

選擇控制及句法學

台師大通識教育課程

文本分析與程式設計

授課：卓騰語言科技 _ PeterWolf

指代消解 (Coreference Resolution): Who is Who?

指代 (coreference) 為語言學中為了避免已經出現的字詞重複出現在文章句子上，導致語句結構過於贅述和語意不夠清晰，所以使用**代詞** (pronouns) 或是**普通名詞** (common nouns) 來代替已經出現過的字詞謂之。

在**計算語言學**中，**指代消解** (Coreference Resolution) 是一個很重要的議題，一般被應用於處理**資訊檢索**中的前處理部份，主要是找回原先被替換過的字詞，為了避免重要的字詞因指代的因素而造成權重計算降低的問題

實例說明：

舉例：

The handsome boy stared at Mary and said nothing: he seemed offended by her manner.

此處的"The handsome boy"與"he"意指為同一人；"Mary"與"her"意指為另同一人。

英文怎麼做：

#觀察資料

inputSTR = "The handsome boy stared at Mary and said nothing: he seemed offended by her manner. "

#流程思考：

#要被指代的對象是 "boy" 和 "Mary"

#找出 "boy" 的位置, 找出 "Mary" 的位置

#要消解的對象是 "he" 和 "her"

#找出 "he" 的位置, 找出 "her" 的位置

#如果 "he" 的位置在 "boy" 之後, 那麼 "he" 應該是指 "boy";

#如果 "her" 的位置在 "Mary" 之後, 那麼 "her" 應該是指 "Mary"

"boy" in inputSTR?

"Mary" in inputSTR?

str.index("boy") or str.find("boy")

str.index("Mary") or str.find("Mary")

str.index("he") or str.find("he")

str.index("her") or str.find("her")

if ("he" 的位置 > "boy"): he=> boy

if ("her" 的位置 > "Mary"): her=> Mary

英文怎麼做？再思考：

#觀察資料

inputSTR = "The handsome boy stared at John and said nothing: he seemed offended by his manner. "



#流程思考：

#要被指代的對象是 "boy" 和 "John"

#找出 "boy" 的位置, 找出 "John" 的位置

#要消解的對象是 "he" 和 "his"

#找出 "he" 的位置, 找出 "his" 的位置

#如果 "he" 的位置在 "boy" 之後, 那麼 "he" 應該是指 "boy";

#如果 "his" 的位置在 "John" 之後, 那麼 "his" 應該是指 "John"

#如果 "he" 的位置在 "John" 之後, 那麼 "he" 應該是指 "John";

#如果 "his" 的位置在 "boy" 之後, 那麼 "his" 應該是指 "boy"

"boy" in inputSTR?

"John" in inputSTR?

str.index("boy") or str.find("boy")
str.index("John") or str.find("John")

str.index("he") or str.find("he")
str.index("his") or str.find("his")

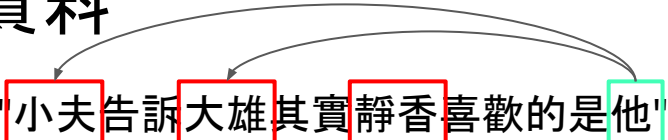
if ("he" 的位置 > "boy"): he=> boy
if ("his" 的位置 > "John"): his=> John

if ("he" 的位置 > "John"): he=> John
if ("his" 的位置 > "boy"): his=> boy

中文怎麼做：

#觀察資料

inputSTR = "小夫告訴大雄其實靜香喜歡的是他"



#流程思考：

#要被指代的對象是"小夫" 和 "大雄"

#找出 "小夫" 的位置, 找出 "大雄" 的位置

#要消解的對象是"他"

#找出 "他" 的位置

#如果 "他" 的位置在 "大雄" 之後, 那麼 "他" 應該是指 "大雄";

#如果 "他" 的位置在 "小夫" 之後, 那麼 "他" 應該是指 "小夫"

"小夫" in inputSTR?

"大雄" in inputSTR?

str.index("小夫") or str.find("小夫")

str.index("大雄") or str.find("大雄")

str.index("他") or str.find("他")

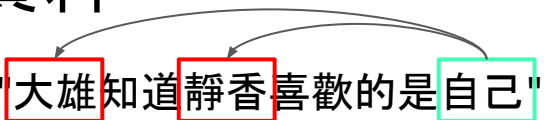
if ("他" 的位置 > "小夫"): 他 => 小夫

if ("他" 的位置 > "大雄"): 他 => 大雄

中文怎麼做，再思考：

#觀察資料

inputSTR = "大雄知道靜香喜歡的是自己"



#流程思考：

#要被指代的對象是"大雄" 和 "靜香"

#找出 "大雄" 的位置，找出 "靜香" 的位置

#要消解的對象是"自己"

#找出 "自己" 的位置

#如果 "自己" 的位置在 "大雄" 之後，那麼 "自己" 應該是指 "大雄"

#如果 "自己" 的位置在 "靜香" 之後，那麼 "自己" 應該是指 "靜香"

"大雄" in inputSTR?

"靜香" in inputSTR?

str.index("大雄") or str.find("大雄")

str.index("靜香") or str.find("靜香")

str.index("自己") or str.find("自己")

if ("自己" 的位置 > "大雄"): 自己 => 大雄

if ("自己" 的位置 > "靜香"): 自己 => 靜香

代名詞的指代消解好像沒什麼嘛！

前前後後線性的確認位置就好啦！





代誌不是憨人想得那麼簡單

imgflip.com

中文怎麼做，再思考：

#觀察資料

inputSTR01 = "大雄聽胖虎說靜香愛的是自己"

inputSTR02 = "大雄聽胖虎的妹妹說靜香愛的是自己"



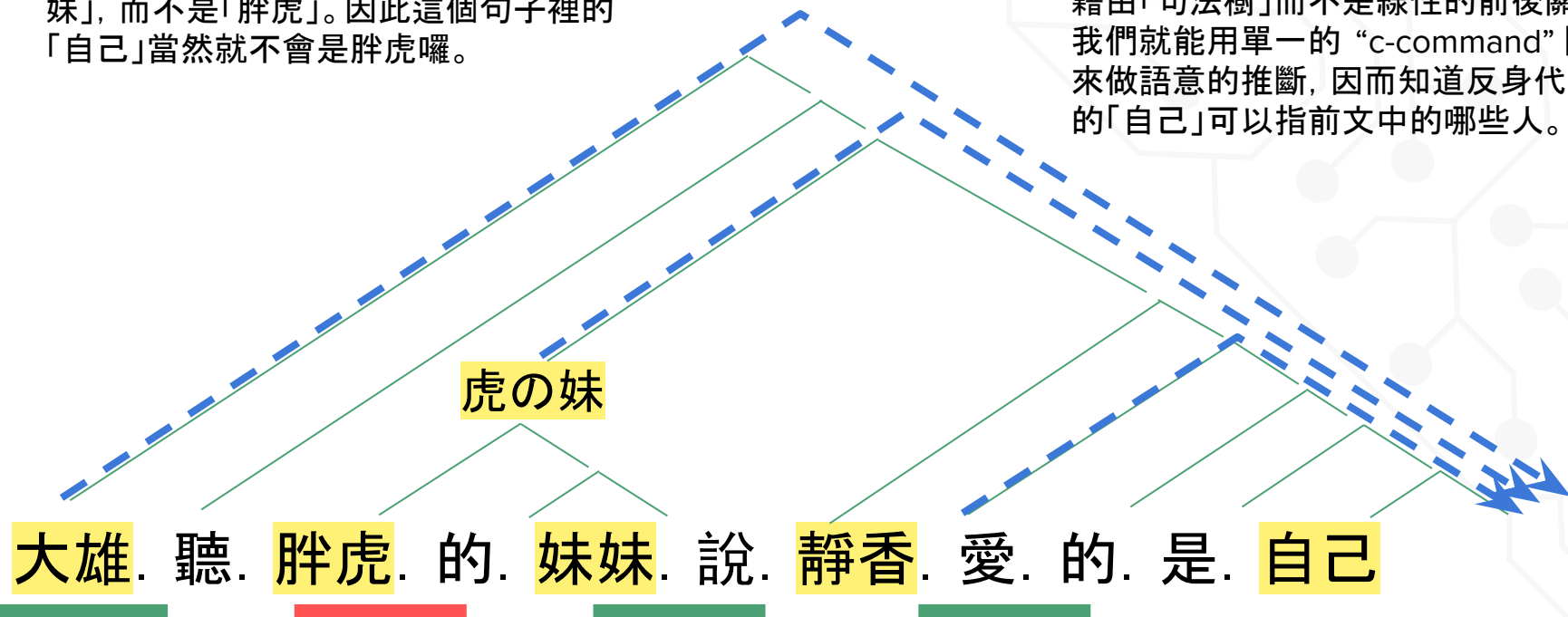
情況從糟糕變成難以理解

線性的前後關係是不足的，別忘了句法樹

C-Command 關係：

對中間這條藍虛線而言，要滿足「只能向上一個節點」的，就只有「胖虎的妹妹」，而不是「胖虎」。因此這個句子裡的「自己」當然就不會是胖虎囉。

藉由「句法樹」而不是線性的前後關係，我們就能用單一的“c-command”關係來做語意的推斷，因而知道反身代名詞的「自己」可以指前文中的哪些人。



Quiz:

指代消解的問題，課堂中做了英文和中文版的演示。試思考以下問題：

1. 英文的是否需要斷詞才能處理指代消解？
2. 中文是否需要先斷詞才能處理指代消解？
3. 線性方向的進行「文本處理 (或自然語言處理)」是否合適？
4. 英文的指代消解問題，下面兩句是否直接也符合 c-command 的演算法？
 - a. Jane heard that Mary's little sister said Debrah believes herself.
 - b. Jane heard that Mary's little sister said "Debrah believes herself."

Assignment:

1. 從課程 github repo 中把課程中提供的 week05 的目錄載下來。
2. 把 week05.py 改名為 **week05_你的學號.py**
3. 在 **week05_你的學號.py** 中，設計你的程式完成以下指定規格：
 - a. 和課堂中使用的 ZH03.py 相比，inputSTR 變成了 inputLIST
 - b. 請調整你的程式，讓它接受 inputLIST 的輸入，並得到和 ZH03.py 一樣的輸出結果。