## 字根分解

張俊盛 清華大學資工系

2020 1203

#### 字根分解與構詞分析

- 輸入:表面形式的字 (e.g., adventure)
- 輸出:前綴(e.g., ad)、字根(e.g., vent)、後綴(e.g., ure)
- 同時提供詞性、構詞規則、解釋、字源等資訊
- 從很難記憶的一長串字母,變成幾個短、有意義、常常出現的詞綴
  - 幫助記憶
  - 幫助猜測陌生單字的詞意 (e.g. circumvent)
  - https://www.learnthat.org/word\_lists/view/13623

#### 字根分解

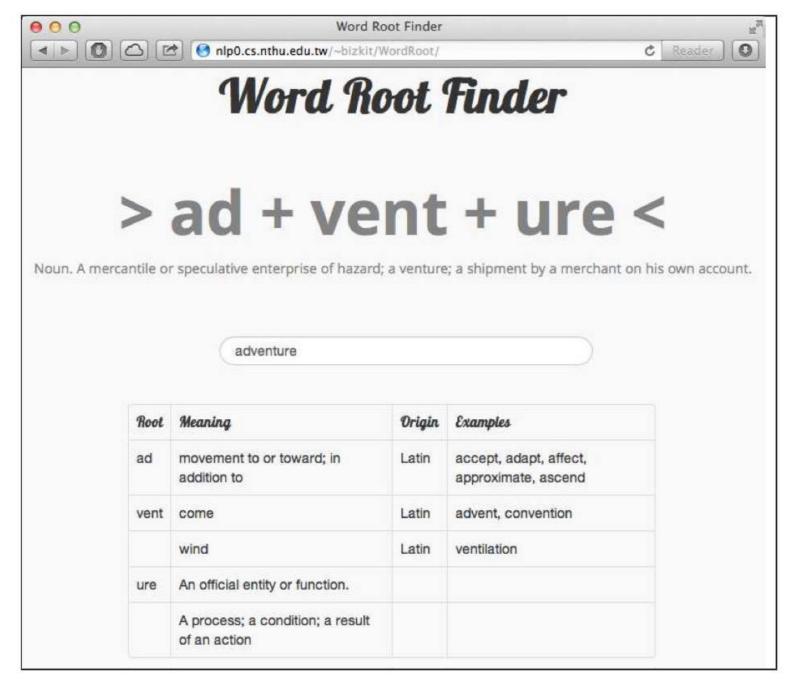
- 對學習外語者很有幫助
- 和中文部首偏旁一樣,可分解成有意義部分,易記易懂
- 對需要快速擴大詞彙量,參加 GRE and TOEFL 標準考試者 特別有用

# 各種方法

- 可以標注字根 (word roots/stems) 和詞綴作為訓練資料
- 估計字根和詞綴的出現機率
- 用機率模型計算機率最大化的字根分解
- 運用最小描述長度法 (minimum description length, MDL) 可以達到非督 導式的字根分解 (Goldsmith, 2000)
- Semantic-based algorithm (Schone and Jurafsky, 2000)
- Conditional Random Fields (Chang and Chang 2012)

# Chang and Chang 2012

Conditional Random Fields (CRF) model



### Chang and Shieh 2020

Statistical model

#### Rooter!

Online dictionary for finding word roots.

Enter a word... Search

#### adventure

#### noun

a wild and exciting undertaking (not necessarily lawful.

Morpheme	Meaning
ad	toward, to, addition
vent	
ure	

#### Problems

- Valid affixes may be applied inappropriately ("ally" stemming to "all")
- Morphological ambiguity may arise ("rating" conflating with "rat" instead of "rate")