-of-synonyms/</a>

**Table 1. Source web sites, number of unique aPairs and URLs of antonym training and test set**

### 1. Collection of Antonyms

A program is developed to:

- collect aPairs from various antonym sources
- unify aPairs (to remove duplications) from above collections
- identify the source of antonyms

- sort antonyms by sources first, then alphabetical order.

This antonym collection includes 1000+ unique aPairs. This program is detailed as follows.

### 1.1 Algorithm – CollectAntonyms.java

- Input:
  - `${ANTONYM}/2.TtSet/${YEAR}/input/antonymSource.data:`  
Includes the source antonym files from `./Sources/`
  - `${ANTONYM}/2.TtSet/${YEAR}/input/Sources/*.data` (files listed in `antonymSource.data` that contain antonyms)

ant1	ant2
------	------

- Identify Sources and unify the aPair

Four antonym sources are identified by computer programs (AntObj.java). The algorithm is described as follows:

- LEX:
  - Lexical records with POSs of [adv|pron|aux|modals|prep|det|conj] and the negative and broad negative tags are used as aPairs and tagged with source of LEX. For an aPair with source of LEX, the negative antonym is stored as ant2. For example, [with|without] is a LEX aPair and the negative antonym [without] is stored as ant2. If the aPair from TT is in the LEX aPair set, the source is automatically identified as LEX.
- SD:
 

The algorithm for identifying a SD (suffixD) aPair is described as follows:

  - if one and only one of the antonyms ends with suffix “-less”
  - the root of the antonym that ends with suffix “-less” is also the root of the other antonym.
  - set the antonym ending with suffix “-less” as ant2 (negative), such as [careful|careless]
- PD:
 

The algorithm for identifying a PD (prefixD) aPair is described as follows:

  - $\text{ant1} \neq \text{ant2}$
  - The prefix belongs to the set of: a-, an-, anti- contra-, counter-, de-, dis-, dys-, il-, im-, in-, ir-, mis-, non-, un-, under-
  - ant1 is the root of ant2 or ant2 is the root of ant1.
  - Set the antonym with prefix to ant2 (negative), such as [possible|impossible]
- CC:
 

Collocates in a Corpus, our first attempt is to use the terms co-occurring in MEDLINE). If an aPair does not belong to above source, it is assigned as SN (semantic network). These are aPairs retrieved by co-occurring patterns from a corpus.

  - Set the ant1 and ant2 by alphabetic order, such as [accept|refuse]  
In other words, ant2 is not necessary the negative antonym.
- SN:
 

Semantic opposite in corpora. These are aPairs retrieved from a semantic network. Patterns are yet to be developed.

  - Set the ant1 and ant2 by alphabetic order, such as [admit|deny]  
In other words, ant2 is not necessary the negative antonym.

- Sort all aPair by Source, ant1|ant2

### 1.2 Program - TtSet.CollectAntonyms.java

```
Shell> cd ${ANTONYM}/bin
```

```
Shell> GetAntonyms ${YEAR}
```

```
40
```

### 1.3 Output – Antonym Pair Format and Examples

The output file (\${2.TtSet}/\${YEAR}/output/antonymTtSet.data) includes 1000 unique aPairs (without EUI or POS information). Table 2 shows examples and 3-field format of retrieved aPairs.

Ant1	Ant2	Source
accept	refuse	CC
start	stop	CC
bye	hi	SN
admit	deny	SN
with	without	LEX
always	never	LEX
able	unable	PD
possible	impossible	PD
careful	careless	SD
worthy	worthless	SD
...	...	...

**Table 2. Examples of antonym with sources from antonym training and test set**

## 2. Generate Antonym Candidates

A computer program is developed to generate antonym candidates from the collected antonym pairs. These antonym candidates are then tagged by our linguists. They are described below.

### 2.1. Algorithm - GenAntCandFromTtSet.java

- Input:
  - aPair: \${4.TtSet}/\${YEAR}/output/antonymTtSet.data
  - infl\_vars: \${1.Lexicon}/\${YEAR}/input/inflVars.data
  - tagged file: \${0.Antonym}/\${YEAR}/input/antCand.data.tag.\${YEAR}
  - domain: \${0.Antonym}/\${YEAR}/input/domain.data
- Go through all aPairs retrieved from the antonym training and test set
- Add EUI and POS information

- If neither antonym has EUI, assign [POS\_TBD] to POS field
  - Assign [EUI\_TBD] to the EUI field for those antonyms without EUI
- If only one antonym has EUI(s), use information with all POSs.
  - Assign [EUI\_TBD] to the EUI field for those antonyms without EUI
- If both antonyms have EUI(s), only use information with the same POS.
  - If already tagged (use ant1|EUI1|ant2|EUI2|POS as key), use previous tag
  - If not tagged, assign tag
    - ⇒ Update source to SD|PD|CC if it is LEX  
 LEX was assigned by ant1|ant2 without POS information in process-40. Accordingly, SOURCE is re-assigned to SD|PD|CC because they are not tagged in LEX (the one tagged in LEX has different POS).
- The source of output is temperately assigned to [TT], run option 41 to retag [TT] to [CC] or [SN] depends if the aPair are collocates in MELDINE or not, respectively. This is a fix due to the updates on new source tag of [SN].

## 2.2. Program

```
Shell> cd ${ANTONYM}/bin
Shell> GetAntonyms ${YEAR}
42
```

## 2.3. Output – Candidate Format and Examples:

The output file (\${4.TtSet}/\${YEAR}/output/antCandTtSet.data) is further split into the following two files:

- ./output/antCandTtSet.data.tag (already tagged)
- ./output/antCandTtSet.data.tbd (candidate file)

Output antonym candidate file (\${4.TtSet}/\${YEAR}/output/antCandTtSet.data.tbd). Table 3 shows examples of output file from antonym candidates derived from Antonym training and test set. Please note that blank cells are automatically filled up with [XX-TBD], such as [EUI\_TBD], [POS\_TBD], [CANON\_TBD], [TYPE\_TBD], [NEG\_TBD] and [DOMAIN\_TBD].

Ant1	EUI1	Ant2	EUI2	POS	Canon	Type	Negation	Domain	Source
accept	E0006699	refuse	E0052488	verb					CC
start	E0057681	stop	E0058047	noun					CC
start	E0057680	stop	E0058048	verb					CC
bye	E0785958	hi	E0785959	noun					SC
admit	E0007437	deny	E0021749	verb					SN
with	E0065517	without	E0065527	conj					LEX -> CC

always	E0008403	never	E0042565	adv					LEX -> CC
able	E0006510	unable	E0062907	adj					PD
possible	E0049058	impossible	E0033808	adj					PD
careful	E0015340	careless	E0015344	adj					SD
worthy	E0065647	worthless	E0065643	adj					SD
...	...	...	...	...					...

**Table 3. Examples of antonym candidates from antonym training and test set**

### 3. 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.

### 4. Validate and auto-fix Antonym Candidates 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-Tags for details.

### 5. 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 2-4 until:
  - all candidates have valid tags (antCandPrefixD.data.tbd = 0), Please refer to document 1-2.LexAntonym-Tags for details.
  - All candidates left in antCandPrefixD.data.tbd do not have EUI (EUI\_TBD).
    - A tagged antCand can't be located if EUI1 or EUI2 is [EUI-TBD]. That is ANT1 or ANT2 are not in the Lexicon.
    - For those ANT1 and ANT2 not in the Lexicon, use them as candidates to add them to the Lexicon. Then, rerun process-41 to generate antonym candidates.