

Week 9

Automatic recognition of grammar patterns

for providing writing suggestions to learners

April 24, 2018

Verb patterns and Dictionary

- In Collins Cobuild English Dictionary, each sense entry of is annotated with patterns observed in a corpus
- Other publishers followed suit and provide reference information similar patterns in different format
- Learner's writing errors can be attributed to ...
 - **V n** * *We discussed about the issues.*
 - **V on n** * *He commented the situation.*
- Therefore, it is a good idea to display writing suggestions as the learners write away
 - grammar patterns of the last word typed
 - relevant examples
 - related collocations

Existing Resources: Macmillan English Dictionary

ability

ability to do something: Tiredness can seriously impair your ability to drive.

above

above all else: Above all else, the government must keep the promises it has made.

absence

in someone's absence : Mark will be in charge in my absence.

absence of: a complete absence of humor

in the absence of something: In the absence of any contrary agreement, the firm accepts full liability.

abstract

abstract idea/concept/principle/notion: Mathematics is concerned with understanding abstract concepts.

accept

accept that: Most scientists accept that climate change is linked to pollution.

generally/widely accepted : His views on genetics are not now widely accepted.

accept blame/responsibility/liability: We cannot accept liability for items stolen from your car.

accept that: For a long time, he simply could not accept that she was dead.

accept someone as something: Mexico was accepted as a member of the OECD in 1994.

accept someone into something: She was desperate for the children to accept her into the family.

Collins COBUILD Grammar Patterns

- discuss-V
 - **V n** *discuss the issues*
 - **V wh to v** *discussed what to do next*
 - **V wh** *discussed when*
- Problems
 - Not domain specific (academic or business)
 - No priority or frequency-based
- Solution (for Academic Writing)
 - Use Academic corpus (COCA-Acad or CiteSeerx)

- Focus on key vocabulary (Academic Keyword List)
- Generate grammar patterns
- Filter and retain high-frequency patterns, much like collocation extraction method (Smadja 1995)

Application of induced grammar patterns

- Provide writing suggestions
 - *WriteAhead* (writeahead.nlpweb.org)
- Induce bilingual Synchronous Pattern Grammar
 - harmful-ADJ
 - * ADJ to n | 對 n 有害
 - * ADJ to n | 在 n 上 有害
- Machine translation

Worksheet

- Purpose
 - Expand your code from Week 8 (to cover nouns and adjectives)
 - Handle additional elements of Pattern Grammar, including the *amount* class and other function words
 - Generate a list of common grammar patterns for headwords in a word list (e.g., AKL)
 - Each pattern is associated with a GDEX example
- Input:
 - UM-Corpus.en.200k.tagged.txt.zip: a set of tagged sentences (using Genia tagger)
 - AKL.py: a list of keywords to be considered as headwords

- Starter code:
- grammarpattern.py: (do not import) reference the code when consider the noun and adjective cases
- grampat.py (use it to expand from handling verbs to verbs, nouns, and adjectives)
- Method:
 - Reuse the *getHighCounts* function
 - Normalize the count by give a high weight (e.g., $\text{length}^{**1.5}$) to long pattern
- Output:
 - A list of patterns with examples

Sample Output

ability-N

N to v (468) the ability to walk

value N

N to v (16) application values to run

N of amount (15) book value of 20.3

discuss-V

V in n (47) discussed in this article

V n (270) discusses the importance

V wh to v (15) discusses how to aggrandize

favour-V

V n (26) favour your own group

V by n (5) favoured by many migrants

classify-V

V into n (8) are classified into groups

V as n (12) classified as a type

useful-ADJ

ADJ to v (30) useful to have

ADJ for n (20) especially useful for microscopy

certain-ADJ

ADJ of n (23) certain of one thing

Bonus

- Detect passive sentence and change the pattern
 - dominate-V **V by n** (sth) *be dominated by matter*
→ **V n** *dominate sth*

Open Problems

- Filter the patterns based on supervised learning with features such as
 - pattern count
 - pattern length
 - pattern type
- Resolve pp attachment ambiguity and keep/remove pp
 - Consider "V in n: discussed in this article".

- PP "in this article" does not attach to the verb "discuss"
- Therefore,
 - * Remove "in n" part, or
 - * Ignore the whole pattern (V in n)