

# 程式人《十分鐘小論文》



## 人造交談語言

( 使用有 BNF 的口語透過機器翻譯和外國人溝通 )

陳鍾誠

2017 年 3 月 11 日

# 最近

- 人工智慧 AI 這個研究領域又開始受到強烈關注！

# 我記得上次 AI 受到強烈關注

- 大約是 20 年前的事情！

# 上上次

- 則是  $20+20$  年前的事情

# 人工智慧領域

- 似乎每 20 年就會走紅一次 …

# 而人工智慧領域裏

- 最難攻克的關卡，目前看來很可能是《自然語言》
- 像是《機器翻譯》之類的問題。

# 雖然 Siri 這類的交談程式

- 已經可以抓住一些人類語言的意義

# 甚至

- IBM 的 Watson 曾經在 Jeopardy 這個益智節目裏擊敗頂尖的人類對手

IBM's Watson Supercomputer Destroys Humans in Jeopardy | Engadget  
[https://www.youtube.com/watch?v=WFR3lOm\\_xhE](https://www.youtube.com/watch?v=WFR3lOm_xhE)



# 但是

- 我們目前仍然看到《機器翻譯》系統，在翻譯文章時通常很糟糕！

# 我們還很難

- 透過機器翻譯和外國人順利的交談。

# 特別是

- 像中文和英文這樣差異頗大的語言
- 目前的機器就是很難翻譯得夠好

# 莫非

- 這代表了我們非得要學英文才能和英美人士交談嗎？

# 難道我們

- 不能透過電腦的輔助，讓自己在不懂對方語言的情況下，和外國人自在的聊天嗎？

# 雖然

- 電腦還無法翻譯得很好

但是人腦可是理解高手啊？

# 人類的弱點

- 是很難記住大量單字

# 所以

- 我們從國中小就開始背單字



# 連續背了幾十年

- 單字量常常還是不超過一萬

但是背單字對電腦來說

- 完全不是問題...

# 要把一本字典讓電腦記住

- 只需要不到一秒

複製進去就行了！

# 而且

- 只要硬碟沒壞，電腦就不會忘掉！

# 透過這樣的想法

- 人類應該可以和電腦攜手合作。

# 由電腦來背單字

- 由人來理解，這樣就行了！

# 但是目前機器翻譯的作法

- 通常是全部交給電腦去做
- 希望電腦能夠給出正確翻譯
- 人只要閱讀翻譯結果就好了！

# 問題是

- 電腦常常無法給出很好的翻譯
- 甚至會給出非常誤導的翻譯結果



# 讓我們看幾個例子

- 就以 Google 翻譯為範本好了！

# 第一個例子是小學課本內容

英文 中文 日文 偵測語言

↔

中文(繁體) 英文 中文(簡體)

翻譯

小明有 5 個蘋果，給了小華 3 個蘋果，請問  
他還剩幾個蘋果？

小明和小華一起吃蘋果。

天這麼黑，風這麼大，爸爸捕魚去，為甚麼還  
不回家？

🔊 🎤 拼

75/5000

Xiao Ming has five apples, gave Xiaohua 3  
apples, I ask him left a few apples?

Xiao Ming and Xiaohua eat apples.

Days so black, the wind is so big, my father  
fishing, why not go home?

☆ 📄 🔊 🔗 ✎

Xiǎomíng yǒu 5 gè píngguǒ, gěile xiǎo huá 3 gè píngguǒ,  
qǐngwèn tā hái shèng jǐ gè píngguǒ?

Xiǎomíng hé xiǎo huá yīqǐ chī píngguǒ.

Tiān zhème hēi, fēng zhème dà, bàba bǔ yú qù,  
wéishénme hái bù huí jiā?

# 您可以看到

- Google 其實翻得還可以
- 但是有些明顯的錯誤！

# 錯誤的區域如下

英文 中文 日文 偵測語言



中文(繁體) 英文 中文(簡體)

翻譯

小明有 5 個蘋果，給了小華 3 個蘋果，請問  
他還剩幾個蘋果？

小明和小華一起吃蘋果。

天這麼黑，風這麼大，爸爸捕魚去，為甚麼還  
不回家？

🔊 🎤 拼 ▾

75/5000

Xiǎomíng yǒu 5 gè píngguǒ, gěile xiǎo huá 3 gè píngguǒ,  
qǐngwèn tā hái shèng jǐ gè píngguǒ?

Xiǎomíng hé xiǎo huá yīqǐ chī píngguǒ.

Tiān zhème hēi, fēng zhème dà, bàba bǔ yú qù,  
wéishènmé hái bù huí jiā?

Xiao Ming has five apples, gave Xiaohua 3  
apples, I ask him left a few apples?

Xiao Ming and Xiaohua eat apples.

Days so black, the wind is so big, my father  
fishing, why not go home?

☆ 📄 🔊 ➦



# 如果讓 Google 翻譯北風與太陽

中文 英文 日文 偵測語言

↔

英文 中文(繁體) 中文(簡體)

翻譯

The North Wind and the Sun were disputing which was the stronger, when a traveler came along wrapped in a warm cloak. They agreed that the one who first succeeded in making the traveler take his cloak off should be considered stronger than the other.  
Then the North Wind blew as hard as he could, but the more he blew the more closely did the traveler fold his cloak around him; and at last the North Wind gave up the attempt. Then the Sun shined out warmly, and immediately the traveler took off his cloak.  
And so the North Wind was obliged to confess that the Sun was the stronger of the two.

594/5000

北風和太陽正在爭論，這是更強大的，當一個旅客來包裹在一件溫暖的斗篷。他們同意，首先成功地使旅行者把他的斗篷關閉的人應該被認為比另一個更強。然後，北風吹得盡可能的硬，但他越是打得越緊密，旅行者折疊他的斗篷在他周圍；最後北風放棄了嘗試。然後太陽熱烈地發光，立即旅行者脫下他的斗篷。所以北風不得不承認太陽是兩者中更強的。

☆ □ 🔊 ↻

Běifēng hé tàiyáng zhèngzài zhēnglùn, zhè shì gèng qiángdà de, dāng yīgè lǚkè lái bāoguǒ zài yī jiàn wēnnuǎn de dǒupéng.  
Tāmen tóngyì, shǒuxiān chénggōng dì shǐ lǚxíng zhě bǎ tā de dǒupéng guānbì de rén yīnggāi bèi rènwéi bǐ lìng yīgè gèng qiáng.  
Ránhòu, běifēngchuī de jǐn kěnéng de yìng, dàn tā yuè shì dǎ de yuè jǐnmì, lǚxíng zhě zhédié tā de dǒupéng zài

相信您會覺得還是很難懂

# 為了解決這個問題

- 我做了一些嘗試！

# 第一個嘗試是

- 只進行逐字翻譯，不改變字詞順序，而且可以透過網頁的 ruby tag 上下對齊。



# 以下是該系統翻譯北風與太陽的結果

● 中翻英 ● 英翻中

The North Wind and the Sun were disputing which was the stronger, when a traveler came along wrapped in a warm cloak.

They agreed that the one who first succeeded in making the traveler take his cloak off should be considered stronger than the other.

Then the North Wind blew as hard as he could, but the more he blew the more closely did the traveler fold his cloak around him;

and at last the North Wind gave up the attempt. Then the Sun shined out warmly, and immediately the traveler took off his cloak.

And so the North Wind was obliged to confess that the Sun was the stronger of the two.

the 北風和the太陽were爭議ing哪個was the 強er  
The North Wind and the Sun were disputing which was the stronger  
何時a旅客到來ed沿包裹in a溫暖斗篷.  
when a traveler came along wrapped in a warm cloak.

他們同意that the 一誰第一得手in making the旅客  
They agreed that the one who first succeeded in making the traveler  
拿取他的斗篷off應該be考慮強er than the其他.  
take his cloak off should be considered stronger than the other.

然後the北風吹ed as硬as he可以, but the更多he吹ed  
Then the North Wind blew as hard as he could, but the more he blew  
the更多密切did the旅客折他的斗篷周圍him;  
the more closely did the traveler fold his cloak around him;

和at最後the北風給ed up the嘗試. 然後the太陽  
and at last the North Wind gave up the attempt. Then the Sun  
閃耀d out溫暖ly, 和馬上the旅客拿取ed off他的  
shined out warmly, and immediately the traveler took off his  
斗篷.  
cloak.

和so the北風was有責任to承認that the太陽was the  
And so the North Wind was obliged to confess that the Sun was the  
強er of the 二.  
stronger of the two.

# 您會注意到

- 我們保留了很  
多詞沒翻譯  
  
(通常是介系詞或  
難以翻譯的詞)

the 北 風 和 the 太陽 were 爭議ing 哪個 was the 強 er  
The North Wind and the Sun were disputing which was the stronger  
, 何時 a 旅客 到來ed 沿 包裹 in a 溫暖 斗篷 .  
, when a traveler came along wrapped in a warm cloak .

他們 同意 that the 一 誰 第一 得手 in making the 旅客  
They agreed that the one who first succeeded in making the traveler  
拿取 他的 斗篷 off 應該 be 考慮 強 er than the 其他 .  
take his cloak off should be considered stronger than the other .

然後 the 北 風 吹ed as 硬 as 他 可以 , but the 更多 他 吹ed  
Then the North Wind blew as hard as he could , but the more he blew  
the 更多 密切 did the 旅客 折 他的 斗篷 周圍 him;  
the more closely did the traveler fold his cloak around him;

和 at 最後 the 北 風 給ed up the 嘗試 . 然後 the 太陽  
and at last the North Wind gave up the attempt . Then the Sun  
閃耀d out 溫暖ly , 和 馬 上 the 旅客 拿取ed off 他的  
shined out warmly , and immediately the traveler took off his  
斗篷 .  
cloak .

和 so the 北 風 was 有責任 to 承認 that the 太陽 was the  
And so the North Wind was obliged to confess that the Sun was the  
強 er of the 二 .  
stronger of the two .

# 而且也有很多翻錯的情況

xiao ming have 5=? 個 mac , give 了 little wah 3=? 個 mac ,  
小明 有 5 個蘋果 , 給了 小 華 3 個蘋果 ,  
excuse me he still leftover several mac ?  
請 問 他 還 剩 幾個蘋果 ?

xiao ming 和 little wah together eat mac 。  
小明 和 小 華 一 起 吃 蘋果 。

day this what pinot , wind this what big , papa fish going , 為  
天 這 麼 黑 , 風 這 麼 大 , 爸爸 捕魚 去 , 為  
what still 不 go home ?  
甚麼 還 不 回 家 ?

# 如果要翻得更好

- 應該要有更好的字典
- 而且要能從多個候選字中選取最好的一個！

# 不過這種英漢逐字對照的方法

- 可以不考慮語句結構調整的問題

因為調整後就無法《逐字對照》了！

# 從這種

- 不改變詞彙結構的翻譯方式出發

我們還能做些甚麼呢？

# 我想到

- 人類雖然記不住單字
- 但是卻可以學會程式語法

程式語言具有明確嚴格的 BNF 語法

# 那麼

- 我們能不能設計一種 BNF 語法
- 不是用來描述程式語言
- 而是描述《用來說的口語》呢？



# 如果這種口語的 BNF 很簡單

- 簡單到我們可以在幾分鐘之內學會
- 那麼借助機器的逐字翻譯，我們有可能在學會語法後和外國人溝通嗎？

# 於是

- 我開始設計這樣的語法

# 並且努力的簡化語法

# 最後我發現

- 似乎可以用很簡單的規則
- 就設計出一套《口語 BNF》

規則有多簡單呢？

# 出乎意料的

- 我發現似乎只需要三條規則

# 這三條規則如下

- $S = Q? NP^* VP^* .$

句子 = 疑問詞 ? 名詞短語 \* 動詞短語 \* 句尾號

- $VP = VP (c VP)^* \\ | (v^* V+ v^*)^+ NP^*$

動詞短語 = 動詞短語 ( 連接詞 動詞短語 ) \*  
| ( 動詞修飾 \* 動詞 + 動詞修飾 \* ) + 名詞短語 \*

- $NP = NP (c NP)^* \\ | (n^* N+)$

名詞短語 = 名詞短語 ( 連接詞 名詞短語 ) \*  
| ( 名詞修飾 \* 名詞 + )

# 如果改成對照寫法或許會更清楚

- $S = Q? \quad NP^* \quad VP^* \quad .$   
句子 = 疑問詞 ?      名詞短語 \*      動詞短語 \*      句尾號

- $VP = VP \quad (c \quad VP)^* \quad | \quad (v^* \quad V+ \quad v^*)^+ \quad NP^*$   
動詞短語 = 動詞短語 ( 連接詞 動詞短語 ) \*    |    ( 動詞修飾 \* 動詞 + 動詞修飾 \* ) +    名詞短語 \*

- $NP = NP \quad (c \quad NP)^* \quad | \quad (n^* \quad N+)$   
名詞短語 = 名詞短語 ( 連接詞 名詞短語 ) \*    |    ( 名詞修飾 \* 名詞 + )



# 然後、我根據這些規則 寫了一個有 BNF 的翻譯程式，只有 130 行

```
https://mdbookspace.com/view/ai/mtBnfSpoken.md ☆
巴哈姆特電玩資訊站 B 弑天劍仙 - 143.第143 首頁 :: 英雄聯盟基地 L 嗶姆播台 - 光暈戰記- 从前有座灵剑山 第二 其他

有 BNF 的口語翻譯系統

程式：BnfSpokenMt.js

// S = Q? NP* (VP)* .
// VP = VP (c VP)* | (v* V+ v*)+ NP*
// NP = NP (c NP)* | (n* N+)

// 小明 有 5 個 蘋果 ， 給 了 小華 3 個 蘋果 ， 請問 他 還 剩 幾 個 蘋果 ？
// NP   V   n   NP       V   v   NP   n   NP       Q   NP   v   V   n   NP
//      VP                VP                        VP

var c = console;

var tagMap={
  N : ["小明", "小華", "小莉", "大雄", "爸爸", "魚", "天", "家", "風", "日", "旅人", "人", "衣", "他", "你", "我", "蘋果", "橘子", "柳丁", "番茄"],
  V : ["有", "給", "剩", "吃", "去", "捕", "回", "至", "爭", "勝", "吹", "抓", "繫", "敗", "勝", "還", "了", "一起", "不"],
  v : ["還", "了", "一起", "不"],
  c : ["又", "和", "且", "或", "與"],
  n : ["幾", "這", "那", "5", "3", "個", "隻", "條", "黑", "大", "的"],
  Q : ["請問", "為甚麼"],
  "." : [",", " ", "?", "。"]
}
```

程式位於：<https://mdbookspace.com/view/ai/mtBnfSpoken.md>

# 其語法剖析器程式碼如下

```
// S = Q? NP* (VP)* .  
function S() {  
    if (isTag("Q"))  
        next("Q");  
  
    while (isTag("n") || isTag("N"))  
        NP();  
  
    while (isTag("V") || isTag("v")) {  
        VP();  
    }  
    next(".");  
}
```

```
// VP = VP (c VP)* | (v* V+ v*)+ NP*  
function VP() {  
    while (isTag("v") || isTag("V")) {  
        while (isTag("v")) next("v");  
        do { next("V") } while (isTag("V"));  
        while (isTag("v")) next("v");  
    }  
    while (isTag("n") || isTag("N"))  
        NP();  
}
```

```
// NP = NP (c NP)* | (n* N+)  
function NP() {  
    while (isTag("n")) next("n");  
    do { next("N") } while (isTag("N"));  
    if (isTag("c")) {  
        next("c");  
        NP();  
    }  
}
```

# 該程式可以進行 語法驗證和翻譯動作

```
function parse(sentence) {  
  words=sentence.split(" ");  
  wi = 0, mtWords=[];  
  c.log("中文:", words.join(" "));  
  T(words);  
  c.log("英文:", mtWords.join(" "));  
}
```

```
parse("小明 有 5 個 蘋果， 給了 小華 3 個 蘋果， 請問 他 還 剩 幾 個 蘋果？");
```

```
parse("小明 和 小華 一起 吃 蘋果。");
```

```
parse("黑 黑 的 天， 大 大 的 風， 爸爸 去 捕 魚， 為甚麼 還 不 回 家？");
```

```
// 全文: https://www.facebook.com/photo.php?fbid=1464494203561879&set=p.1464494203561879&type=3&\_\_ftid=1464494203561879
```

```
// parse("聽 狂 風 怒 吼， 真 叫 我們 害怕。 爸爸! 爸爸! 我們 心理 多麼 牽掛， 只要 您 早點 兒
```

```
// parse("我的 好 孩子， 爸爸 回來 啦！ 你看 船艙 裡， 裝 滿 魚 和 蝦， 努力 就 有 好 收穫
```

```
parse("風 與 日。 風 日 爭， 旅 人 至， 脫 者 勝， 風 狂 吹， 人 緊 衣， 風 敗， 日 暖 照，
```

# 但是

- 語句有時必須要調整一下  
才能符合那三條 BNF 語法

$$S = Q? NP^* VP^* .$$
$$VP = VP (c VP)^* \mid (v^* V^+ v^*)^+ NP^*$$
$$NP = NP (c NP)^* \mid (n^* N^+)$$

# 例如

- 天這麼黑，風這麼大，爸爸捕魚去，為甚麼還不回家？



- 黑黑的天，大大的風，爸爸去捕魚，為甚麼還不回家？

# 為甚麼要這樣調整呢？

- 天這麼黑，風這麼大，爸爸捕魚去，為甚麼還不回家？



- 黑黑的天，大大的風，爸爸去捕魚，為甚麼還不回家？

# 那是因為要符合 BNF 語法

- 天這麼黑，風這麼大，爸爸捕魚去，為甚麼還不回家？



- 黑黑的天，大大的風，爸爸去捕魚，為甚麼還不回家？

n n n N

NP

S

n n n N

NP

S

N

NP

S

V V N

VP

Q

v v V N

VP

S

# 然後我們就可以 得到下列的翻譯結果

```
===== 中文 =====
黑 黑 的 天 ， 大 大 的 風 ， 爸 爸 去 捕 魚 ， 為 甚 麼 還 不 回 家 ？
===== 剖析 =====
黑:n 黑:n 的:n 天:N ，:..
大:n 大:n 的:n 風:N ，:..
爸爸:N 去:U 捕:U 魚:N ，:..
為甚麼:Q 還:v 不:v 回:U 家:N ?:..
===== 英文 =====
black black _de sky ， big big _de wind ， Papa go hunt fish ， why still not ba
ck home ？
```



# 當然也可以檢驗翻譯其他語句

```
D:\js>node mtBnfSpoken.js
```

```
===== 中文 =====
```

```
小明 有 5 個 蘋果， 給了 小華 3 個 蘋果， 請問 他 還 剩 幾 個 蘋果？
```

```
===== 剖析 =====
```

```
小明:N 有:U 5:n 個:n 蘋果:N，:..
```

```
給:U 了:v 小華:N 3:n 個:n 蘋果:N，:..
```

```
請問:Q 他:N 還:v 剩:U 幾:n 個:n 蘋果:N？:..
```

```
===== 英文 =====
```

```
ShaoMin have 5 _ge apple， give _le ShaoHua 3 _ge apple， Q he still own _gi _ge apple？
```

```
===== 中文 =====
```

```
小明 和 小華 一起 吃 蘋果。
```

```
===== 剖析 =====
```

```
小明:N 和:c 小華:N 一起:v 吃:U 蘋果:N。:..
```

```
===== 英文 =====
```

```
ShaoMin and ShaoHua together eat apple。
```

# 或許也可以試試古文

===== 中文 =====

風與日。風日爭，旅人至，脫者勝，風狂吹，人緊衣，風敗，日暖照，人脫衣，日勝。

===== 剖析 =====

風:N 與:c 日:N 。:.

風:N 日:N 爭:V ，:.

旅人:N 至:V ，:.

脫:V 者:N 勝:V ，:.

風:N 狂:V 吹:V ，:.

人:N 緊:V 衣:N ，:.

風:N 敗:V ，:.

日:N 暖:V 照:V ，:.

人:N 脫:V 衣:N ，:.

日:N 勝:V 。:.

===== 英文 =====

wind and3 sun 。 wind sun argue ， travler come ， take\_off guy win ， wind wild  
blow ， people tighten cloth ， wind lose ， sun warm shine ， people take\_off  
cloth ， sun win 。

# 這樣

- 我們就發明了一個具有 BNF 的口語
- 而且這種口語應該很容易檢驗與翻譯

# 只是我們不知道

- 是否人類想表達的語句
- 都可以改寫成符合這三條語法的句子

# 我們也不知道

- 人們是否會想用這種人造口語和外國人溝通！

===== 中文 =====

小明 有 5 個 蘋果， 給 了 小華 3 個 蘋果， 請問 他 還 剩 幾 個 蘋果？

===== 剖析 =====

小明:N 有:V 5:n 個:n 蘋果:N，:..

給:V 了:v 小華:N 3:n 個:n 蘋果:N，:..

請問:Q 他:N 還:v 剩:V 幾:n 個:n 蘋果:N？:..

===== 英文 =====

ShaoMin have 5 \_ge apple， give \_le ShaoHua 3 \_ge apple， Q he still own \_gi \_ge apple？

===== 中文 =====

小明 和 小華 一起 吃 蘋果。

===== 剖析 =====

小明:N 和:c 小華:N 一起:v 吃:V 蘋果:N。:..

===== 英文 =====

ShaoMin **and** ShaoHua together eat apple。



# 還有外國人

- 是否能在簡易的訓練之後看懂翻譯！

===== 中文 =====

黑 黑 的 天 ， 大 大 的 風 ， 爸 爸 去 捕 魚 ， 為 甚 麼 還 不 回 家 ？

===== 剖析 =====

黑:n 黑:n 的:n 天:N ，:..

大:n 大:n 的:n 風:N ，:..

爸爸:N 去:V 捕:V 魚:N ，:..

為甚麼:Q 還:v 不:v 回:V 家:N ？:..

===== 英文 =====

black black \_de sky ， big big \_de wind ， Papa go hunt fish ， why still **not** back home ？

# 但是我們知道

- 電腦要翻譯這種有 BNF 的口語
- 會比翻譯自然語言容易得多
- 而且語意也會更精確！

# 如果能編一本

- 每個英文詞都只對應一個  
漢語詞彙的雙語詞典



# 那麼

- 整個語言就會變得非常精確
- 不會再有任何歧義！

但這會是一個工程

我想把這樣的技術商品化

# 並且設計出

- 讓人可以不需要背單字，就可以和外國人溝通的軟體

# 如果您有興趣

- 把這樣的技術商品化！

請您跟我聯絡！

# 我是陳鍾誠

- 我在金門大學資訊工程系教書

# 我的聯絡方式為

- Email: ccckmit@gmail.com
- Facebook: <https://www.facebook.com/ccckmit>
- 手機號碼：0988068390  
( 國際碼 :886-988068390)



# 由於我是大學老師

- 我不能開公司，  
不能當董監事，  
不能做生意！

# 但是

- 我還是希望能完成這件商品！

# 這除了需要程式

- 還需要一些人工（編字典等等）
- 以及有能力進行商品化的公司參與！

# 歡迎您

- 和我一起來完成這樣一個翻譯軟體
- 並且一起經營讓人們可以不需要學外語就能相互溝通的社群。

# 我相信

- 科技始終來自於惰性...

有了這樣的工具

# 或許

- 在未來的某一天

我們就可以



不需要再學英文

也能和外國人溝通了

只要

我們記住

# 下列的三條語法

$S = Q? NP^* VP^* .$

$VP = VP (c VP)^* \mid (v^* V^+ v^*)^+ NP^*$

$NP = NP (c NP)^* \mid (n^* N^+)$

然後

剩下的

就交給機器了！



希望

那一天

能趕快到來！

讓我們一起攜手

# 創造一個翻譯工具

讓那一天早點到來吧！

這就是我們今天的

# 十分鐘小論文



希望您會喜歡！

我們下回見！

Bye Bye!