

音節

Im-chiat

*Syllable*

羅馬字編碼：*Unicode*

Lô-má-jī pian-bé：*Unicode*

有的羅馬字在 *Unicode* 裡，由兩個  
或三個符號組合而成

Ū-ê Lô-má-jī tī *Unicode* lâi-tóe, sī  
iōng n̄ng-ê iah-sī sa<sup>n</sup>-ê hū-hō  
tàu--khí-lâi ê.

*Unicode* 完全支援台語羅馬字  
(2004)之前，有其它編碼

*Unicode* ôan-chôan chi-ôan Tâi-gú  
Lô-má-jī (2004) chìn-chêng, ū  
kî-tha ê pian-bé.

羅馬字輸入工具：

*Taiwanese Package*

*<http://www.phahng.idv.tw>*

備站 *<http://taigi.fhl.net/TP/>*

Lô-má-jī su-jīp kang-khū :

*Taiwanese Package*

爲了不同編碼的交換需求，採用標準 *ASCII* 爲中間格式

Ūi-tiòh bô kâng pian-bé ê kau-ōa<sup>n</sup>  
su-kiû, chhái-iōng phiau-chún ê  
*ASCII* chòe tiong-ng keh-sek.

聲調用數字(第 1、4 聲音無調符，  
數字可省略)放在音節最後面

Sia<sup>n</sup>-tiāu ēng sò-jī (tē-1/4-sia<sup>n</sup> bô  
tiāu-hû, sò-jī ē-tàng kán-séng) khng  
tī im-chiat siōng āu-piah.

a	á	à	ah	â	ā	àh
<i>a(1)</i>	<i>a2</i>	<i>a3</i>	<i>ah(4)</i>	<i>a5</i>	<i>a7</i>	<i>ah8</i>



鼻音“n”用“*N*” 或 “*nn*”

Phī<sup>n</sup>-im ēng “*N*” iah-sī “*nn*”

phī <sup>n</sup>
<i>phiN7 / phinn7</i>

後元音“o”用 “*ou*”

Āu-gôan-im “o” ēng “*ou*”

hō·lō·
<i>hou5-lou5</i>

台語音節數：2,500~3,000

Tâi-gú im-chiat sò : 2,500~3,000

比較：華語大概 1,200

Pí-kàu : Hôa-gú t̄ai-khài 1,200

工具程式 1：台語羅馬字拼字檢查

Kang-khū thêng-sek 1 : Tâi-gú lô-má-jī  
pheng-jī kiám-cha

輸入：羅馬字音節，輸出：*T/F*

Su-jip : Lô-má-jī im-chiat, su-chhut :  
*T/F*

eg : goa2 (O)      gie8 (X)

方法 1 : *Finite state machine* (稍微  
囉唆 😞 )

Hong-hoat 1 : *Finite state machine*  
(sió-khóa lo-so 😞 )

方法 2 : 查表(約 3,000 個 entries)

Hong-hoat 2 : chhâ pió (tāi-khài ū  
3,000 ê entries)

方法 3：把聲母、韻母、聲調反向

拆開檢查(**table size**小)

Hong-hoat 3 : kā sia<sup>n</sup>-bú, ūn-bú,

sia<sup>n</sup>-tiāu tò-thâu thiah-khui

kiám-cha (**table size** khah sòe)



反向的原因：台語有零聲母的音節，m/ng同時可當聲母(ma/ ngeh)及韻母(m7/ng5)，左到右讀的時候無法判別，右到左讀(聲調已先處理)則可以。

Tò-thâu ê gôan-in : Tâi-gú ū  
lêng-sia<sup>n</sup>-bú ê im-chiat, m/ng  
tông-sî ē-tàng tòng-chòe sia<sup>n</sup>-bú  
(ma/nga) kah ūn-bú (m/ng), ùi  
tò-pêng thak bô-hoat-tō phò<sup>n</sup>-tōan,  
ùì chià<sup>n</sup>-pêng thak (sia<sup>n</sup>-tiāu í-keng  
seng chhú-lí a) tō ē-sái.

# Program(ASP):

```
Function SiPOJimchiat(pojich)
```

```
    tngtou = Len(pojich)
```

```
    pojtmp = pojich
```

```
    ' Te 1 e Char kai cho sio-sia
```

```
    ch = Left(pojtmp, 1)
```

```
    If ch >="A" and ch <= "Z" Then
```

```
    pojtmp = LCase(ch) & Right(pojtmp, tngtou - 1)
```

```
    End If
```

```
    'SiaN-tiau e pou-hun
```

```
    st = Right(pojtmp, 1)
```

```
    If st > "8" or st <= "1" Then
```

```
    If st="p" or st="t" or st="k" or st="h" or Right(pojtmp, 2)="hN"
```

```
    Then
```

```
        st = "4"
```

```

Else
    st = "1"
End If
Else
tngtou = tngtou - 1
pojtmp = Left(pojtmp, tngtou)
End If
'response.write "siaN-tiau : " & st & "<br>"
'Unbu e pou-hun
ubflag = False
ub = Right( pojtmp, 5 )
If ub = "oaihN" Then
ubflag = True
Else
ub = Right( ub, 4 )
If ub="iang" or ub="iong" or ub="iahN" or ub="iauh" or

```

```
ub="oaiN" or _
  ub="oang" or ub="iauN" Then
  ubflag = True
Else
  ub = Right( ub, 3 )
  If ub="ahN" or ub="aiN" or ub="ang" or ub="auh" or
ub="ehN" or ub="eng" or _
  ub = "auN" or _
  ub="iaN" or ub="iah" or ub="iam" or ub="ian" or ub="iap" or
ub="iat" or _
  ub="iau" or ub="iau" or ub="ioh" or ub="iok" or ub="ioN" or
ub="iuN" or _
  ub="iak" or ub="iou" or ub="iuh" or _
  ub="ngh" or _
  ub="ouN" or ub="oaN" or ub="oah" or ub="oai" or ub="oan" or
ub="oat" or _
```

```

ub="oeh" or ub="ouh" or ub="ong" or _
ub="uiN" or ub="uih" Then
ubflag = True
Else
ub = Right( ub, 2 )
If ub="aN" or ub="ah" or ub="ai" or ub="ak" or ub="am" or
ub="an" or _
    ub="ap" or ub="at" or ub="au" or _
    ub="eN" or ub="eh" or ub="ek" or ub="iN" or ub="ia" or
ub="im" or _
    ub="in" or ub="io" or ub="ip" or ub="it" or ub="iu" or ub="ih"
or _
    ub="mh" or ub="ng" or _
    ub="ou" or ub="oN" or ub="oa" or ub="oe" or ub="oh" or
ub="ok" or _
    ub="om" or ub="op" or _

```

```
    ub="uh" or ub="ui" or ub="un" or ub="ut" Then
    ubflag = True
Else
    ub = Right( ub, 1 )
    If ub="a" or ub="e" or ub="i" or ub="m" or ub="o" or ub="u"
Then
    ubflag = True
    End If
End If
    End If
End If
    End If
SiPOJimchiat = ubflag

'SiaNbu e pou-hun
If SiPOJimchiat Then
```

```
sbflag = False
tngtou = tngtou - Len(ub)
sb = Left(pojsmp, tngtou )
If sb="chh" or sb="ph" or sb="th" or sb="kh" or sb="ch" or
sb="ng" or _
    sb="p" or sb="m" or sb="b" or sb="t" or sb="n" or sb="l" or
sb="k" or _
    sb="h" or sb="g" or sb="s" or sb="j" or sb="" Then
sbflag = True
End If
SiPOJimchiat = sbflag
End If
End Function
```



**工具程式 2 : 音節數統計(syllable  
count SC),**

**Kang-khū thêng-sek 2 : im-chiat sò  
thóng-kè**

輸入：羅馬字文本，輸出：音節數

Su-jip : Lô-má-jī bûn-pún, su-chhut :  
im-chiat-sò

**Compiler , get\_token ( )**

注意：數字處理方式(eg：2007)一  
個數字當做一個音節

Chù-ì : sò-jī chhú-lí hong-sek (eg:2007)  
chit-ê sò-jī tòng-chòe chit-ê im-chiat

漢羅文本：一漢字一音節

Hàn-lô bûn-pún : chit Hàn-jī chit  
im-chiat

線上系統：Sòa<sup>n</sup>-téng hē-thóng

<http://iug.csie.dahan.edu.tw/nmtl/dadwt/count.htm>

# (Program)

```
function clSyllCount (clTxt)
```

```
    dim prevIsSyll, curr, count, c, c1
```

```
    prevIsSyll = false
```

```
    curr = 1
```

```
    count = 0
```

```
    do while curr <= len(clTxt)
```

```
        c = mid(clTxt, curr, 1)
```

```
        ' check if [a-zA-Z0-9]
```

```
        if ( "a" <= c and c <= "z") or ("A" <= c and c <= "Z") then
```

```
            if prevIsSyll then
```

```
' within the syllable, do nothing
else
' a new syllable
  count = count + 1
  prevIsSyll = true
end if
elseif ("0" <= c and c <= "9") then
  if prevIsSyll then
    ' peek the next char
    if curr + 1 <= len(clTxt) then
      c1 = mid(clTxt, curr + 1, 1)
```

```
else
    c1 = ""
end if
' check if the tone
if ("1" <= c and c <= "9") and not ("0" <= c1 and c1 <= "9") then
    ' the digit is taken with the previous letters as a syllable, do
not count it
else
    ' the previous letters is a syllable, the digit is a new syllable
    count = count + 1
end if
```

```
    prevIsSyll = false
else
    ' a digit is counted as a syllable
    count = count + 1
end if
elseif (c = "<") then
    if (curr + 3) <= len(cITxt) then
        if lcase(mid(cITxt, curr, 4)) = "<br>" then
            ' html break is not a syllable, skip them
            curr = curr + 3
        end if
    end if
end if
```



```
    end if
    prevIsSyll = false
else
    ' not a syllable character
    prevIsSyll = false
end if
' check the next one
curr = curr + 1
loop
clSyllCount = count
end function
```

工具程式 3：音節模糊查詢(**similar  
syllable query, SSQ**)

**Kang-khū thêng-sek 3 : im-chiat  
hâm-hô chhâ-sûn**

原因：**Gôan-in**：

(1) 使用者對聲調無熟

**sú-iōng-chiá tùi sia<sup>n</sup>-tiāu bô sèk**

**-2 -3 -h -5 -7 -h8 拿掉後比對**

**-2 -3 -h -5 -7 -h8 thèh-tiāu**

**liáu-āu pí-tùi**

(2) 入聲-p -t -k -h 不太會分辨

jip-sia<sup>n</sup> -p -t -k -h bô-siá<sup>n</sup> ē-hiáu

khu-piát

-p -t -k -h 拿掉後比對

-p -t -k -h thèh-tiāu liáu-āu pí-tùi

### (3) 想要找押韻的音節

siū<sup>n</sup>-beh chhē tùi-kù (ah-ūn) ê  
im-chiat

省略聲母聲調，只取韻母比對

kā sia<sup>n</sup>-bú, ūn-bú thèh-tiāu,  
kan-na lâu ūn-bú lâi pí-tùi

台華線上辭典羅馬字無聲調查詢

<http://iug.csie.dahan.edu.tw/iug/Ungian/soannteng/chil/Taihoa.asp>

台語字典無聲調/入聲/韻母查詢

<http://iug.csie.dahan.edu.tw/TG/jitian/tgjt.asp>

# 工具程式 4：數字轉調符

Kang-khū thêng-sek 4 : sò-jī chóan  
tiāu-hû

eg : koai3 --> kòai, tiau7 --> tiāu

原則：聲調符號在韻母的主要元音  
之上(文獻寫法不一致)

Gôan-chek : sia<sup>n</sup>-tiāu hû-hō tī ūn-bú ê  
chú-iàu gôan-im téng-bīn  
(bûn-hiàn siá-hoat bô it-tì)  
a > o > e > u > i      (o>a>...)



# 演算法

**lán-sòan-hoat**

輸入：用數字表示的白話字音節

**(eg: nng7, mi5, koai2, ...)**

**Su-jip : iōng sò-jī piáu-sī ê Pèh-ōe-jī  
im-chiat**

想法：填入這些元音出現的位置

Siū<sup>n</sup>-hoat : kā chia-ê gôan-im

chhut-hiān ê só-chāi thiām --jip-khì

<b>o</b>	<b>a</b>	<b>e</b>	<b>u</b>	<b>i</b>	<b>n</b>	<b>m</b>

**step 1** : 從左到右依序將出現上  
述符號的位置填入上面的表

**step1** : ùi tò-pêng kàu chià<sup>n</sup>-pêng,  
chiong chia-ê hû-hō chhut-hiān ê  
só-chāi kā thiām jip-khì

<b>nng<sup>7</sup></b>	<b>o</b>	<b>a</b>	<b>e</b>	<b>u</b>	<b>i</b>	<b>n</b>	<b>m</b>
	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>2</b>	<b>0</b>

<b>mi<sup>5</sup></b>	<b>o</b>	<b>a</b>	<b>e</b>	<b>u</b>	<b>i</b>	<b>n</b>	<b>m</b>
	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>2</b>	<b>0</b>	<b>1</b>

<b>koai<sup>2</sup></b>	<b>o</b>	<b>a</b>	<b>e</b>	<b>u</b>	<b>i</b>	<b>n</b>	<b>m</b>
	<b>2</b>	<b>3</b>	<b>0</b>	<b>0</b>	<b>4</b>	<b>0</b>	<b>0</b>

**step 2** : 找到第一個非 0 的數字，  
並傳回

**step 2** : chhē-tiòh tē-it-ê hui 0 ê sò-jī,  
thoan --tng-lâi

加速的想法：從左到右，只要找到  
**o, a, e, u** 時，調符一定是在這  
個字母上面，所以可直接傳  
回，上表也可減少為

<b>i</b>	<b>n</b>	<b>m</b>
----------	----------	----------

Ka-sok ê siū<sup>n</sup>-hoat : ùi tò-pêng kàu  
chià<sup>n</sup>-pêng, chí-iàu chhē-tiòh o, a,  
e, u ê-sî, tiāu-hû it-tēng sī tī  
chit-ê jī-bú téng-thâu, só-í ē-tàng  
tīt-chiap thōan --tng-khì, pió  
pì<sup>n</sup>-chòe

<b>i</b>	<b>n</b>	<b>m</b>
----------	----------	----------

注意：chù-ì

(1) "ou"的部分，可考慮之前先處理，或之後再處理。

"ou" ê pō-hūn, ē-sái khó-lū  
chìn-chêng seng chhú-lí, iah-sī  
āu--lâi chiah chhú-lí



kou5 --> kôu --> kô

(2) 鼻音"nn/N" 也要處理

phī<sup>n</sup>-im mā ài chhú-lí

# (Program in C)

```
int flag[3];
int i;
char ch;
int ToneMarkerPosition ( char *syllable )
{
    for ( i=0; i<3; i++ ) flag[i] = 0; //Initial
    for ( i=0; i < strlen(syllable)-1; i++ ) //最後位置是數字，不
        必處理
    {
        ch = syllable[i];
        if ( ch=='o' || ch=='a' || ch=='e' || ch=='u' ) {
            return (i+1);
        }
    }
}
```

```
    } else if ( ch == 'i') {
        flag[0] = i+1;
    } else if ( ch == 'n') {
        flag[1] = i+1;
    } else if ( ch == 'm') {
        flag[2] = i+1;
    }
}
for ( i=0; i<3; i++ )
    if (flag[i] > 0 ) return( i );
return (-1); //error
}
```

# (Program in ASP) Asc2TU.asp

<%

' Asc2TU.asp/POJ\_dict.asp: Converts ASCII Peh-oe-ji (POJ)  
to Taiwanese Package (TP) Taigi Unicode (TU) POJ  
encoding.

' Version 1.0

' Copyright (C) 2005 Henry Hong-Giau Tan-Tenn, Iunn  
Un-gian

' E-mail: ungian@{domain hidden}

'

' Update and documentation, if any, are available at

' <http://betelnut.org/>

'

' PURPOSE

'

' (See above.)

'

' INPUTS

'

' strJipLai: String to convert.

'

' RETURNS

'

' strJipLai with vowel and nasalization glyphs in TU POJ  
encoding.

'

' LICENSE

'

' This library is free software; you can redistribute it and/or  
' modify it under the terms of the GNU Lesser General Public  
' License as published by the Free Software Foundation;  
either

' version 2.1 of the License, or (at your option) any later  
version.

' This library is distributed in the hope that it will be useful,  
' but WITHOUT ANY WARRANTY; without even the implied  
warranty of

' MERCHANTABILITY or FITNESS FOR A PARTICULAR  
PURPOSE. See the GNU

' Lesser General Public License for more details.

' You should have received a copy of the GNU Lesser General

Public

' License along with this library; if not, write to the Free  
Software

' Foundation, Inc., 59 Temple Place, Suite 330, Boston, MA  
02111-1307 USA

Dim objRE

Function Asc2TU(strJipLai)

Dim strChhutKhi, strExpandedRE, i  
strChhutKhi = strJipLai



Set objRE = New RegExp

For i = 0 To UBound(arrRE)

    strExpandedRE = "\b" & conInitCons & arrRE(i) &  
conEnding

    strChhutKhi = RE(strExpandedRE, strChhutKhi)

Next

Set objRE = Nothing

Asc2TU = strChhutKhi

End Function

%>

# poj\_TU\_dict.asp

```
<%
```

```
' Version 0.2 (updated 11/03/05)
```

```
'
```

```
' *****
```

```
' * Regular expression components *
```

```
' *****
```

```
Const conInitCons =
```

```
    "(Chh | Ch | Kh | Ng | Ph | Th | B | G | H | J | K | L | M | N | P | S | T  
     | chh | ch | kh | ng | ph | th | b | g | h | j | k | l | m | n | p | s | t | \b)"
```

```
Const conEndCons = "(ng | [hk mnpt])?"
```

```
Const conNasal = "(nn | NN | N)?"
```

```
Const conTone = "([1-578] | \b)"
```

```
Dim conEnding
```

```
conEnding = conEndCons & conNasal & conTone
```

```
' *****
```

```
' * Load array with regular expressions *
```

```
' * for vowels *
```

```
' *****
```

```
Dim arrRE(8)
```

```
arrRE(0)="(ou | ng | [maeiou])()
```

arrRE(1)="(Ou | Ng | [MAEIOU])0"

arrRE(2)="(a)([iu])"

arrRE(3)="(A)([iu])"

arrRE(4)="(u)(i)"

arrRE(5)="(U)(i)"

arrRE(6)="(o)(ai | [ae])"

arrRE(7)="(O)(ai | [ae])"

arrRE(8)="([iI])(au | [aou])"

' \*\*\*\*\*

' \* ASCII-TU Dictionary

\*

```
' * Destroy object in main script *
```

```
' *****
```

```
Dim objTU
```

```
Set objTU = CreateObject("Scripting.Dictionary")
```

```
objTU.CompareMode = 0 ' Case-sensitive
```

```
objTU.Add "A2", "&#193;"
```

```
objTU.Add "A3", "&#192;"
```

```
objTU.Add "A5", "&#194;"
```

```
objTU.Add "A8", "A&#781;"
```

```
objTU.Add "A7", "&#256;"
```

```
objTU.Add "a3", "#224;"  
objTU.Add "a2", "#225;"  
objTU.Add "a5", "#226;"  
objTU.Add "a8", "a#781;"  
objTU.Add "a7", "#257;"
```

```
objTU.Add "au3", "#224;u"  
objTU.Add "au2", "#225;u"  
objTU.Add "au5", "#226;u"  
objTU.Add "au8", "a#781;u"
```

objTU.Add "au7", "&#257;u"

objTU.Add "E2", "&#201;"

objTU.Add "E3", "&#200;"

objTU.Add "E5", "&#202;"

objTU.Add "E8", "E&#781;"

objTU.Add "E7", "&#274;"

objTU.Add "e8", "e&#781;"

objTU.Add "e3", "&#232;"

objTU.Add "e2", "&#233;"



objTU.Add "e5", "&#234;"

objTU.Add "e7", "&#275;"

objTU.Add "I7", "&#298;"

objTU.Add "I8", "I&#781;"

objTU.Add "I3", "&#204;"

objTU.Add "I2", "&#205;"

objTU.Add "I5", "&#206;"

objTU.Add "i3", "&#236;"

objTU.Add "i2", "&#237;"

```
objTU.Add "i5", "&#238;"  
objTU.Add "i7", "&#299;"  
objTU.Add "i8", "i&#781;"
```

```
objTU.Add "M2", "&#7742;"  
objTU.Add "M3", "M&#768;"  
objTU.Add "M5", "M&#770;"  
objTU.Add "M8", "M&#781;"  
objTU.Add "M7", "M&#772;"
```

```
objTU.Add "m2", "&#7743;"
```

objTU.Add "m3", "m&#768;"  
objTU.Add "m5", "m&#770;"  
objTU.Add "m7", "m&#772;"  
objTU.Add "m8", "m&#781;"

objTU.Add "N", "&#8319;"  
objTU.Add "nn", "&#8319;"

objTU.Add "Ng2", "&#323;g"  
objTU.Add "Ng3", "&#504;g"  
objTU.Add "Ng5", "N&#770;g"

objTU.Add "Ng8", "N&#781;g"

objTU.Add "Ng7", "N&#772;g"

objTU.Add "ng2", "&#324;g"

objTU.Add "ng3", "&#505;g"

objTU.Add "ng5", "n&#770;g"

objTU.Add "ng8", "n&#781;g"

objTU.Add "ng7", "n&#772;g"

objTU.Add "O2", "&#211;"

objTU.Add "O3", "&#210;"

objTU.Add "O5", "&#212;"  
objTU.Add "O8", "O&#781;"  
objTU.Add "O7", "&#332;"

objTU.Add "o3", "&#242;"  
objTU.Add "o2", "&#243;"  
objTU.Add "o5", "&#244;"  
objTU.Add "o8", "o&#781;"  
objTU.Add "o7", "&#333;"

objTU.Add "Ou1", "O&#856;"

objTU.Add "Ou2", "&#211;&#856;"  
objTU.Add "Ou3", "&#210;&#856;"  
objTU.Add "Ou5", "&#212;&#856;"  
objTU.Add "Ou8", "O&#781;&#856;"  
objTU.Add "Ou7", "&#332;&#856;"

objTU.Add "ou1", "o&#856;"  
objTU.Add "ou3", "&#242;&#856;"  
objTU.Add "ou2", "&#243;&#856;"  
objTU.Add "ou5", "&#244;&#856;"  
objTU.Add "ou8", "o&#781;&#856;"

objTU.Add "ou7", "&#333;&#856;"

objTU.Add "U2", "&#218;"

objTU.Add "U3", "&#217;"

objTU.Add "U5", "&#219;"

objTU.Add "U8", "U&#781;"

objTU.Add "U7", "&#362;"

objTU.Add "u3", "&#249;"

objTU.Add "u2", "&#250;"

objTU.Add "u5", "&#251;"

```
objTU.Add "u7", "&#363;"
```

```
objTU.Add "u8", "u&#781;"
```

```
' Destroy object in main asp page
```

```
%>
```