

Visualizing the Sense of Language

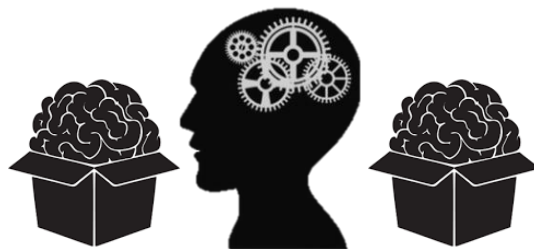
**Through analyzing the usage of
Mandarin approximative *chayidian* (差一點)**

The Problem : The Sense of Language is a Black Box



Input : Utterance

1. 逐句驗證曠日廢時
2. 能處理之語料有限



Process : Sense of Language

1. 語感運作無法被記錄
2. 判斷依據缺乏一致性

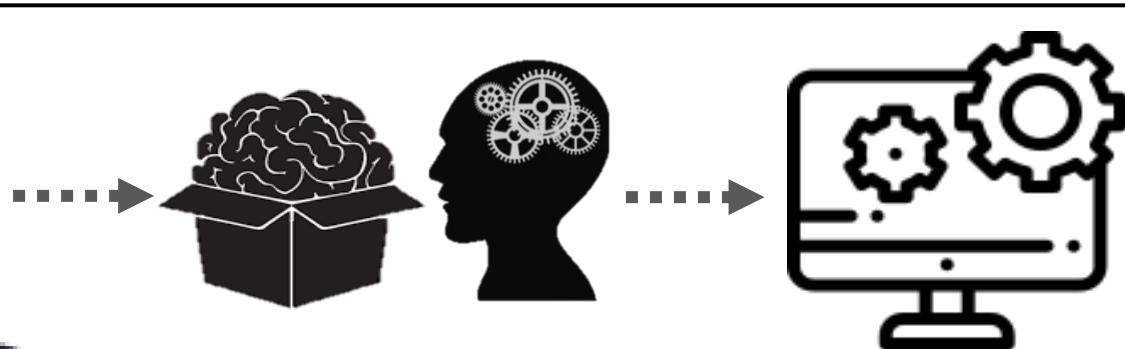


Output

1. 難以複製實驗步驟
2. 缺乏實際操作依據



What If...?



把語感實體化

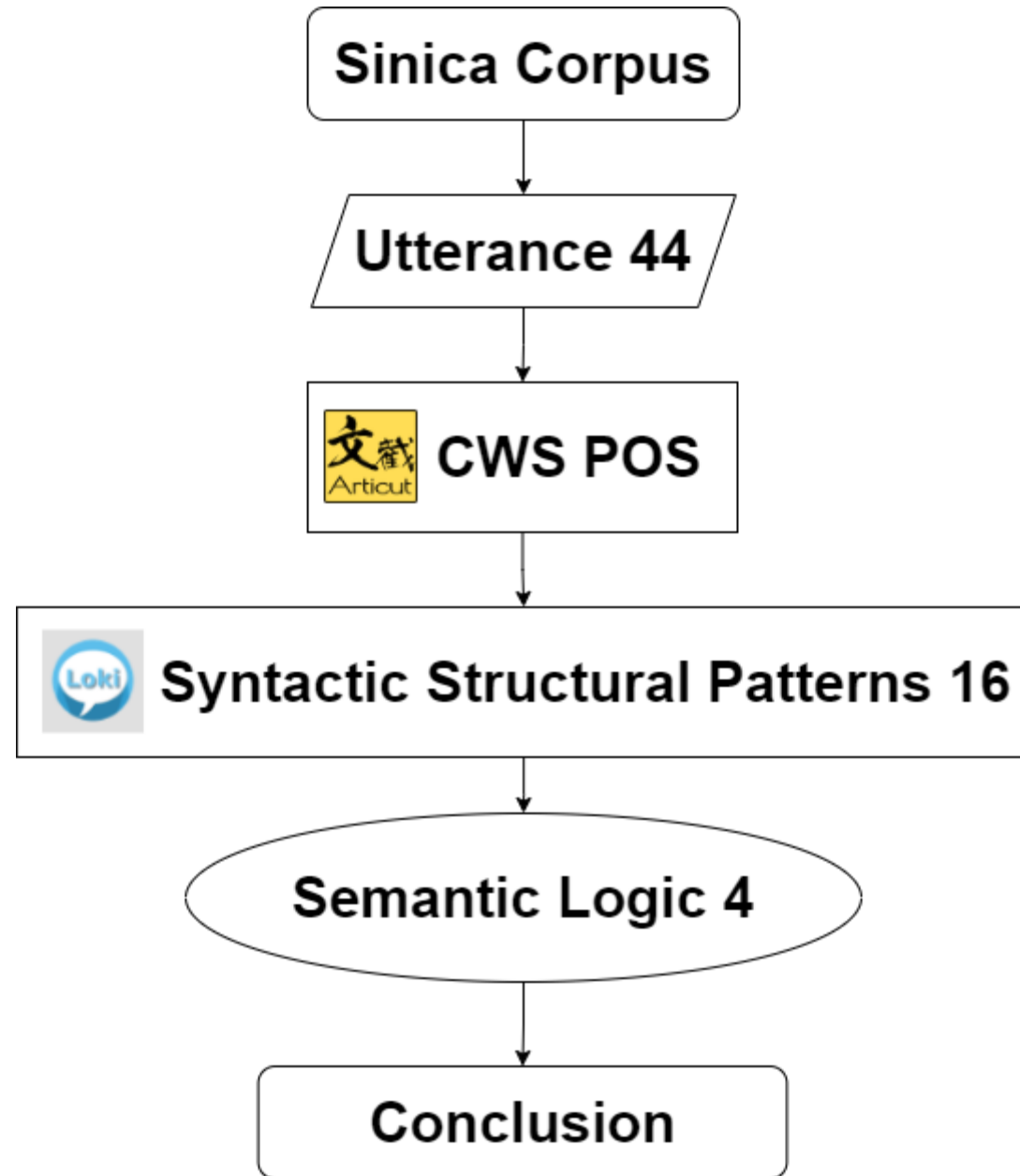
把判斷依據和過程變成程式碼

1. 電腦可以代為執行所有的驗證過程
2. 留下具一致性的步驟依據供人驗證



<https://github.com/Chenct-jonathan>





Sinica Corpus

Utterance 44



CWS POS

 Syntactic
Structural Patterns 16

Semantic Logic 4

Conclusion

平衡語料庫 Sinica Corpus Input Utterance (44)

E.g., 差一點昏倒

爭三連霸的瑞典名將艾柏格則**差一點**落馬
只**差一點**沒和那漂亮女人做成一回好事

差一點見上面

差一點忘了

可是我**差一點**被卡子抓走

我**差一點**被抓去坐牢

把經理**差一點**嚇昏了過去

我真的感動得眼淚都**差一點**掉下來了

差一點沒把手指頭當菜切了

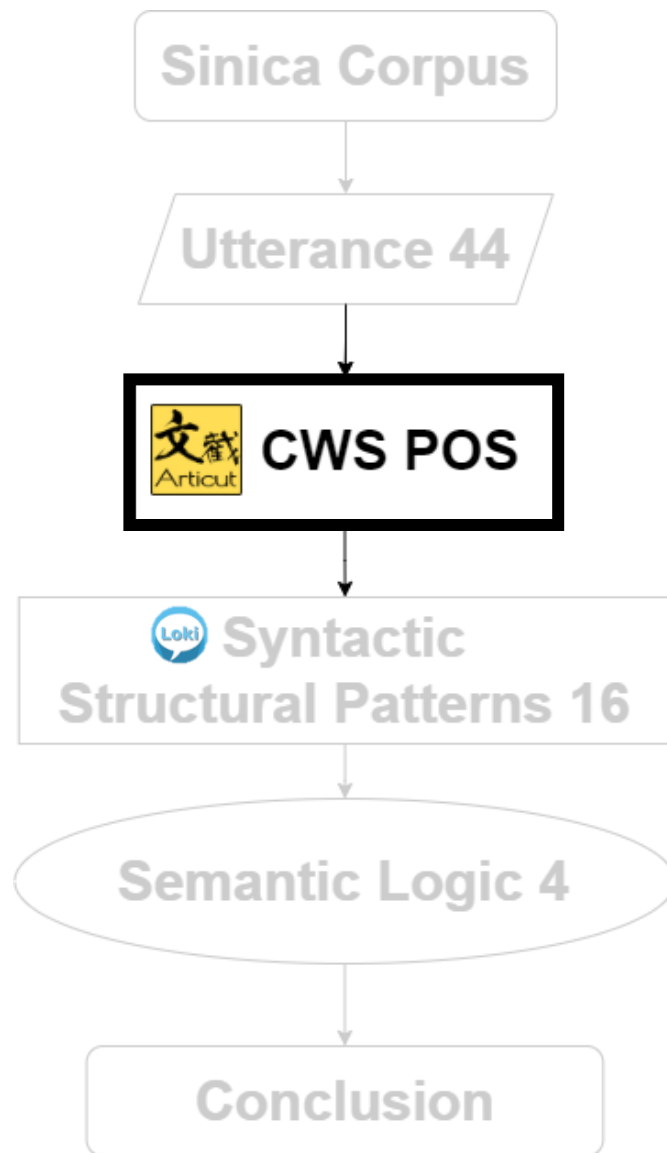
差一點把爸爸心愛的上等酒給打翻了

原來他**差一點**摔倒了

差一點遭到截肢

差一點就車畚斗

...

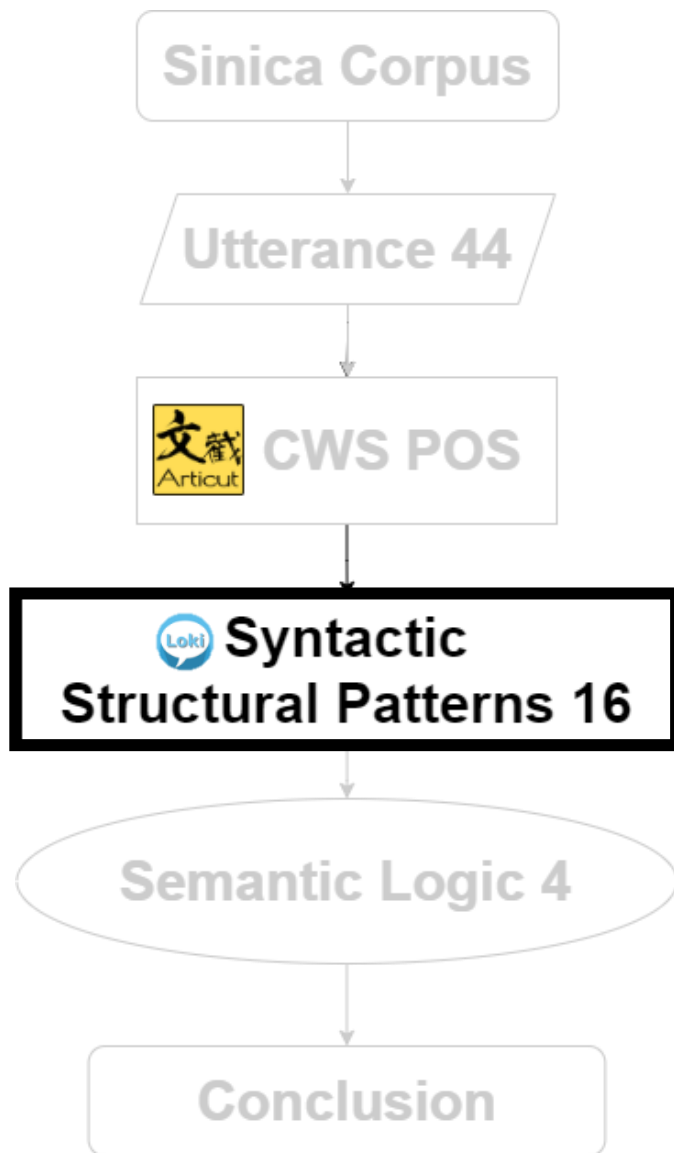


Articut

CWS POS

<MODIFIER>差一點</MODIFIER> <ACTION_verb>昏倒</ACTION_verb>

```
{
  "exec_time": 0.030290603637695312,
  "result_segmentation": "差一點/昏倒",
  "result_pos": [
    "<MODIFIER>差一點</MODIFIER><ACTION_verb>昏倒</ACTION_verb>"
  ],
  "result_obj": [
    [
      {
        "text": "差一點",
        "pos": "MODIFIER"
      },
      {
        "text": "昏倒",
        "pos": "ACTION_verb"
      }
    ]
  ],
}
```



Loki

Syntactic Structural Pattern

1. 差一點昏倒

<(MODIFIER|ModifierP)>[^<]*?</(MODIFIER|ModifierP)><(ACTION_verb|VerbP)>昏倒
</(ACTION_verb|VerbP)>

任意字詞 任意結構 任意時間 任意功能詞 任意範圍 任意名詞 任意修飾詞 AI 推薦動詞

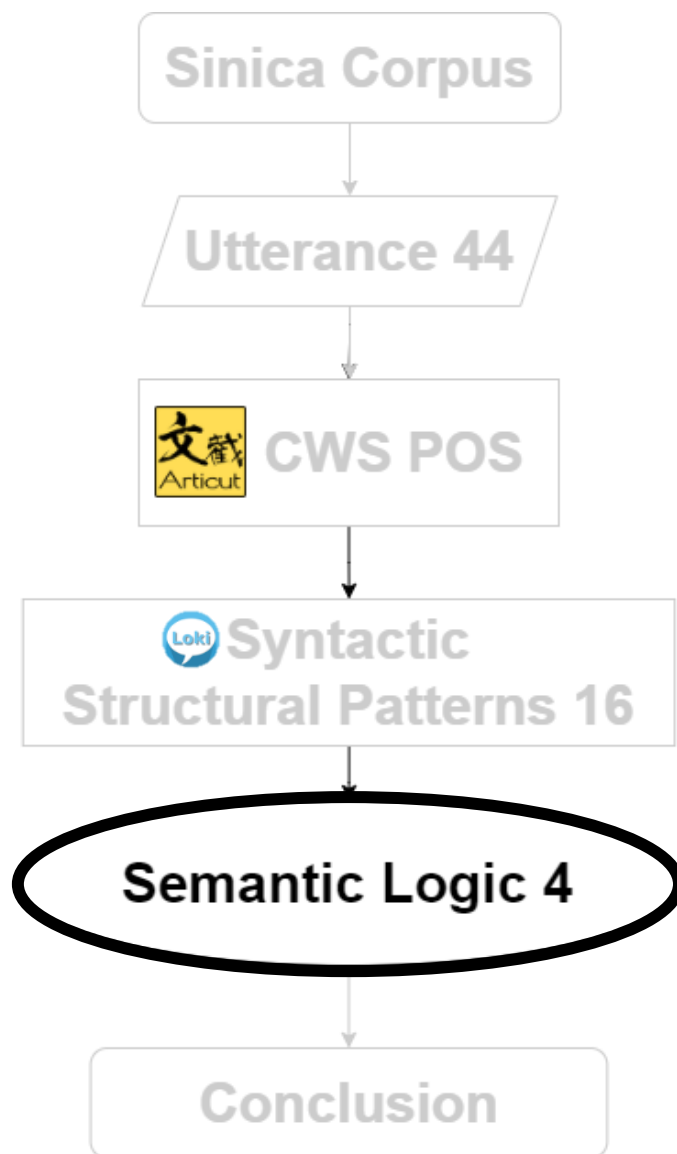
分析 檢驗

1. 差一點昏倒

<MODIFIER>差一點</MODIFIER><(ACTION_verb|VerbP)>[^<]*?[倒死]</(ACTION_verb|VerbP)>

任意字詞 任意結構 任意時間 任意功能詞 任意範圍 任意名詞 任意修飾詞 AI 推薦動詞

分析 檢驗



Generate Semantic Logic In Python

```
134 = .....if utterance == "差一點昏倒": # 只有符合此 pattern 的句子會進此 code block 處理
135 .....tmpInputSTR = inputSTRsplitter(inputSTR) # 只取 input 字串中從 [差一點] 開始的字
136 .....tmpPosSTR = formMSG(tmpInputSTR, pat)
137 = .....if re.search(pat, tmpPosSTR).group(5) == None: # 以位置為依據找尋第一個動詞
138 .....resultDICT["First Verb"] = re.search(pat, tmpPosSTR).group(6)
139 = .....if re.search(pat, tmpPosSTR).group(8) == None:
140 .....pass
141 = .....else:
142 .....resultDICT["First Verb"] = resultDICT["First Verb"] + re.search(pat, tmpPosSTR).group(8)
143 = .....elif re.search(pat, tmpPosSTR).group(6) == None:
144 .....resultDICT["First Verb"] = re.search(pat, tmpPosSTR).group(5)
```

```
145 = .....if resultDICT["First Verb"][-1] in ["了", "昏"]:
146 .....resultDICT["reason"] = "因為 [差一點] 後面的第一個動詞 [{}] 為完成貌事件(perfective)語意，故可使用 [差一點]。"
147 .....resultDICT["key"] = "完成貌(perfective)"
148 = .....else:
149 .....resultDICT["reason"] = "因為 [差一點] 後面的第一個動詞 [{}] 為結束體事件(accomplishment)語意，故可使用 [差一點]。"
150 .....resultDICT["key"] = "結束體(accomplishment)"
...
```


Findings and Demo

Grammatical Aspect	Lexical Aspect
完成貌(Perfective)	結束體(Accomplishment)
經驗貌(Experiential)	達成體(Achievement)

12:

[adv_sinica_chayidian] 原來他差一點摔倒了 ==> 差一點昏倒

說明：

因為「差一點」後面的第一個動詞「摔倒了」為完成貌事件(perfective)語意，故可使用「差一點」。

註：完成貌(perfective)帶有一終點。

13:

[adv_sinica_chayidian] 又曾經在泳渡深潭時不慎捲入旋渦差一點溺死 ==> 差一點昏倒

說明：

因為「差一點」後面的第一個動詞「溺死」為結束體事件(accomplishment)語意，故可使用「差一點」。

註：結束體(accomplishment)帶有一終點。

14:

[adv_sinica_chayidian] 卻真實地讓他以為差一點就跌入了萬丈深淵 ==> 差一點昏倒

說明：

因為「差一點」後面的第一個動詞「跌入了」為完成貌事件(perfective)語意，故可使用「差一點」。

註：完成貌(perfective)帶有一終點。



<https://github.com/Chenct-jonathan>

Conclusion



Just a second
Let me show you why