Week 10

Inducing Synchronou Grammar Patterns

for Machine (Assisted) Translation and Bilingual Lexicography

May 1, 2018

Machine Assisted Human Translation

- **Translation Memory** tools, a database of bilingual text segments that could be reused in new translation projects
- **Terminology database/manager**, allowing translators to manage terminology bank electronically.
- **Bilingual Concordancer**, for retrieving instances of a word or an expression, its context in a monolingual, bilingual or multilingual corpus

Source: `en.wikipedia.org/wiki/Computer-assisted_translation`

English Dictionary with Grammar Patterns

- Consider "finish" in a monolingual dictionary (Collins)
  1. **[V -ing]** (finish doing something) *As soon as he'd* **finished eating**, *he excused himself.*
  2. **[V n]** (finish something that you are making) *The consultants had been working to* **finish a report** *this week.*
- Consider "avoid" in a monolingual dictionary (Collins)
  1. (take action in order to prevent something from happening)
     - **[V n]** The pilots had to take emergency action to avoid a disaster.
     - **[V -ing]** Women have to dress modestly, to avoid being harassed by the locals.

Bilingual Dictionary with Grammar Patterns

- Consider "finish" in a bilingual dictionary (Collins)

  [**V v-ing |v V**] (**FINISH v-ing** | **v＋完**) *As soon as he'd* **finished eating**, *he excused himself.* 他一吃完就告退。

  [**V n|V n**] (**FINISH n** | 完成 **n**) *The consultants had been working to* **finish a report** *this week.* 這些顧問本週一直在努力完成一份報告。

- Consider "avoid" in a bilingual dictionary

  [**V n|V n**] (**AVOID n** | 避免 **n**) *The pilots had to take emergency action to* **avoid a disaster**. 飛行員必須採取緊急行動**避免災難**。

[V -ing] (**AVOID -ing** | 避免 **v**) (*Women have to dress modestly, to* **avoid being harassed** *by the locals.*) 為了**避免**被當地人**騷擾**，婦女必須謙虛地穿著。

- Consider "care" and "rave" in a bilingual dictionary

  [V about n|V n] (**CARE about n** | 關心 **n**) (CARE about n|關心 n) **care about the future** 關心未來

  [V about n|對 n V] (**RAVE about n** | 對 **n** 讚不絕口**)** **rave about this book** 對這本書讚不絕口

Worksheet

- Purpose
  - 透過英文和翻譯對應，產生「英文規則」的對應「中文規則」
  - 用前幾週的程式，產生英文文法規則和實例（如 V about n—care about the future）
  - 找英文中心語所對應的中文、詞性
    * care 關心 V
    * about （沒對應，所以沒資料）
    * future 未來 n
  - 產生英文＋中文文法規則
    * V about n | V n
    * care about the future | 關心未來
  - 統計中文文法規則的次數

- 用前幾週的程式，過濾保留高頻（次數高於平均值，幅度為２個標準差）
- Input:
  - 英文語料庫：UM-Corpus.en.200k.tagged.en.txt
  - 平行於英文的中文語料庫：UM-Corpus.en.200k.tagged.ch.txt
  - 詞彙對應資料：align.final.200k
- Output:
  - 英文文法規則
  - 對應的中文文法規則（依照次數，由高到低排列）
  - 對應英文、中文文法規則的實例（3個例子）

Sample Input

英文資料：

[('all the commune members', ',', 'young and old', ',', 'went', 'out', 'to', 'hervest', 'the crops', '.'),
    ('all the commune member', ',', 'young and old', ',', 'go', 'out', 'to', 'hervest', 'the crop', '.'),
    ('PDT DT JJ NNS', ',', 'JJ CC JJ', ',', 'VBD', 'RP', 'TO', 'VB', 'DT NNS', '.'),
    ('I-NP I-NP I-NP H-NP', 'O', 'I-NP I-NP H-NP', 'O', 'H-VP', 'H-PRT', 'H-TO', 'H-VP', 'I-NP H-NP', 'O')]

0-4 all the commune members ,
5-8 young and old ,
9-11 went out to
12 hervest
13-14 the crops
15 .

中文資料：

0全體_N 1老少_N 2社員_N 3都_ADV 4收割_V 5莊稼_N 6去_V 7了_ASP 8。_。

詞彙對應：

0-0 1-0 2-1 2-2 3-0 4-1 5-1 6-1 7-1 8-1 9-6 10-6 11-6 12-5 13-7 14-4 14-5 15-8

0-0 1-0 2-1 2-2 3-0 4-1 all the commune members 全體 全體 老少+社員 全體 老少

5-1 6-1 7-1 8-1 young and old ，老少 老少 老少 老少

9-6 10-6 11-6 went out to 去 去 去

12-5 harvest 莊稼

13-7 14-4 14-5 the crops 了 收割＋莊稼

15-8 .

Sample Output

V n (88601)
-> V n (25868)
finish this task 完成_V 這_DET 個_CL 任務_N
finish this project 做完_V 這_DET 個_CL 計劃_N
-> V (9039)
finish eating 吃完_V
finished it 就_ADV 完成_V
-> V v (3856)
finish my paper 完成_V 報告_V
have finished the exercises 已_ADV 地_DE 做完_V 練習_V
-> n V (3107)
should finish the work we 工作_N 我們_N 應_ADV 完成_V
finishing high school 中學_N 畢業_V

```
===============================================================

V p n (7774)
-> V n (1468)
care about the future 關心_V 未來_N
should care about China 應該_ADV 關心_V 中國_N
-> V (380)
care about them 在乎_V
talk about some 說_V 一些_DET
talk about it 討論_V 呢_SFP
-> V v (292)
care about the accumulation 顧_V 積累_V
talk about detente 講_V 緩和_V
-> V p n (169)
```

talk about the problem 討論_V 在_P 問題_N

went to food market 繼續_V 對_P 糧食_N 市場_N

come to my home 放學_V 到_P 我_N 家_N

-> p n V (130)

is excited about the awards ceremony 對_P 頒獎_Nv 儀式_N 非常_ADV 期待_V

rave about this book 對_P 這_DET 本_CL 書_N 讚不絕口_V

went to her 朝_P 她_N 走去_V

BONUS

- 產生文法規則實例（以搭配詞（如 *(rave, book)*）次數篩選）
- 產生全句實例（以 GDEX 原則篩選）