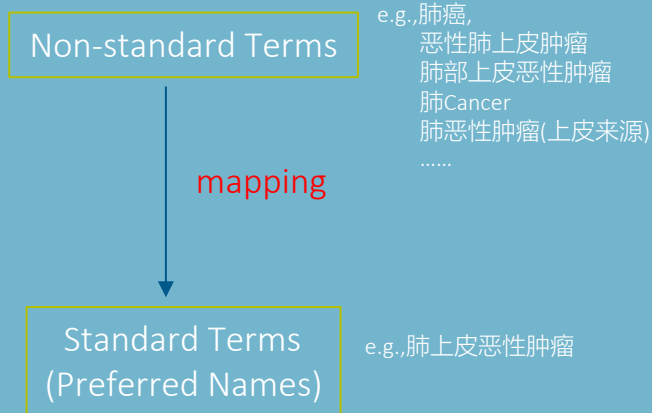


Concept Mapping of Medical Terms

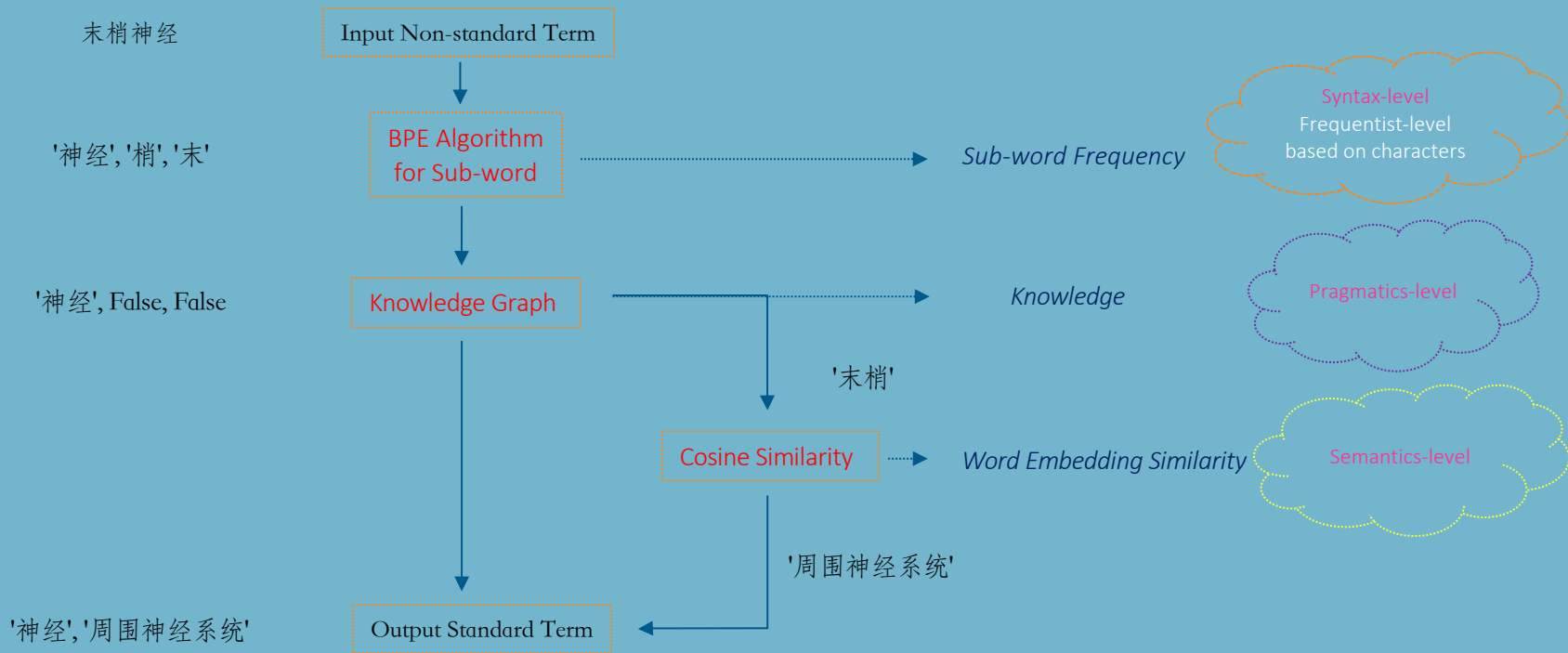
via Syntax, Semantics, and Pragmatics Levels

Shuyue Jia

Objectives

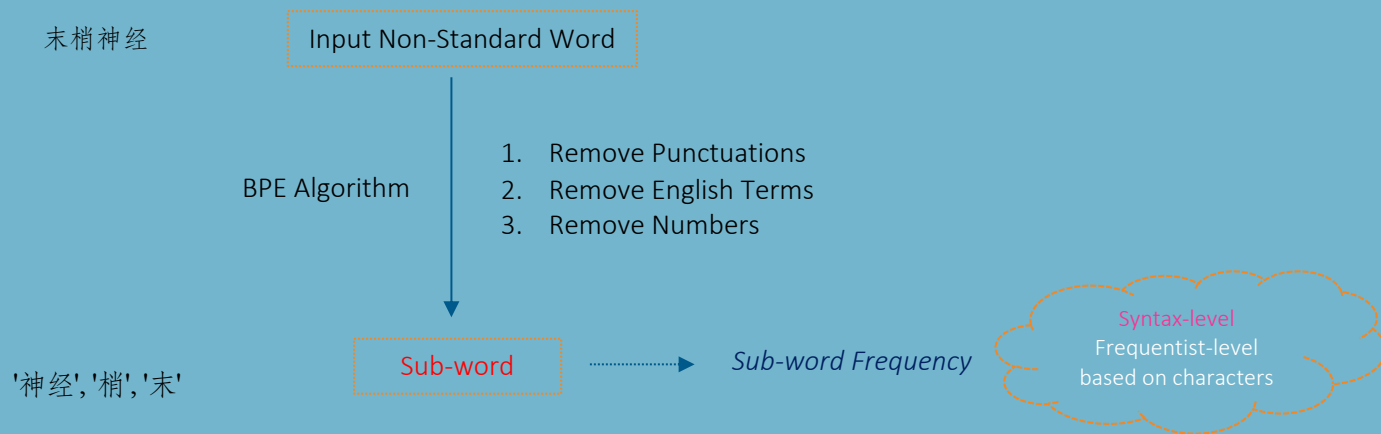


Method



Method – Syntax level Analysis

Why using Sub-words?



1. Learn **compounding** and **transliteration** from sub-word representations
2. Generalize to **translate** and **produce** new words (unseen at training time)

Getting **Sub-words**:

Byte Pair (2-gram) Encoding (BPE) Algorithm Recap:

- 1) Initialize symbol vocabulary with character vocabulary

`vocab = {'l o w .': 5, 'l o w e r .': 2, 'n e w e s t .': 6, 'w i d e s t .': 3}`

- 2) Find the most frequent 2-gram pairs ('A', 'B') from every word

`{('d', 'e'): 3, ('e', 'r'): 2, ('l', 'o'): 7, ('w', '.'): 5, ('w', 'e'): 8, ('e', 'w'): 6, ('r', '.'): 2, ('w', 'i'): 3, ('e', 's'): 9, ('n', 'e'): 6, ('s', 't'): 9, ('i', 'd'): 3, ('t', '.'): 9, ('o', 'w'): 7}`

We find ('e', 's'): 9

- 3) Merge ('A', 'B') $\rightarrow$ ('AB') and repeat 2).

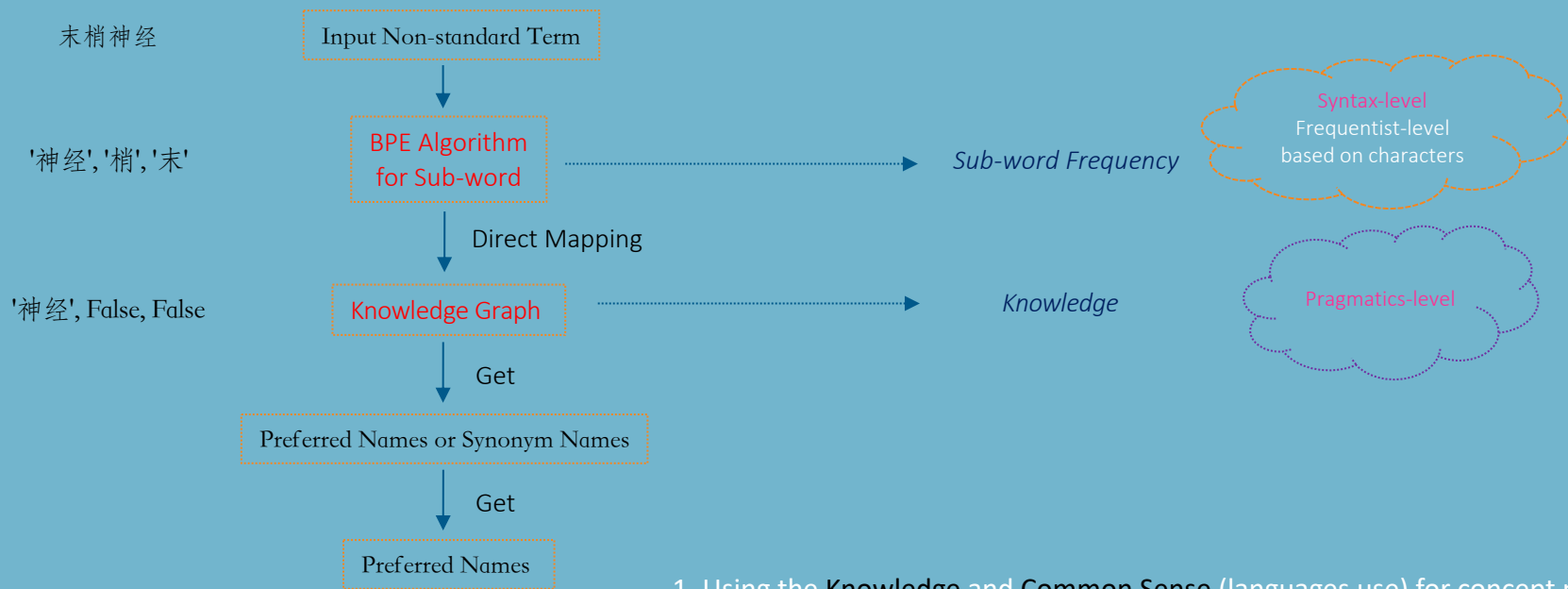
`{'l o w': 5, 'l o w e r': 2, 'n e w e s t': 6, 'w i d e s t': 3}`

Character 2-gram

- 4) Stop merging until reach the *num(merge operation)* or *minimum frequency*

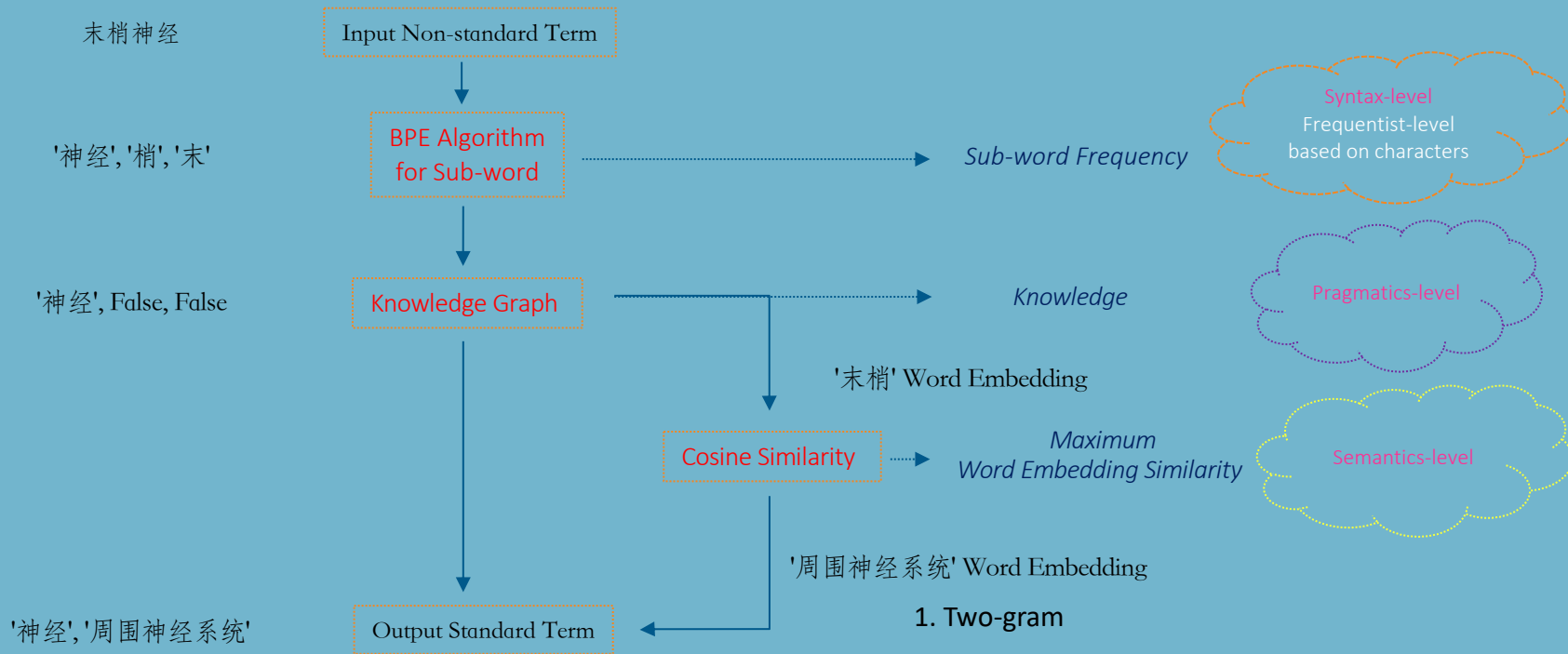
Method – Pragmatics level Analysis

Using Knowledge Graph



1. Using the **Knowledge and Common Sense** (languages use) for concept mapping
2. Pre-defined **Look-up Dictionary** in the field of Medical AI

Method – Semantic level Analysis



Find the most similar concept via word Embedding's Cosine Similarity

1. Two-gram
2. Sub-words
3. Forward-string grams
4. Backward-string grams
5. (If there is) Negative & other character gram

Current Challenges

1. Negative words → Add negative word's embedding to sub-word Embedding '非感染性' VS '感染性'
2. English words & Numbers → Sub-word Max Frequency 'Willis环'
3. Punctuations
4. Abbreviations
5. Isomorphism ← Similar Word Embedding!

PHILIPS

www.philips.com

Thanks and have a nice day!

innovation  you