

台灣南島語數位典藏：詞類與標記原則

Elizabeth Zeitoun 齊莉莎 and Harry Yu 余清華

中央研究院 - 台灣南島語數位典藏 Academia Sinica Formosan Language Archive - Microsoft Internet Explorer

檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)

← 前一頁 下一頁 → 停止 重新整理 首頁 搜尋 我的最愛 媒體 記錄 郵件 列印 編輯 討論

網址(D) <http://www.ling.sinica.edu.tw/formosan/> 移至 連結 >>

中央研究院

台灣南島語數位典藏
Academia Sinica Formosan Language Archive



中文版 ENGLISH

完成 開始 網路網路

Microsoft PowerPoint - [台... 中央研究院 - 台灣南島...

AM 08:08

報告內容

- 介紹網站架構及工作進行
- 解釋語言分析之原則

第一部分

台灣南島語數位典藏之架構

計畫目的

透過WWW建立南島語的數位博物館，旨在收存編輯和傳播該語言的數位化資源，供以存取：

- (i) 南島語言的文本語料
- (ii) 比較語料

網路介面主要特徵

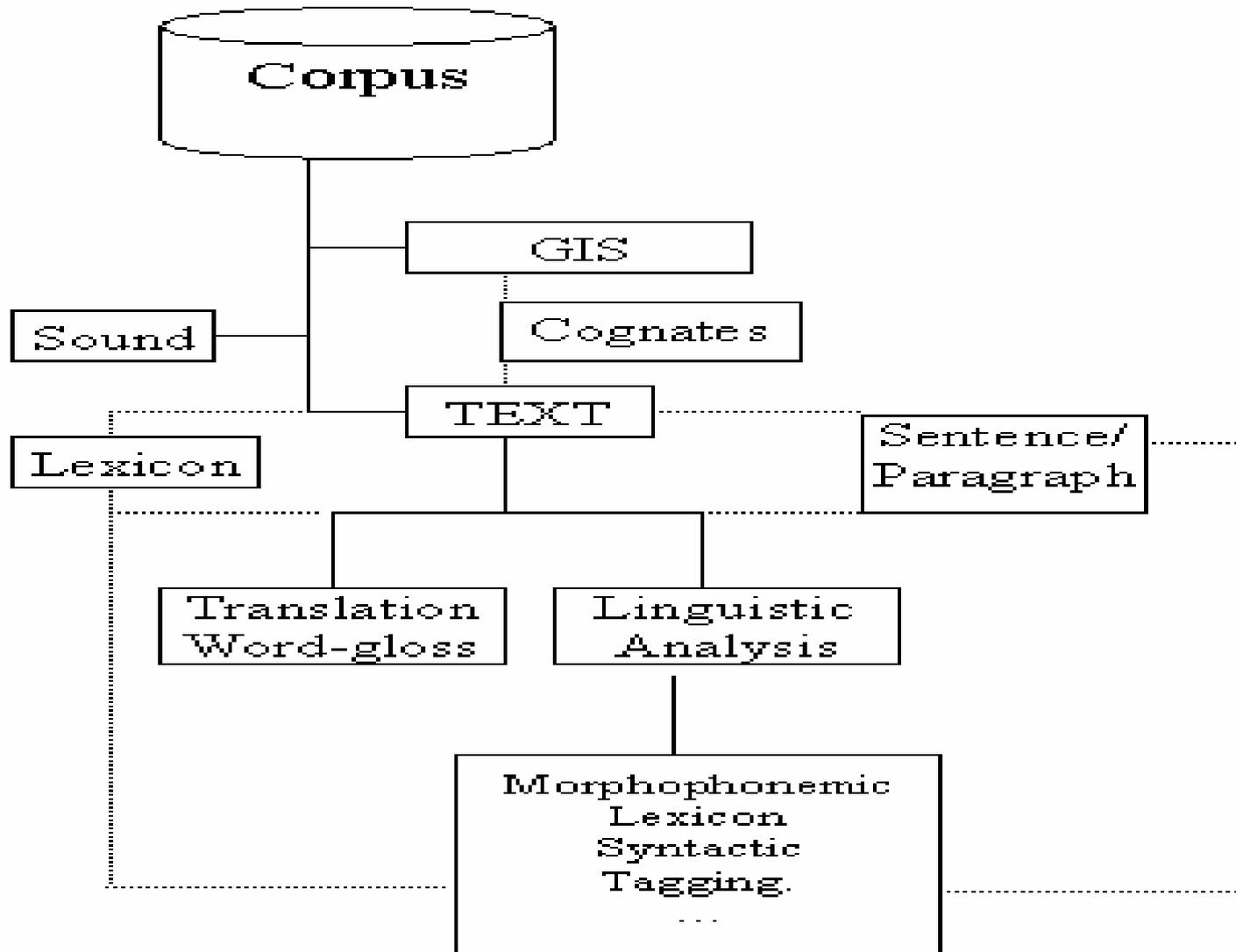
■ 有兩種介面：

1. 中文
2. 英文

■ 有三種關連的資料庫設計：

1. 語料庫
2. 地理資訊系統
3. 書目資料庫

典藏架構圖



工作安排 (語料庫)

■ 2001: 魯凱語萬山方言

■ 2002-2006 :

- 2002: 魯凱語方言 / 雅美語 (沒有放在網站)
- 2003: 鄒語、泰雅語、賽夏語
- 2004: 布農語、阿美語
- 2005: 排灣語
- 2006: 卑南語

■ 此外，沙阿魯阿語、卡那卡那富語、西拉雅語和巴澤語也將陸續數位化及上網。

語料庫介面:長篇語料查詢

搜尋範圍 - Microsoft Internet Explorer

檔案(F) 檢視(V) 我的最愛(A) 工具(T) 說明(H)

http://www.ling.sinica.edu.tw/Formosan/ch/select_corpus.htm

語料庫查詢

[典藏內容](#) [使用說明](#) [回首頁](#)

§ 自訂語料庫搜尋範圍 §

■ 設定語言別搜尋範圍：

全部
 設定範圍：語言別： 方言別：

■ 設定文體搜尋範圍：

全部
 設定範圍： 傳說 口述 對談 歌謠

■ 設定語料呈現方式：

全部
 設定範圍： 原始語料 (Original data) 重新分析語料 (Re-analyzed data)

■ 設定語料搜尋方式：

請選擇：

網際網路

Formosan Corpus - Microsoft Internet Explorer

檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)

http://www.ling.sinica.edu.tw/Formosan/en/select_corpus.htm

Formosan Corpus

[Contents of the archive](#) [Help](#)

■ Mode:

All
 Searching Domain: Language: Dialect:

■ Genre:

All
 Searching Domain: folktale narrative conversation song

■ Display Method:

All
 Searching Domain: Original data Re-analyzed data

■ Searching Method:

Please select:

完成 網際網路

句子查詢

Microsoft Internet Explorer

地址: /www.ling.sinica.edu.tw/Formosan/ch/text_frame.asp

一語言別: 魯凱語/萬山 文體: 全部 語料呈現: 重新分析語料 語料

重新檢索

1. 我們的話 [後設資料](#)

選擇其他文本

顯示

句子 - Sentence

山自稱是萬山人。

maðalaə-nai ta-piʔa-aə-na-ða po-a[acə ʔoronoho m-ia

我們.屬格 處所名物化-動態,非限定:做-處所名物化-還-他.屬格 取-名 萬山 動態,虛擬式

後就(開始)學我們萬山和屏東縣的話。

kaavaʔi-ŋa-ða ʔaomo mani a[ə ʔicaʔocaʔo ʔina vaha-nai ʔi

名物化-動態,非限定:來-已經-他.屬格 日本人 就 動態,非限定:拿 動態,虛擬式:學 這 話-我們.屬格 萬

種語言後,(就發現)我們的話有許多(和)魯凱族是一樣的。

a-pa-rama-ramao kapa-ka-ðao-ðao ta-ʔini[əo ʔi tali-

非限定:弄 動態名物化-動態,非限定:重疊-相同 很-狀態,非限定-重疊-多 主語名物化-動態,非限定:像 屬於-

我們說:「你們是同宗(的)。」

namə patoʔo: maʔ-itoloto-lo[ə-nomi m-ia.

國際網路

View the corpus - Microsoft Internet Explorer

地址: http://www.ling.sinica.edu.tw/Formosan/en/text_frame.asp

Searching Domain: Rukai/Mantauran Genre: All Display: ;re_analyzed Searching method: Browse

Choice of a text: 1. Our language [metadata](#) Select another

View: Sentence Display

01.001.a

Our ancestors used to call (themselves) Mantauran.

onaʔi ʔaamaðalaə-nai ta-piʔa-aə-na-ða po-a[acə ʔoronoho m-

that ancestor-1PE.Gen LocNmz-Dyn.NFin:do-LocNmz-still-3S.Gen give-name Mantauran Dyn

01.001.b

When the Japanese came, they started (to learn) our language as well as (the languages spoken by the

ðonaʔi a-kaavaʔi-ŋa-ða ʔaomo mani a[ə ʔicaʔocaʔo ʔina

that ClsNmz-Dyn.NFin:come-already-3S.Gen Japanese then Dyn.NFin:take Dyn.Subj:learn this

01.001.c

They compared (our) languages and (discovered that) there were many resemblances between Rukai and

mani poa a-pa-rama-ramao kapa-ka-ðao-ðao ta-ʔini[əo

then Dyn.NFin:make ActNmz-Dyn.NFin:Red-identical very-Stat.NFin-many-Red SubJNmz-Dyn.NF

01.001.d

Then they told us: "You share the same ancestry."

mani al-inamə patoʔo: maʔ-itoloto-lo[ə-nomi m-ia.

完成 國際網路

段落查詢

Microsoft Internet Explorer

檢視(V) 我的最愛(A) 工具(T) 說明(H)

http://www.ling.sinica.edu.tw/Formosan/ch/text_frame.asp

語言別: 魯凱語/萬山 文體: 全部 語料呈現: ;重新分析語料 語料 重新檢索

1. 我們的話 [後設資料](#)

段落 - Paragraph

ʔi ʔaamaɔalaə-nai ta-piʔa-aə-na-ɔa po-aʔacə ʔoponoho m-ia ɔonaʔi a-kaavaʔi-
ɔa ʔaomo mani aʔa ʔicaʔocaʔo ʔina vaha-nai ʔoponoho la tali-ʔao-ʔaɔɔo. mani
a-pa-rama-ramao kapa-ka-ɔao-ɔao ta-ʔiniʔao ʔi tali-ɔəkai ʔina vaha-nai.
ni aʔ-inamə patoʔo: maʔ-itolo-toloʔo-nomi m-ia. mani ʔihoʔo-nai ʔina ʔoponoho
ʔa-ta aanai m-ia-nai. ɔonaʔi a-ki-kaavaʔi-na-ɔa ʔaomo, o-ʔihoʔo-ka-nai maʔ-
o-toloʔo-mita ɔəkai m-ia. kasa-ta ʔoponoho vaha-ta om-ia-nai.

的祖先萬山自稱是萬山人。日本人來了以後就（開始）學我們萬山和屏東縣的話。他們比
兩種語言後，（就發現）我們的話有許多（和）魯凱族是一樣的。然後他們告訴我們說：
們是同宗（的）。」我們萬山人（才）知道「原來咱們（是屬於魯凱族）。」日本人還沒
前，我們不知道我們和魯凱族同宗。我們只（認為）咱們的語言（就）是萬山話。

ʔi vaha-nai ʔi ʔoponoho toramoro ka ma-kociŋai. cəəla ʔinaʔi la-a-lakə-nai ta-
ʔisi-ŋa: o-ʔihoʔo-ka-i ka pa-loloɔo ʔinaʔi ta-ia-ə-nai ʔao-vaə-vaha ta-ka-a-
ŋə. cəəla ʔi vaʔavaʔa: lo m-aʔa-nai siri-ri ka mati-ʔihoʔo-ka-i m-iiʔa vaha-nai.
ʔao-vaə-vaha-nai ka-ʔoponoh-aə ocaə ʔinaʔi vaha-nai, ta-ʔihoʔo to-siʔi olo
a-nai, ʔako-ʔoponoh-aə-nai om-iʔa-ka-i ka ma-rama-ramao. toʔa-toloraə-ni

網際網路

View the corpus - Microsoft Internet Explorer

檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)

http://www.ling.sinica.edu.tw/Formosan/en/text_frame.asp

Searching Domain: Rukai/Mantauran Genre: All Display:
;re_analyzed Searching method: Browse

Choice of a text: 1. Our language [metadata](#)

View: Paragraph

1.

onaʔi ʔaamaɔalaə-nai ta-piʔa-aə-na-ɔa po-aʔacə ʔoponoho m-ia ɔonaʔi a-kaavaʔi-ŋa-ɔa ʔa
aʔa ʔicaʔocaʔo ʔina vaha-nai ʔoponoho la tali-ʔao-ʔaɔɔo. mani poa a-pa-rama-ramao kapa-ka-
ta-ʔiniʔao ʔi tali-ɔəkai ʔina vaha-nai. mani aʔ-inamə patoʔo: maʔ-itolo-toloʔo-nomi m-ia. ma
nai ʔina ʔoponoho ka-iʔa-ta aanai m-ia-nai. ɔonaʔi a-ki-kaavaʔi-na-ɔa ʔaomo, o-ʔihoʔo-ka-
itolo-toloʔo-mita ɔəkai m-ia. kasa-ta ʔoponoho vaha-ta om-ia-nai.

Our ancestors used to call (themselves) Mantauran. When the Japanese came, they started (to learn) our
language as well as (the languages spoken by the people living in) Pingtung county. They compared
languages and (discovered that) there were many resemblances between Rukai and our (own) language.
language. Then they told us: "You share the same ancestry." (That's how) we, Mantauran, learned that
actually we were Rukai. Before the Japanese came, we did not know we were related to the Rukai. We
knew that our language was Mantauran.

2.

ʔinaʔi vaha-nai ʔi ʔoponoho toramoro ka ma-kociŋai. cəəla ʔinaʔi la-a-lakə-nai ta-ovaʔisi-ŋa
ʔihoʔo-ka-i ka pa-loloɔo ʔinaʔi ta-ia-ə-nai ʔao-vaə-vaha ta-ka-a-roɔŋə. cəəla ʔi vaʔavaʔa:
nai siri-ri ka mati-ʔihoʔo-ka-i m-iiʔa vaha-nai. olo ʔao-vaə-vaha-nai ka-ʔoponoh-aə ocaə ʔ
vaha-nai, ta-ʔihoʔo to-siʔi olo vaha-nai, ʔako-ʔoponoh-aə-nai om-iʔa-ka-i ka ma-rama-ram
toloraə-ni vaha-nai ʔoponoho toramoro ma-kociŋai.

Our language is very difficult (to learn). Look at the young generation of children: (they) do not
know (how to) repeat (what) the old people say. Look at those outsiders: (even) if they inter-married
with our people, they still do not know (how to) repeat (what) the old people say.

完成 網際網路

原始/重新分析之語料對比

視語料庫 - Microsoft Internet Explorer

編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)

http://www.ling.sinica.edu.tw/formosan/ch/text_frame.asp

搜尋範圍: 一 語言別: 魯凱語/大南 文體: 傳說;;; 語料呈現: 原始語料;重新分析語料 語料搜尋: 瀏覽模式

重新檢索

選擇其他文本

顯示

本別: 1. 豹來了 [後設資料](#)

類別: 句子 - Sentence

Lang/Dial: 魯凱語/大南 Genre: 傳說 Text: 1/豹來了 Disp: 原始語料;重新分析語料
Format: raw Audio: off Sentence translation Using: English

001
I shall tell a story now.
omawlaðako kayasa.
I tell story now

002
There was a family.
kai ka cikikil.
at * village

003
There were two children.
kai ka taDosa ka lavalak.
at * two * children

004
Their mother said, "Boys, get up and go tend our sheep."
tinalini ka, lamataka, anDakala si la wa ʔaorikanɪ ki b
their mother * brothers get up and then go tend * o

005
In those brothers got up in the morning.

Lang/Dial: 魯凱語/大南 Genre: 傳說 Text: 1/豹來了 Disp: 原始語料;重新分析語料
Format: edited Audio: off Sentence translation Using: English

01.001
I shall tell a story now.
ay-to-mawlað-ako kayasa.
will-make-story-1S.Nom now

01.002
There was a family.
y-a-kay ka cəkə-cəkəl.
at-Real-this * Red-village

01.003
There were two children.
y-a-kay ka ta-ɖosa ka la-valak.
at-Real-this * *-two * plur-child

01.004
Their mother said, "Boys, get up and go tend our sheep."
ko tina-lini ka la-ma-taka, anɖakal-a s
* middle-aged woman-3P.Gen Top plur-Rec-elder sibling get up-imp ar

01.005
The (two) brothers got up in the morning.

成

開始

檢視語料庫 - Mic... Microsoft PowerPo... 台灣南島語數位... 檔案總管 - Public... 台灣南島語數位... 網際網路 PM 12:4

詞彙分佈

視語料庫 - Microsoft Internet Explorer

編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)

http://localhost/formosan/ch/text_frame.htm

別: 萬山方言 選擇其他方言

別: 我們的話 選擇其他文章

別: 逐字索引 - Concordance 顯示

功能分析: 尋找詞彙、詞根、語意 顯示

onaʔi ʔaamaɖalæ-nai ta-piʔa-æ-na-ɖa po-a|acə
cæla ʔi va|ava|æ lo m-a|a-nai siri-ri
-vao-vaha-nai ka-ʔoponoh-æ ocao ʔinaʔi vaha-nai ta-|ihoʔo
mani poa a-pa-rama-ramao kapa-ka-ɖao-ɖao ta-ʔini|ao
-nai ta-ovalisi-ŋa o-|ihoʔo-ka-i ka pa-loloɖo ʔinaʔi ta-ia-ə-nai
-ka-i ka pa-loloɖo ʔinaʔi ta-ia-ə-nai ʔao-vao-vaha ta-ka-a-roɖaŋə
pa a-pa-rama-ramao kapa-ka-ɖao-ɖao ta-ʔini|ao ʔi tali-ɖəkai
aomo mani a|a ʔicaʔocaʔo ʔina vaha-nai ʔoponoho la tali-|ao-|aoɖo
a-toloræ-li ʔinaʔi vaha-nai ʔoponoho alaka-i ʔinaʔi ta-ovalisi-ŋa
cæla ʔinaʔi la-a-lakə-nai ta-ovalisi-ŋa
onaʔi ʔaamaɖalæ-nai ta-piʔa-æ-na-ɖa
toʔa-toloræ-li
toʔa-toloræ-li
toʔa-toloræ-li vaha-nai ʔoponoho toramoro
ʔinaʔi vaha-nai ʔi ʔoponoho toramoro
to-siʔi
ʔoponoho m-ia
ka mati-|ihoʔo-ka-i m-iiʔa vaha-nai
to-siʔi olo vaha-nai ʔako-ʔoponoh-æ-na
ʔi tali-ɖəkai ʔina vaha-nai
ʔao-vao-vaha ta-ka-a-roɖaŋə
ʔina vaha-nai
a-o|olai ka o-|ihoʔo-ka-i ka ʔoponoho va
o-|ihoʔo-ka-i ka pa-loloɖo ʔinaʔi ta-ia-ə
po-a|acə ʔoponoho m-ia
ʔinaʔi vaha-nai ʔoponoho alaka-i ʔinaʔi ta
vaha-nai ʔoponoho toramoro ma-kociŋai
ma-kociŋai
ka ma-kociŋai
to-siʔi

近端內部網路

Microsoft PowerPoint - [工... 檢視語料庫 - Microsoft I...

上午 11:15

檢索畫面

資料庫 - Microsoft Internet Explorer

檢視(V) 我的最愛(A) 工具(T) 說明(H)

搜尋 我的最愛 媒體

http://www.ling.sinica.edu.tw/formosan/ch/srch_form.asp

內容檢索

[回上頁](#) [使用說明](#) [回首頁](#)

語言別: 全部 文體: 全部 語料搜尋方式: Keyword Search

詞(原文):

詞(中譯):

選單:

代名詞選單:

關鍵字示例:

1. 關鍵字以原文輸入時,若使用到 IPA 音標,請點選對應的代碼如下:

IPA	ð	ŋ	ɔ	ə	l	ʔ	+	θ
代碼	240	331	598	601	621	660	616	952
IPA	ø	ʃ	æ	œ	β	ʰ	ʎ	
代碼	216	643	230	339	946	649	404	
2. 關鍵字以中文輸入時,請輸入可能出現的字(如:酒、芒草、書)

執行 清除

網際網路

Searching the archive - Microsoft Internet Explorer

檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)

上一頁 搜尋 我的最愛 媒體

網址(D) http://www.ling.sinica.edu.tw/Formosan/en/srch_form.asp

Searching the archive

[Back](#) [Help](#) [Home](#)

Searching Domain (language/dialect): All Genre: All Searching method: Search

Usage:

- 1. To enter an IPA character using original language, please press the code button instead:

IPA	ð	ŋ	ɔ	ə	l	ʔ	+	θ
Code	240	331	598	601	621	660	616	952
IPA	ø	ʃ	æ	œ	β	ʰ	ʎ	
Code	216	643	230	339	946	649	404	
- 2. To enter the English keyword, please type a word that may occur (such as book, wine, etc.)

Keyword (original):

Keyword (English):

Lexical Category:

Personal Pronoun:

Search Clear

完成 網際網路

檢索結果

搜尋結果 - Microsoft Internet Explorer

檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)

← 上一頁 → 搜尋 我的最愛 媒體

網址(D) <http://www.ling.sinica.edu.tw/formosan/ch/srch.asp?PageNum=9> 移至

■ 搜尋範圍—語言別: 全部 文體: 全部 語料搜尋方式: Keyword Search 目標詞: maa [內容檢索](#)

搜尋結果

[第一頁](#) [上一頁](#) [下一頁](#) [最後一頁](#) 選擇頁次:

201.  [魯凱語-萬山-口述](#)

原文: lo **maa** vanao-mao toʔaraki ʔa-paavavanao la ʔaihaihai.

註解: 如果 動態.虛擬式:洗澡-無人稱.主格 使用 工具名物化-動態.非限定:洗澡.重疊 和 毛巾

中譯: 洗澡(時也有)香皂和毛巾(可)使用。

英譯: When we bathe, we can use a soap and a towel.

202.  [魯凱語-萬山-口述](#)

原文: a-[opəŋə-ŋa-nai korokoro, mani po-inama solata

註解: 分句名物化-動態.非限定:完成-已經-我們.主格 動態.虛擬式:挖 就 動態.非限定:弄-我們.斜格 動態.虛擬式:寫

中譯: 挖完後(政府)就讓我們自己去抽一張籤,看(抽中什麼)名字。

英譯: When we had finished digging, (the officers) wrote our names and rolled (each piece of) paper and each

完成 網際網路

建立語料庫

轉換機制：

自由格式 → 機讀格式 → xml & 資料庫格式

網頁顯示 IPA 符號：使用 numeric code references (如 #601;) 取代小圖形 (如 schwa - ə)

設計自動化程式辨認段落、句子及單字

資料庫綱目：Metadata Set

多語句子對齊：使用隱藏表格分開原句及中、英譯句

句子層級資料庫

location	c_fretran	e_fretran	soundpath	textid
001a	我們的祖先自稱萬山人。	Our ancestors used to call themselves Mantauran.	Mantauran/001a.mp3	1

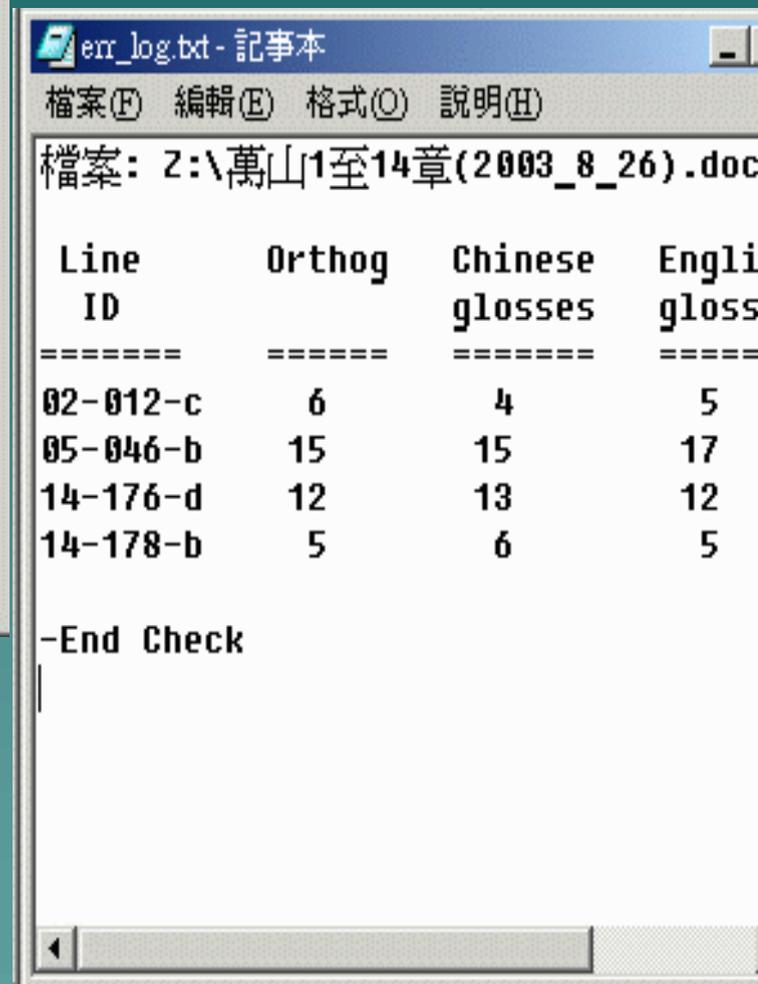
單字層級資料庫

Position	Wordorder	orthog	cgls	egls	textid
01a	0	onaʔi	那	that	1
01a	1	ʔaamaðalaə-nai	祖先-我們.屬格	ancestor-1PE.Gen	1
01a	2	ta-piʔa-aə-na- ða	處所名物化-動態.非限定: 做-處所名物化-還-他.屬 格	LocNmz-Dyn.NFin:do- LocNmz-still-3S.Gen	1
01a	3	po-aɭacə	取-名	give-name	1
01a	4	ʔoponoho	萬山	Mantauran	1
01a	5	m-ia	動態.虛擬式-這樣	Dyn.Subj-so	1

e: D:\Rukai Converter\Rukai.mdb ... Run

ction	c_freetra	e_freetra	soundfile	textid	lang_id	dial_id	genre	location	order	orthog	punc	left_t	right	open	close	cgl	egls	tex									
Tu_01_001_a	所謂撒海	The ancest	01_001_a.a	1	賽夏語	東河	口述	SAITu_01_001_a	0	koʃa&							動態,非限定	Dyn.NFin.call-									
								SAITu_01_001_a	1	ʃawan												家名	family name				
								SAITu_01_001_a	2	ka-k-hay													時間名物化	TempNmz-Sta			
								SAITu_01_001_a	3	k-in-tatini														狀態,非限定	Stat.NFin-Obj		
								SAITu_01_001_a	4	minʔal														動態,限定	Dyn.Fin.AF.co		
								SAITu_01_001_a	5	ila														動貌	Asp		
								SAITu_01_001_a	6	haw													*	*			
								SAITu_01_001_a	7	ʔi														在	at		
Tu_01_001_b	田美上ノ	The upper	01_001_b.	1	賽夏語	東河	口述	SAITu_01_001_b	0	tien-moi								地名	place name								
								SAITu_01_001_b	1	raysʔ												地名	place name				
								SAITu_01_001_b	2	koʃa&													上面	up			
								SAITu_01_001_b	3	ʔami&														動態,非限定	Dyn.NFin.call-		
								SAITu_01_001_b	4	noka														地名	place name		
								SAITu_01_001_b	5	lipon														屬格	Gen		
								SAITu_01_001_b	6	koʃa&														日本	Japan		
								SAITu_01_001_b	7	ʔami&														動態,非限定	Dyn.NFin.call-		
SAITu_01_001_b	8	a														地名	place name										
Tu_01_002_a	住在Am	How were	01_002_a.a	1	賽夏語	東河	口述	SAITu_01_002_a	0	o...								驚嘆語	Excl								
								SAITu_01_002_a	1	maʔ													也	also			
								SAITu_01_002_a	2	ʔino&														何處	where		
								SAITu_01_002_a	3	ʔihan														*	*		
								SAITu_01_002_a	4	ila														動貌	Asp		
								SAITu_01_002_a	5	ʔihan														*	*		
								SAITu_01_002_a	6	noka														屬格	Gen		
								SAITu_01_002_a	7	lipon														日本	Japan		
								SAITu_01_002_a	8	pay														動態,非限定	Dyn.NFin.pass		
								SAITu_01_002_a	9	ila														動貌	Asp		
								SAITu_01_002_a	10	hiniʔi															這-受事焦點	this-PF	
								SAITu_01_002_a	11	ʔi															在	at	
								SAITu_01_002_a	12	rayʔin															地名	place name	
								SAITu_01_002_a	13	ray															在	at	
								SAITu_01_002_a	14	pagaaʃ															地名	place name	
								SAITu_01_002_a	15	ka															主格	Nom	
SAITu_01_002_a	16	maʔji&															人	person									

檢查一致性



第二部分

語言分析之主要原則

主要原則

- (1) 書寫系統
- (2) 記錄原則
- (3) 子句、句子及段落之界線
- (4) 闡釋、詞綴及詞類分析原則

書寫系統

■ 書寫系統: 國際音標

■ 理由:

- 不同書寫系統之不一致性:

- 台灣南島語之間之書寫系統不一致, 比如: 阿美語 d = [d̚], 但是 = 魯凱語 [d]

- 學者、專家也不太一致, 比如: Blust 使用 c 紀錄 [c] 因古南島語 *C > 邵語 c 但是大部分的專家學者使用 th 來紀錄 [c]。

- 國際音標符合國際標準 (參見 Rosetta 計畫)

記錄原則

- 主要標準以構詞為主（**morphophonemic analysis**）。
 - ▶ 詞之認定以「詞素」（morpheme）為原則。
 - ▶ 台灣南島語的詞素可分為：詞根、詞綴、依附詞。
- 這種分析可以讓我們解釋：
 - 詞根的主要意思（cf. “**core meaning**” of a root）；解釋詞的意思不止於該詞彙出現的地方（meaning is NOT “context-based”，比如：魯凱語萬山方言的 *taka* 不表示「哥哥」或「姊姊」而表示「哥姐」。
 - 詞綴的意思與分佈
 - 語音轉換（比如：魯凱語萬山方言的 *o-va*  *ai* 給 vs. *vaa*  *i* 或 *voo*  *i*。）

子句、句子及段落界線

- 兩種因素可以協助我們決定子句、句子的界線

(一) 語調

(二) 句法結構

闡釋及詞類原則

- 闡釋是我們作語言分析通常使用的. 有些符號需要解釋
 - : = divisible affix (but not easily divisible in corpus)
 - . = portmanteau morpheme
 - - = bound morpheme (no distinction so far between affixes and clitics)
 - <> = infix

詞綴及詞類：原則及分析

因此除了闡釋提供詞義及句法功能之外（如：是否動詞為限定或非限定，代名詞為屬格或主格等），我們提供其他兩種查詢方式：詞綴分析及詞類別分析。

原則為利用構詞規則

搜尋詞庫時，可顯示所有關聯的詞彙與句子。

詞類標記：

- 表示詞類
- 語意
- 構詞上之特色（如：名詞是否指的是人[+ pers] 或非人[- pers]）
- 說明詞根是否為自由式或附著式
- 可找出詞類別及相關詞彙

詞類標記

(詞庫使用的) 中文詞類及台灣南島語的詞類比較

Tagging

1. A
2. C
3. ADV
4. ASP
5. N
6. DET
7. M
8. T
9. P
10. Vi
11. Vt
12. POST
13. FW
14. U

English

- Adjective
- Conjunction
- Adverb
- Aspect
- Noun
- Determiner
- Measure
- Particle
- Preposition
- Intransitive Verb
- Transitive Verb
- Postposition
- Foreign Words
- Undecided

Rukai

- $\frac{3}{4}$
- ✓
- $\frac{3}{4}$
- ✓
- ✓
- ✓
- ✓
- ✓
- ✓
- ✓
- $\frac{3}{4}$
- ✓
- ✓

Other Fn lgs

- (✓)
- ✓
- (✓)
- ✓
- ✓
- ✓
- ✓
- ✓
- ✓
- ✓
- $\frac{3}{4}$
- ✓
- ✓

(詞庫尚未使用的) 中文詞類及台灣南島語務必使用的詞類比較

Tagging	English	Rukai	Other Fn Igs
1. AUX	Auxiliary verb	$\frac{3}{4}$	✓
2. NEG	Negator	✓	✓
3. TOP	Topic	✓	✓
4. MOD	Modality	✓	✓
5. Tns	Tense	$\frac{3}{4}$	(✓)

計畫工作小組

計畫主持人：E. Zeitoun

工作小組

語言分析：E. Zeitoun 齊莉莎, Lin Hui-chuan 林惠娟、Hsin Tien-hsin

邢天馨 (Rukai)

Chu Tai-hwa 朱黛華 (Saisiyat)

Yeh Yu-ting 葉郁婷 (Atayal)

地理資訊系統：Hua Jia-jing 華加婧、Bai Bing-ling 白璧玲

工程師：Yu Ching-hua 余清華

數位化：Lin Zhi-xian 林志憲 (Bunun, Tsou, Pazih)

Lin Cui-wei 林翠緯 (Amis)

後設資料：Weng Cui-xia 翁翠霞

書目資料庫：Liu Chiu-yun 劉秋雲

Hua Jia-jing 華加婧

國科會贊助以及中研院語言所部分補助

**Thank you for your attention,
Comments welcome!**

<http://www.ling.sinica.edu.tw/formosan>