

---

本章內容：

\* yacc 解析器

# 出版序

把這本書當作電腦工具書來看，其實只對了一半；可能一半都不到。

## 名稱

UNIX 工具程式的名稱都大有來頭，和 DOS/Windows 平台下的直覺式（且商業化）的程式名稱對照一下，不難發現這是兩種哲學下的產物。fdisk 就是 format disk（規劃磁碟），沒有比這個更簡單的；awk 呢？Aho、Weisberger 和 Kernighan 三位大師！不盡然，很多人說它代表 awkward（笨手笨腳的）；Expect 是會期望（expect）對方輸入的工具；shell（殼層）是包在核心（kernel）的外層，而非位於低階系統的上層；Tcl 唸起來是 tickle（搔癢），因為它只能碰觸到系統的表面；cat 是 concatenate（一隻貓的尾巴接上另一隻貓的鼻子；依此類推）。UNIX 顯然偏向意象化的命名，DOS/Windows 則是極端的具體化。

## 翻譯

在拉丁文的字根裡面，lex 代表的是法律（law），經過兩千年的演化，在這裡變成了 lexical（關於詞彙）；重點在於：怎樣讓電腦明白你在說什麼？就理解力而言，print 和 shit 這兩個字對電腦是完全沒有兩樣的。lexer 似乎可以譯為詞彙解析器，但“解析”這個字眼值得商榷：一般流行的字是 parse，或是 analyse。從拉丁文字根來看，parse 其實是“拆解”，而 analyse 是帶有某些鑽研色情意味的動詞（如果你很喜歡 analyse 的話）；硬把 parse 譯作“解析”，analyse 譯成“分析”，這種程度很適合大學聯考。

至於要 parse（或是 analyse）的對象，不外 syntax、semantics、grammar 和 pattern；組成這些東西的元素大概是 token、symbol，或是乾脆一點：lex。yacc 可以判斷你的 parse 方式有沒有問題，如果它不爽，你就要 yuk 了！假使把 grammar 理所當然的譯為“文法”，那就理所當然的毫不合理；syn 意味著“同化”，syntax 是“語法”，譯為“句法”似乎更貼近原意；如果你的興趣在於 semantics（語意學），應該去看看 Aho 的 compiler 天龍寶典，“lex & yacc”只會讓你日子好過一點。

如果 token 是“符號”，那 symbol 又是什麼？或許之間只有一些具象化程度上的差距。幾個 token 湊起來就成了 expression，雖然裡面一堆橫來豎去的牙籤，但如果這是大家都看得懂的玩意兒，表示它是個 regular expression，雖然它一點也不 regular。好吧！只好用“常規表示法”勉強湊合一下。Regular expression 必須和某些 pattern 比對，如果中獎（比對符合）了就採取某些行動，這個 pattern 稀奇古怪，姑且譯為“樣式”，反正它本質上和壁紙的圖案沒什麼差別。

如果讀者懷疑我們何必拿瑣碎的名詞大作文章，請翻到封面，上面寫著兩個斗大的字：“lex & yacc”；如果您不是這麼 lexical，就不會想 parse；既然想 parse，就必須 analyse；analyse 的對象就是這堆一般想當然爾的瑣碎東西。如果您 match 這種預設類型的 lexical pattern，而且 exclude 某些 cynical tokens，自然能體會我們的 expression。

在技術翻譯的領域中碰到的困難，“lex & yacc”是相當具有代表性的（註一）。倒不是因為它的專有名詞最多，而是它具備了各種翻譯名詞分歧性（ambiguity）的典型。和電腦技術本身的距離是：翻譯不是精密科學，它是一種經驗科學（註二）；翻譯者的工作，是儘可能的縮短這個距離。

## 後記

我綜合一下網路上對本書提出的意見：

優點：

1. 討論了許多通用的版本。
2. 整章的篇幅特別討論 SQL。
3. 可以建立一些專屬語言的 compiler。

缺點：

1. 缺少直接對 lex 和 yacc 兩者間內部溝通的討論。
2. 未涵蓋與 C 程式的介面。
3. 沒有提到 Windows 版本下的 PCY ACC。
4. MS-DOS 版本的 yacc 範例程式編譯有問題。

---

註一 不過，還是比“Programming Perl”簡單一點。“lex & yacc”是在研究文字的搬弄，“Programming Perl”則不少是在搬弄文字。

註二 最近看到一本書把 Interrupt Vector（中斷向量）譯為“干擾因子”（八成是把 Interrupt 當成 Interfere，vector 當作 factor）。可惜譯者並未詳述其玄妙之理念。據此推論，IVT（Interrupt Vector Table）或許可譯為“干擾因子的桌子”。

