Speaking of *kong*^{2}

A corpus-based description

Joren Pronk

Student number: s1044664
Supervisor: dr. J.M. Wiedenhof
May 30, 2017
MA thesis
Research Master in Linguistics
Leiden University

Cover illustration: *The character* 講 *kong*[2].
Source: Chinese-tools (2016)

Table of contents

List of abbreviations & symbols

-	(in glosses) indicates an affix boundary (in transcriptions) indicates a syllable has undergone tone sandhi and is placed between the number of the tone as pronounced in isolation and the one of the tone pronounced in its sandhi form, e.g. *hi*$^{1\text{-}7}$*bang*^{7} 'to hope' (in translations) indicates syllable boundaries in the Tai$^{5\text{-}7}$lo^{5} romanization system
--	indicates hesitation by the speaker or interruption from an interlocutor
~	indicates variants of the same form and comes between the form with the tone as pronounced in isolation and the one pronounced in its sandhi form, e.g. *kong*^{2} ~ *kong*$^{2\text{-}1}$
=	indicates a clitic boundary
A + B	indicates A is followed by B, optionally with something in between
A > B	indicates A becomes B
*	(preceding an example sentence) indicates the following example is either ungrammatical or does not correspond with the intended meaning
.	(in glosses) indicates the division of two English words that correspond to the same morpheme in TSM
,	(between meanings) indicates related meanings corresponding with on form in the target language
:	(in glosses) indicates separate meanings for fused morphemes, e.g. *tse*^{1} is a fusion of *tsit*^{4} 'this' and e^{5} 'entity' and is accordingly glossed 'this:entity'
;	(between meanings) indicates the end of the first meaning and the start of the next leaving the question of the relatedness of the different meanings aside
?	(in transcriptions) indicates I am not certain whether following form is transcribed properly or not (in glosses) indicates I am not sure whether the preceding gloss is the right one or not
'text'	indicates a meaning
"text"	indicates a quotation
text	indicates a linguistic form or a technical term that is explained for the first time
text	(in transcriptions) indicates a highlighted form, i.e. a for that particular discussion relevant form
TEXT	indicates a gloss with a grammatical meaning
<text>	indicates a written form

	(in transcriptions) indicates a Mandarin loan
/text/	indicates phonological form
[text]	indicates phonetic form
0	(in transcriptions) *khin$^{1-7}$siann^{1}*, Neutral tone
1	(in glosses) 1st person
	(in transcriptions) *im$^{1-7}$ping^{5}* tone, first tone
2	(in transcriptions) *siong^{7}* tone, second tone
3	(in glosses) 3rd person
	(in transcriptions) *im$^{1-7}$khi^{3}* tone, third tone
4	(in transcriptions) *im$^{1-7}$jip^{8}* tone, fourth tone
5	(in transcriptions) *iong$^{5-7}$ping^{5}* tone, fifth tone
7	(in transcriptions) *iong$^{5-7}$khi^{3}* tone, seventh tone
8	(in transcriptions) *iong$^{5-7}$jip^{8}* tone, eighth tone
9	(in transcriptions) high rising tone
AGR	agreement, an interjection used to express agreement
ATT	attention, an interjection used to express that the speaker is about to say something
AUX	auxiliary
BKCH	backchannel, interjections used by the listener to express that he hears and understands what the speaker is saying
C	common gender
cf.	*confer* 'compare'
CFM	confirmation, an interjection used to express confirmation and agreement
COMPL	complementizer
DASS	dogmatic assertion, an interjection used to express the speaker feels something is and should be known to all ('as we all know') (cf. Chao 2011 [1968]: 803-804)
DEF	definite
DIR	directional, a preposition used to express lative and ablative meaning
Du.	Dutch
DUR	durative, a clitic used to describe a certain event as being continuous (cf. Wiedenhof 2015: 231)

EC	expected continuation, an interjection used to express the expectation that the preceding utterance will get a follow-up by either the speaker or the listener (cf. Wiedenhof 2015: 242-244)
ed.	editor
eds.	editors
e.g.	*exempli gratia* 'for example'
et al.	*et alii* ‘and so forth’
etc.	*et cetera* 'and others'
EXP	experiential, a clitic used to express the aspectual meaning that someone has experienced the action expressed by the verb in a time preceding the narrated time not directly, but with an interval in between (cf. Wiedenhof 2015: 228)
EXST	existential, a verb commonly used to present entities that in the context are considered unknown, e.g. $u^{7\text{-}3}$ gu^{5} 'there is a buffalo', or an auxiliary used to express perfect aspect
f.	female
FAM	familiarity, a prefix used to attach to nouns to indicate familiarity usually of the speaker to the referent of the noun this prefix attaches to
FUT	future, an auxiliary verb used to indicate planned activities or events in the relative future
haha	laughter
HST	hesitation
ibid.	*ibidem* 'in the same place'
i.e.	*id est* 'this is'
INCL	inclusive person
m.	male
M	measure word
MoE	ministry of Education
N	noun
NEG	negation
No.	*numéro* 'number'
O	object
p.	page
PL	plural
pp.	pages

PROG	progressive aspect
PRTC	participant, expressed by the morpheme *ka*^{7} 'PRTC', which introduces an extra participant to the verbal event
Pt.	Portuguese
QW	question, a morpheme used to express polar interrogativity ('yes or no')
RLV	relevance, an interjection used to express that the speaker feels the preceding utterance is relevant to the context in which it occurs
RLT	relational, a conjunction used between clauses and sentences to express both contrasting ('but') and non-contrasting ('and') coordinativity
ROC	Republic of China
S	subject
SG	singular
SRPS	surprise, an interjection used to express the speaker's surprise
sth.	something
SUB	subordination, a morpheme used to express subordination of preceding to following morphemes, e.g. *gu*^{5} *e*$^{5\text{-}7}$ *bak*$^{8\text{-}4}$*tsiu*^{1} 'the eyes of the buffalo'
TML	Tai$^{5\text{-}7}$uan$^{5\text{-}7}$ Bu$^{2\text{-}1}$gi$^{2\text{-}1}$ Lian$^{5\text{-}7}$bing^{5} 'Alliance for the Native Languages of Taiwan'
TSM	Taiwanese Southern Min
V	verb
viz.	*videlicet* 'namely'
WARN	warning, a interjection used to express the speaker's strong belief that the interlocutor must be aware of something
WWII	Second World War
XPC	expectation, an interjection used to express the speaker's expectation that the listener is aware of the preceding information

1 The study of Taiwanese

Figure 1: *Kóng ~ káng* in Douglas (1899: 246)

kóng (R. káng), to say; to talk; to discourse upon; to discuss; to explain by discoursing (Cn. "káng," to discourse or discuss; Cn. "söh," to say, to talk).

kóng-kè, to ask more than the proper price, gradually letting it down through haggling. **kóng-hô**, to discuss matters till agreed, as persons at feud, or as two men plotting against a third.

kóng-phoà, to tell the particulars. **kóng-chheh** (Cn. káng-tsü), to discourse on a book, explaining it. **kóng tō-lí**, to discourse on doctrine; to preach.

Figure 2: *Kóng ~ káng* in Lôo (2011: 191)

講 白 kóng 白 káng

一、說。例：1. 講話kóng-uē、2. 講袂出喙kóng bē tshut tshuì（說不出口）、3. 講甲喙角全波kóng kah tshuì-kak tsuân pho（說得嘴角冒泡；說得天花亂墜）、4. 講情kóng-tsîng（說情；說項）。

Figure 1 above shows an excerpt from the entry *kóng* in the monumental Hokkien dictionary by Carstairs Douglas (1899: 246). The first meanings given are 'to say' and 'to talk'.

More than a century later, the voluminous 實用台語詞典 *Sit$^{8-4}$iong$^{7-3}$ Tai$^{5-7}$gi$^{2-1}$ su$^{5-7}$tian^{2} / Shíyòng Táiyǔ cídiǎn*, Figure 2, by Lôo Kńg-tsiânn 盧廣誠 (2011: 191) gives a similar meaning in its first description under the entry "講 *kóng*":

> "Say. Examples: 1. 講話 *kóng-uē* 'speak', 2. 講袂出喙 *kóng bē tshut tshuì* 'can't get the words out', 3. 講甲喙角全波 *kóng kah tshuì-kak tsuân pho* 'speak in such a way that bubbles of saliva form in the corner

of the mouth', 4. 講情 *kóng-tsîng* 'intercede; put in a good word for someone'."

Here, the first meaning given for *kong*^{2} is its equivalent in modern Mandarin: 說 *shuō* 'to speak, to talk, to utter, to say' (Zhāng 2009: 1410).

Example (1) is a Taiwanese Southern Mǐn (TSM) sentence from my spoken corpus in which we find the morpheme *kong*$^{2-1}$ ~ *kong*^{2}:

(1) A^{0} ka$^{3-2}$ tsit$^{4-8}$ tsun^{7}, eh-- eh-- tau$^{3-2}$te^{2} =*ne*, lan$^{2-1}$ i$^{2-1}$tsing$^{1-7}$ u$^{7-3}$ **kong$^{2-1}$** =kue$^{3-2}$
RLT arrive this moment HST HST in.the.end =RLV 1INCL already[1] EXST **say** =EXP

siann$^{2-1}$mih$^{8-3}$ hue^{7}?
what thing

'So and up to this moment eh- eh- what exactly have we already talked about?' (5A; 06:23)

The usage of *kong*^{2} in (1) seems to corroborate both accounts in these dictionaries: *u*$^{7-3}$ *kong*$^{2-1}$-*kue*$^{3-2}$ translates to the English 'have talked about'. Let us, however, consider example (2) below, taken from the same corpus:

(2) A^{0} soo$^{2-1}$i$^{2-1}$ kho$^{2-1}$ling^{5} =*ne*, eh-- koh$^{4-1}$ ui$^{7-3}$ thiann$^{1-7}$tsiong$^{3-2}$ ping$^{5-7}$iu^{2} e$^{7-3}$ kam$^{2-1}$kak$^{4-8}$
RLT therefore maybe =RLV HST every place listening.audience friend will feel

kong$^{2-1}$ m^{0} khak$^{4-8}$sit^{8} =*ne*, tsin$^{1-7}$tsiann^{2} tsainn$^{1-7}$iann$^{2-1}$-- beh$^{4-1}$ tsainn$^{1-7}$iann$^{2-1}$
speak AGR indeed =RLV really know want.to know

i$^{1-7}$ e$^{5-7}$ i$^{3-2}$su^{3}.
3 SUB meaning

'So it may be that, uhm-- our listeners will feel that, hm yes, they really know-- want to know its meaning.' (3A; 05:40)

In example (2) too, we encounter *kong*$^{2-1}$. This time, however, it does not seem clear that *kong*$^{2-1}$ means 'say' or 'speak'.

This raises the following question: in what ways is *kong*^{2} ~ *kong*$^{2-1}$ used in TSM? The aim of the present article is to answer this question by giving a description of *kong*^{2} ~ *kong*$^{2-1}$ in TSM

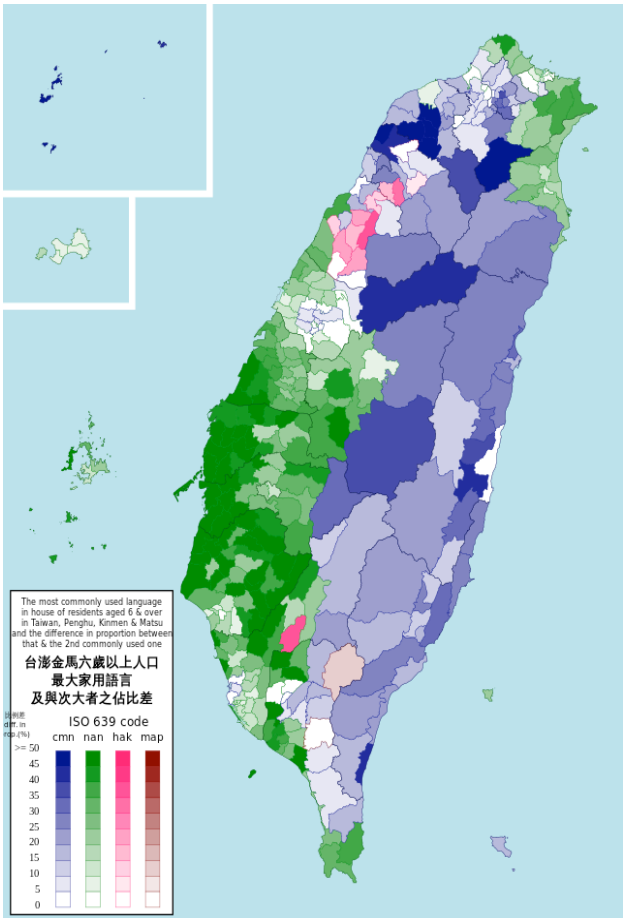
1 Note that the speaker uses *i*$^{2-1}$*tsing*^{1} to mean 'already'. The regular Taiwanese pronunciation is *i*$^{2-1}$*king*^{1}. This is probably due to influence of the Mandarin form *yǐjīng* 'already'.

accounting for all the instances in my corpus. The description will provide an analysis firmly rooted in modern spoken data.

1.1 Taiwanese Southern Mĭn[2]

Figure 3: Most commonly used language at home in 2010

(Source: Wikimedia 2015)



The term Taiwanese is used to refer to the varieties of Southern Mĭn as spoken on the island of Taiwan and a number of smaller islands in its vicinity, such as the Pescadores and Quemoy. This is the area that is *de facto* governed by the government of the Republic of China (ROC). Taiwanese currently has around 14,000,000 speakers (Lín 2008: 5). Figure 3 presents the area's where Taiwanese is the most commonly spoken home language in green; in the blue area's, Mandarin is the most commonly spoken language at home, in the pink area's Hakka and in the brown area Austronesian languages.

For a long time, speakers of almost exclusively Austronesian languages lived on Taiwan. Around the mid 17th century, many speakers of Sinitic languages, predominantly from the area around the cities of 漳州 Tsiang$^{1\text{-}7}$tsiu^{1} and 泉州 Tsuan$^{5\text{-}7}$tsiu^{1}, situated in the modern province of 福 建 Hok$^{4\text{-}8}$kian^{3} (Mand. *Fújiàn*), settled on Taiwan. These people brought along their own languages, especially the Tsiang$^{1\text{-}7}$tsiu^{1} and Tsuan$^{5\text{-}7}$tsiu^{1} varieties of Southern Mĭn. In addition to these Southern Mĭn varieties, considerable groups of Hakka speakers also crossed over to Taiwan (Klöter 2005: 1).

From 1895 to 1945, Taiwan was a colony of the Japanese empire. When Japan surrendered

2 The discussion in this section has partly been borrowed from Pronk (2014: 8-11).

at the end of the Second World War (WWII), Taiwan was ceded to the ROC. In the years between the end of WWII and 1949, the ROC was in civil war between nationalists and communists. The nationalists lost and after the establishment of the People's Republic of China (PRC) on October 1, 1949, they fled to the island of Taiwan, which the ROC had acquired just a few years earlier from Japan (Schoppa 2000: 100). During this historic exodus, soldiers from all over China came to Taiwan. Already before the WWII, the official language of the ROC was Mandarin. After their retreat to Taiwan, the nationalists held on to this policy and drastically restricted the use of Taiwanese, the native language of about 73% of its population, as well as of Hakka and the aboriginal Austronesian languages spoken on Taiwan (Klöter 2005: 1-2).

In the past, the Tsiang$^{1\text{-}7}$tsiu^{1} and Tsuan$^{5\text{-}7}$tsiu^{1} dialects were clearly distinguishable from each other, but in the last couple of decades, the speakers of these dialects have been intermingling due to a reduction of the rural-urban gap. Within Taiwan, the movement of speakers from one dialect area to another is one of the causes of the change which varieties of Southern Mǐn spoken in Taiwan have undergone (Lín 2008: 5-6). For some fifteen years, a small group of mainly Western scholars have started using different terminology for the different dialects of Southern Mǐn that are referred to as a group with the name *Taiwanese*. These scholars think the names Tsiang$^{1\text{-}7}$tsiu^{1} dialect and Tsuan$^{5\text{-}7}$tsiu^{1} dialect are not appropriate anymore since these are the names of the ancestor dialects spoken in the respective cities in Hok$^{4\text{-}8}$kian^{3}. Although the Southern Mǐn varieties spoken in Taiwan originate in these cities, they have been developing independently since the 17th or 18th century (Klöter 2005: 3-4).

Table 1: Japanese loanwords in Taiwanese

Taiwanese		**Japanese**		
Form	**Meaning**	**Form**	**Meaning**	***Kanji*/*Kana***
ban^{3}lian^{7}pit^{4}	'fountain-pen'	*mannenhitsu*	'fountain-pen'	万年筆
bian^{3}tong^{1}	'lunch box'	*bentō*	'lunch box'	弁当
ne^{2}tsiang^{7}	'waitress, female shopkeeper'	*jochū*	'female servant, maid'	女中
oo^{7}ba^{1}sang^{2}	'flibbertigibbet, middle-aged woman'	*obasan*	'aunt'	伯母さん
bi^{2}lu^{3}	'beer'	*bīru* (< Du. *bier*)	'beer'	ビール
phang^{2}	'bread'	*pan* (< Pt. *pão*)	'bread'	パン

One of the more striking characteristics of Taiwanese is the large-scale adoption of Japanese loanwords, which the dialects of Tsiang$^{1\text{-}7}$tsiu^{1} and Tsuan$^{5\text{-}7}$tsiu^{1} have not adopted. This mainly happened from 1895 till 1945, when Taiwan was under Japanese control. The examples in Table 1 are from Lín (2008: 123-126).

The Japanese words in Table 1 are given both in the romanization system *Rōmaji* and in their customary Japanese orthography in *kanji* 漢字 'Chinese characters' and/or *kana* 仮名 'Japanese syllabic writing'. The final two words *bi^{2}lu^{3}* 'beer' and *phang^{2}* 'bread' are examples of words of Western origin, Dutch and Portuguese respectively, that have entered the Taiwanese language through Japanese.

1.2 Motivation

The present study is my thesis for the Research Master in linguistics at Leiden University. I chose to write about the Taiwanese language for a number of reasons. First of all, in the fall of 2012 and the spring of 2013, I studied a year at the National Taiwan Normal University in Taipei. My main goal there was Mandarin acquisition, but during my stay there I came into contact with the fascinating language that the people there call *Tai$^{5\text{-}7}$gi^{2}* 'Taiwanese'; it is "their" language. In the fall of 2012, I took a course in Taiwanese where I learned the pronunciation and the Tai$^{5\text{-}7}$lo^{5} transcription system. When I returned to the Netherlands, I met a Taiwanese student of linguistics whose mother tongue was Taiwanese. It was from him that I learned to speak Taiwanese to the extent of being able to use Taiwanese to communicate and engage in ordinary conversation. The following year, in the summer of 2014, I went to the village of 柳營 *Liu$^{2\text{-}1}$iann^{5}* in 臺南直轄市 *Tai$^{5\text{-}7}$lam$^{5\text{-}7}$ tit$^{8\text{-}4}$hat$^{4\text{-}8}$tshi^{7}* the District of Tainan where I lived with a mother and her son for five weeks in order to improve my conversation skills. I learned a lot from Khóo Lāu-su, who on his own made the whole trip, and its problems, worth while.

These past few years Taiwan, its people and their languages, most notably Taiwanese Southern Mĭn, have come to hold a special place in my studies. It makes me sad that the future of Taiwanese, as is the case with most of the languages of Taiwan, is as uncertain as it is. I hope this thesis may serve to contribute to the documentation and preservation of the Taiwanese cultural heritage.

The idea for the topic of this thesis came to me when I needed to translate a Dutch text into Mandarin for the school board of the Algemene Chinese School Utrecht in my capacity of teacher of Mandarin there. I noticed that I often used the verb *shuō* 'say' following another verb, a

collocation which other teachers there did not use in their conversations with me. This was the first time that it struck me that, in this respect, the Mandarin I had learned in Taiwan somehow differed from that of the other teachers, who were all from China. I subsequently noticed extensive use of a similar collocation in a Taiwanese language television series I watched. There, it was *kong*^{2} ~ *kong*$^{2\text{-}1}$ 'say', the equivalent of Mandarin *shuō*, which followed the verb. It is because of these observations that I chose to write a description of the use of *kong*^{2} ~ *kong*$^{2\text{-}1}$ in TSM for my Master's thesis.

1.3 TSM phonology & Tai$^{5\text{-}7}$lo^{5}

For the transcription of Taiwanese Southern Mǐn, an adaptation of the Tai$^{5\text{-}7}$uan$^{4\text{-}5}$ Lo$^{5\text{-}7}$ma$^{2\text{-}1}$ji$^{7\text{-}3}$ Phing$^{1\text{-}7}$im$^{1\text{-}7}$ Hong$^{1\text{-}7}$an^{3} 臺灣羅馬字拼音方案 'Taiwanese Romanization System' (Tai$^{5\text{-}7}$lo^{5} in short) will be used. First, I will discuss TSM phonology and the way it is written in the Tai$^{5\text{-}7}$lo^{5} transcription system. Then, I will introduce my adaptations to Tai$^{5\text{-}7}$lo^{5}, as used in this article. The introduction to the Tai$^{5\text{-}7}$lo^{5} transcription system will for a large part be based on MoE (2011a). Unless stated otherwise, my phonological analysis follows Klöter (2005: 7-17).

Tones

TSM is a tonal language with eight phonemic tones. Minimal and near-minimal pairs are listed in Table 2. When transcribing tones phonetically (i.e. when they are written between square brackets), I will use the numbers 1 till 5 to indicate the relative start and end pitch of the tone, 1 representing the lowest pitch and 5 the highest. This notation system was devised by Yuen Ren Chao for the transcription of tones in all Sinitic languages (Chao 1930). Consider the TSM word for 'bed' [bin^{33}ts^{h}ŋ^{14}]. The first syllable has a medium pitch, indicated by the number 3. Both the first and the last digit are a 3, which means that the starting and ending pitch are the same. This indicates that the medium pitch does not change and the result is a medium level pitch. The second syllable starts low (indicated as 1) and ends mid-high (4), it thus has a low rising contour. If the two digits are underlined, it means that the syllable is very short and abrupt. These syllables are the two *jip*^{8} tone syllables (more details below).

Three things should be noted in the numbering of the tones. First of all, the numbering implemented here is the one commonly used for Southern Mǐn tonal systems. The TSM tonal system consists of seven tones which derive from a four-way division reconstructed for Middle Chinese. In traditional Chinese phonology (*im*$^{1\text{-}7}$*un*$^{7\text{-}3}$*hak*^{8} 音韻學), the names of these four tones are

平 *ping*^{5}, 上 *siong*^{2}, 去 *khi*^{3}, 入 *jip*^{8}. This four-way tonal system eventually split into a higher register, 陰調 *im*$^{1\text{-}7}$*tiau*^{7}, and a lower register, 陽調 *iong*$^{5\text{-}7}$*tiau*^{7}, resulting in a system with eight phonetic tones. The registers originated in the relative height of pitches. Syllables which started with voiced consonants were consistently pronounced at lower pitch than syllables that started with

Table 2: TSM tones

No.	TSM name	English name	Phonemic	Phonetic	Meaning
1	*im*$^{1\text{-}7}$*ping*^{5} 陰平	High Level tone	/kaŋ^{1}/	[kaŋ^{55}]	‘day’
2	*siong*^{2} 上	Rising[3] tone	/kaŋ^{2}/	[kaŋ^{52}]	‘chat’
3	*im*$^{1\text{-}7}$*khi*^{3} 陰去	High Falling tone	/kaŋ^{3}/	[kaŋ^{31}]	‘descend’
4	*im*$^{1\text{-}7}$*jip*^{8} 陰入	High Entering tone	/kak^{4}/	[kɑk$^{\neg 32}$]	‘horn’
5	*iong*$^{5\text{-}7}$*ping*^{5} 陽平	Low Level tone	/kaŋ^{5}/	[kaŋ^{14}][4]	‘the same’
7	*iong*$^{5\text{-}7}$*khi*^{3} 陽去	Low Falling tone	/kaŋ^{7}/	[kaŋ^{33}]	‘play tricks on’
8	*iong*$^{5\text{-}7}$*jip*^{8} 陽入	Low Entering tone	/kak^{8}/	[kɑk$^{\neg 55}$]	‘compete’
0	*khin*$^{1\text{-}7}$*siann*^{1} 輕聲	Neutral tone	/a^{0}/	[a^{31}]	‘PERF’

voiceless consonants. Presumably, this difference in pitch height gradually gained phonological prominence, while in most varieties voiced stops, affricates and fricatives lost their voicedness and merged with their voiceless counterparts, resulting in an eight-way phonemic tonal system (Norman 2012: 53). This happened in most Sinitic languages, with the *Wú* 吳 varieties as a notable exception (Yip 2002: 185-189).

In the transcription of Southern Mĭn tones, the numbers 1 to 4 are used for the four tonal categories in the higher register, and the numbers 5 to 8 are used to refer to the same four tonal categories in the lower register. This is the traditional numbering used by Western missionaries describing Chinese dialects. It differs from the numbering commonly used in the comparative description of Sinitic languages, which first gives the higher and lower register of the *ping*^{5} 平 tone (1 & 2), then those of the *siong*^{2} 上 tone (3 & 4), the *khi*^{3} 去 tone (5 & 6) and the *jip*^{8} 入 tone (7 & 8). In synchronic descriptions of Southern Mĭn the first system, sometimes termed the missionary system (Klöter 2005: 7), is used. I will follow this missionary system here. The way the two

3 Note that the English names are translations of the Chinese names, which in turn reflect Middle Chinese in which the *siong*^{2} tone appears to have had a ‘rising’ contour. In modern TSM, however, the Rising tone has a falling contour, the Low Level tone a rising contour and the Low Falling tone a level contour.

4 On the basis of my own observations, I provide tone values for [kaŋ^{14}, kɑk^{55}, a^{31}], different from Klöter (2005: 9, 11).

systems use numbers to transcribe the eight tones recognised by traditional Chinese phonology is presented in Table 3.

Table 3: Tone numbering

TSM name	***im*$^{1\text{-}7}$*ping*^{5}**	***im*$^{1\text{-}7}$*siong*^{2}**	***im*$^{1\text{-}7}$*khi*^{3}**	***im*$^{1\text{-}7}$*jip*^{8}**	***iong*$^{5\text{-}7}$*ping*^{5}**	***iong*$^{5\text{-}7}$*siong*^{2}**	***iong*$^{5\text{-}7}$*khi*^{3}**	***iong*$^{5\text{-}7}$*jip*^{8}**
Missionary system	1	2	3	4	5	6	7	8
Comparative system	1	3	5	7	2	4	6	8

Secondly, because in most Southern Mĭn varieties, the *iong*$^{5\text{-}7}$*siong*^{2} 陽上 tone, i.e. the *siong*^{2} 上 tone of the lower register, tone 6 of the missionary system, has partially merged with the *im*$^{1\text{-}7}$*siong*^{2} 陰上 and partially with the *iong*$^{5\text{-}7}$*khi*^{3} 陽去 tones and subsequently ceased to exist (Yáng 1991: 34), by way of convention the number 6 is skipped when marking tones. Seeing that the present description is synchronic in nature, those original *iong*$^{5\text{-}7}$*siong*^{2} 陽上 tones that have merged with the *im*$^{1\text{-}7}$*siong*^{2} 陰上 tone are indicated with the number 2, and those that have merged with the *iong*$^{5\text{-}7}$*khi*^{3} 陽去 tone with the number 7. In the present study, my approach will be strictly synchronic and I will not further specify whether an *im*$^{1\text{-}7}$*siong*^{2} or *iong*$^{5\text{-}7}$*khi*^{3} tone derives from a *iong*$^{5\text{-}7}$*siong*^{2} tone.

Finally, in its analysis of tones, traditional Chinese phonology does not discuss the so-called Neutral tone, *khin*$^{1\text{-}7}$*siann*^{7} 輕聲. In TSM, however, minimal pairs such as *au*$^{7\text{-}3}$*jit*^{8} 'hereafter' and *au*^{7}*jit*$^{8\text{-}0}$ 'day after tomorrow', and *kong*$^{2\text{-}1}$ *nng*$^{7\text{-}3}$*ku*^{3} 'utter two sentences' and *kong*^{2} *nng*$^{7\text{-}0}$ *ku*$^{3\text{-}0}$ 'to utter a couple of sentences' bare witness to the fact that the Neutral tone can engender phonological contrasts and therefore lead to a difference in meaning. To me, if two semantically different lexical items can only be distinguished by a single phonetic element – be it a difference in vowel quality, a different consonant or a difference in tonal contour – this element must be considered phonemic. In light of such minimal pairs, I consider the Neutral tone a phonemic tone, as shown in Table 2. The Neutral tone has been assigned the number 0. For more information on the phonetic value of the Neutral tone and the details about its occurrence, the reader is referred to Zhèng (1994).

In TSM, the phonetic value of a tone changes depending on the morphosyntactic context it appears in. This process is known as tone sandhi. Usually, all syllables except for the final syllable in a clause exhibit tone sandhi. In (3), a sample TSM sentence is transcribed phonetically. Below the gloss of every syllable that exhibits tone sandhi, the form of the tone when the syllable is spoken

in isolation is given. This tone will be referred to as the citation tone. The form of the tone when it has undergone tone sandhi will be called either the sandhi tone or the sandhied tone:

(3) [**a**^{33}iəŋ^{55} **bɤ̯**^{33} **ɕũ**^{31}**be**^{55} **ka**^{55} **gua**^{55} **tau**^{52}tin^{33} **kʰi**^{55} **ɕu**^{33} tswi^{52}]
FAM-PNIng NEG.EXST want.to and 1SG together go swim water
[a^{11}] [bɤ̯^{13}] [ɕũ^{33}] [beʔ$^{\neg 32}$] [kaʔ$^{\neg 32}$] [gua^{52}] [tau^{31}] [kʰi^{31}] [ɕu^{14}]

'A-ing does not want to go swimming with me.'

If we compare the first row in (3) with the third row, we see that out of twelve syllables, nine exhibit tone sandhi – these have been highlighted in bold. The exact rules for when tone sandhi occurs and what phonetic form the tone exhibiting tone sandhi has are complex. Here, I will only discuss the TSM tone sandhi circle, which forms the basis of the tone sandhi system (Kuo 2013: 3).

Figure 4: TSM tone sandhi circle

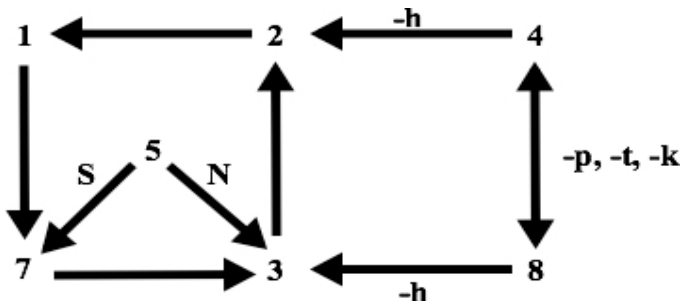


Figure 4 shows the TSM tone sandhi circle. In TSM, whenever a second tone syllable sandhies, it sandhies to a high level contour which is usually identified with the first tone. Syllables with a third tone sandhi to a high falling contour which is identified with the second tone. Syllables with a seventh tone sandhi to a low falling contour which is identified with the third tone. And syllables with a first tone sandhi to a mid level contour which usually is identified with the seventh tone. The fifth tone shows regional variation: in the south of Taiwan (S in Figure 4), it sandhies to a mid level contour, the seventh tone; but in the north (N), it sandhies to a low falling contour instead, the third tone. The tones of the *jip*^{8} category sandhi to the opposite register when the syllable ends in /-p, -t, -k/: the fourth tone sandhies to a short high tone (the eighth tone), whereas the eighth tone sandhies to a short low tone (the fourth tone). If however, the syllable with a fourth tone ends in a glottal stop /-h/ ([-ʔ$^{\neg}$]), it sandhies to a high falling contour, which is identified with the second tone, whereas the eighth tone sandhies to a low falling contour, which is identified with the third tone. In my

experience, in these last two cases, the glottal stop gets lost in the sandhi process. However, there are reports of dialects where the glottal stop is not lost in the process (Klöter 2005: 9).

It should be noted that it is common practice to identify the sandhied tone of the first tone with the seventh tone, etc. Studies show, however, that there are phonetic differences between an original tone and a sandhi tone of the same contour. The differences are that original tones are generally longer in duration, have a creakier quality of voice and a wider F0 range. Experiments have shown that native speakers are able to distinguish between original tones and sandhi tones on the basis of these phonetic differences even though the tonal contours are basically the same (Kuo 2013: 111-117). It is not clear from these studies which part of the speech community is involved. Kuo's findings seem to suggest a phonemic difference between an original tone and a sandhi tone of similar contour. However, in the absence of studies geared towards the different tonemes – i.e. phonemic tones – in TSM, in this study, I will follow the more common practice of identifying the sandhi tone of the third tone with the second tone and the sandhi tone of the second tone with the first tone, etc., as is shown in the sandhi tone circle in Figure 4.

Table 4 gives examples of the tone sandhi patterns discussed above.

Table 4: Tone Sandhi

No.	**In isolation**	**Meaning**	**Sandhi form**	**Meaning**
1	[tɕʰɪəŋ^{55}]	'clear'	[tɕʰɪəŋ^{33} tsau^{52}kʰa^{55}]	'clean the kitchen'
2	[kɔŋ^{52}]	'say'	[kɔŋ^{55} ue^{33}]	'speak'
3	[kɔŋ^{31}]	'hit with sth. heavy'	[kɔŋ^{52} pɑŋ^{31}kju^{14}]	'play baseball'
4 (-p, -t, -k)	[tɕit$^{\underline{32}}$]	'this'	[tɕip$^{\underline{55}}$ bwe^{55} hi^{14}]	'this fish'
4 (-h)	[aʔ$^{\underline{32}}$]	'duck'	[a^{52}kɑŋ^{55}]	'drake'
5 (south)	[kʰja^{14}]	'ride'	[kʰja^{33} be^{52}]	'ride a horse'
5 (north)	[kʰja^{14}]	'ride'	[kʰja^{31} be^{52}]	'ride a horse'
7	[twa^{33}]	'big'	[twa^{31} ljap$^{\underline{32}}$ ɕi^{33}kwe^{55}]	'a big water melon'
8 (-p, -t, -k)	[ap$^{\underline{55}}$]	'box (M)'	[ab$^{\underline{32}}$a^{52}]	'box (N)'
8 (-h)	[taʔ$^{\underline{55}}$]	'step on'	[kʰa^{33}ta^{31}tɕʰa^{55}]	'bicycle'

Apart from the basic tone sandhi circle, I will mention one additional special case of tone sandhi here as well, since its transcription involves the additional number 9, which I have used in the transcription of my corpus on occasion. The contour of this sandhi form is a high rising pitch, as shown in Table 5. This contour manifests itself in triple-adjective constructions and in the merger of certain syllables.

Table 5: Triple-adjective constructions

No.	Adjective	Meaning	Triple adjective	Meaning
1	[ɕɻ^{55}]	‘hot’	[ɕɻ^{35} ɕɻ^{33} ɕɻ^{55}]	‘extremely hot’
3	[p^{h}ɔŋ^{31}]	‘round and elastic’	[p^{h}ɔŋ^{35} p^{h}ɔŋ^{35} p^{h}ɔŋ^{31}]	‘extremely round and elastic’
5	[ɑŋ^{14}]	‘red’	[ɑŋ^{35} ɑŋ^{33} ɑŋ^{14}]	‘extremely red’
7	[lɪəŋ^{33}]	‘loose’	[lɪəŋ^{35} lɪəŋ^{31} lɪəŋ^{33}]	‘extremely loose’
8	[peʔ$^{\underline{55}}$]	‘white’	[pe^{35} pe^{31} peʔ$^{\underline{55}}$]	‘extremely white’

Triple adjective-constructions are constructions in which an adjective is reduplicated twice, giving the meaning of ‘extremely [adjective]’. The high rising tone manifests itself on the first of the three syllables of twice reduplicated adjectives of which the isolation tone is the first, fifth, seventh or eight tone. When adjectives which in isolation are pronounced with a third tone are reduplicated twice, the first two syllables sandhi to this high rising contour (Yáng 1991: 142-143). Just like Tai$^{5\text{-}7}$lo^{5}, I use the number 9 to indicate this high rising tone (see Table 12).

Apart from the triple adjective construction, the high rising tone is also manifested on syllables that have come about through the merger of two syllables: *bua^{9}kin^{2}* ‘there is nothing serious’ < *bo$^{5\text{-}7}$ iau$^{3\text{-}2}$kin^{2}*, and *hong^{9} ~ hong^{2}* ‘give someone’ < *hoo$^{7\text{-}3}$ lang^{5}*.

For more information on tone sandhi in TSM, the reader is referred to Âng (2002).

Consonants

Tai$^{5\text{-}7}$lo^{5} uses seventeen different letters and letter combinations to mark the fourteen different phonemic consonants of TSM that can occur in syllable initial position. Table 6 shows these phonemes.

As shown in the first two columns of Table 6, the voiced stops /b, l, g/, the affricates /d͡z, t͡s, t͡s^{h}/ and the sibilant /s/ are all phonemes with two or multiple allophones. In the phonemic system of TSM, /l/ [l, d, n] functions as the voiced alveolar stop, corresponding to /b/ in the labial series and /g/ in the dorsal series, as shown in Table 7. Sometimes, these allophones are written in Tai$^{5\text{-}7}$lo^{5}, and sometimes not. For example, the voiced stops have nasal and oral allophones in complementary distribution, with nasal plosives followed by nasal vowels, and oral plosives by oral vowels exclusively. In Tai$^{5\text{-}7}$lo^{5}, nasal vowels that follow a nasal initial are not marked for nasality. Contrastive pairs can be seen in Table 7.

Table 6: Tai$^{5\text{-}7}$lo^{5} consonants in syllable initial position

Phonological form	**Phonetic form**	**Tai$^{5\text{-}7}$lo^{5}**
/b/	[b]	b
	[m]	m
/p/	[p]	p
/p^{h}/	[p^{h}]	ph
/l/	[l, d]	l
	[n]	n
/t/	[t]	t
/t^{h}/	[t^{h}]	th
/g/	[ɡ]	g
	[ŋ]	ng
/k/	[k]	k
/k^{h}/	[k^{h}]	kh
/h/	[h]	h
/d͡z/	[dz, dʑ]	j
/t͡s/	[ts, tɕ]	ts
/t͡s^{h}/	[ts^{h}, tɕ^{h}]	tsh
/s/	[s, ɕ]	s

Table 7: Oral and nasal allophones of TSM voiced stops

Oral	**Tai$^{5\text{-}7}$lo^{5}**	**Meaning**	**Nasal**	**Tai$^{5\text{-}7}$lo^{5}**	**Meaning**
[bi^{33}]	*bi^{7}*	‘smell, taste’	[mĩ^{33}]	*mi^{7}*	‘noodles’
[la^{33}]	*la^{7}*	‘stir, mix’	[nã^{33}]	*na^{7}*	‘if’
[ge^{33}]	*ge^{7}*	‘art’	[ŋẽ^{33}]	*nge^{7}*	‘hard’

The allophones [l] and [d] are both possible forms of the phoneme /l/ when occurring before oral vowels, and are in free variation there. Together, they form [l] and [d] on the one hand, in complementary distribution with [n] on the other hand, the latter occurring only before nasal vowels. The affricatives /d͡z, t͡s, t͡s^{h}/ and fricative /s/ have palatalized allophones [dʑ, tɕ, tɕ^{h}, ɕ] respectively when occurring before the high front vowel /i/.

In Table 8, a complete inventory of TSM phonetic (a), phonemic (b) and Tai$^{5\text{-}7}$lo^{5} (c) consonants is shown. The consonants that can occur syllable-finally have been added in brackets. Tai$^{5\text{-}7}$lo^{5} *-h*, *-p*, *-t*, *-k* represent unreleased stops [-ʔ̚, -p̚, -t̚, k̚] respectively. The phonemic status of the final consonants, especially of *-h*, *-m*, *-n* and *-ŋ*, is a point of ongoing debate. This is mainly

for two reasons: 1) complementairy distribution between [-p̚, -t̚, k̚] in syllables with *jip*^{8} tones and [-m, -n, -ŋ] in non-*jip*^{8} syllables, and 2) the status of initial [m, n, ŋ] as allophones of /b, l, g/ before nasal vowels.

Table 8: TSM consonants

PHONETIC		**Bilabial**	**Alveolar**	**Velar**	**Glottal**
Plosive & affricate	**unaspirated**	p, -p̚	t; ts, tɕ; -t̚	k, -k̚	-ʔ̚
	aspirated	p^{h}	t^{h}; ts^{h}, $tɕ^{h}$	k^{h}	
	voiced	b, m, -m	l, n; dz, dʑ; -n	g, ŋ, -ŋ	
Fricative			s, ɕ		h

PHONEMIC		**Bilabial**	**Alveolar**	**Velar**	**Glottal**
Plosive & affricate	**unaspirated**	p, -p	t, t͡s, -t	k, -k	
	aspirated	p^{h}	t^{h}, $t͡s^{h}$	k^{h}	
	voiced	b, -b	l, d͡z, -l	g, -g	
Fricative			s		h

TAI$^{5\text{-}7}$LO^{5}		**Bilabial**	**Alveolar**	**Velar**	**Glottal**
Plosive & affricate	**unaspirated**	p, -p	t, ts, -t	k, -k	-h
	aspirated	ph	th, tsh	kh	
	voiced	b, m, -m	l, n; j; -n	g, ng, -ng	
Fricative			s		h

Regarding the first reason, some authors argue in favor of an analysis of *jip*^{8} tones as phonemic and the glottal stop as being a phonetic feature of the realization of syllables with a *jip*^{8} tone ending in a vowel. In this analysis, a form such as *ah*^{4} 'duck' with the phonetic realization [aʔ̚$^{\underline{32}}$] is phonemically /a^{4}/. The form *at*^{4} 'break; scoop with a spoon' with the phonetic realization [at̚$^{\underline{32}}$] would be analyzed as /at^{4}/. The difference between the two being that phonemically, *ah*^{4} 'duck' ends in a vowel and *at*^{4} 'break; scoop with a spoon' ends in the consonant /t/. On the other hand, in this analysis *ah*^{4} 'duck' phonemically contrasts with the prefix *a*^{1}- 'FAM' in tone exclusively: /a^{4}/ 'duck' as opposed to /a^{1}/ 'FAM' (Chung 1996: 20-21). According to Chung (*ibid.*), Luó (1931) is a proponent of analyzing [-ʔ̚] as a phoneme. Chung does not, however, indicate where exactly Luó states this and I have been unable to trace it in the second edition of his work (1956). Perhaps Chung inferred this from the fact that in the coda, the glottal stop is present in the transcription

system devised by Luó. He indicates the glottal stop with <q> (1956: 32, 40). I will follow Chung here, in analyzing the glottal stop as a phonetic feature of the two *jip*^{8} tonemes in syllables ending in a vowel.

In analyzing the status of the initials [m, n, ŋ] as opposed to that of [b, l, g], Luó and Chung disagree as well. Luó (*ibid.*: 52) analyzes /m, n, ŋ, b, l, g/ as six different phonemes, whereas Chung (*ibid.*: 34-46) proposes that they the nasals are allophones of the oral stops. Here too, I follow Chung.

Finally, the consonants [m] and [ŋ] can constitute a syllable either with a tone or with a tone and other consonants, but without the need for vowels Examples of these are listed in Table 9. As discussed above, Chung (1996: 210-214) analyzes [m] and [ŋ] as allophones of /b/ and /g/ respectively. Here, I will follow his practice: /b, l, g/ are realized as [b, l, g] syllable initially before oral vowels and as [m, n, ŋ] before nasal vowels, syllable finally and – in the case of [m] and [ŋ] – when functioning as the syllable nucleus.

Table 9: [m] and [ŋ] constituting syllables

Phonetic form	**Tai$^{5\text{-}7}$lo^{5}**	**Meaning**	**Phonetic form**	**Tai$^{5\text{-}7}$lo^{5}**	**Meaning**
[m̥m̩ʔ$^{\overline{55}}$]	*hmh*^{8}	‘gloom’	[m̩^{33}]	*m*^{7}	‘NEG’
[sŋ̍^{55}]	*sng*^{1}	‘sour’	[pŋ̍^{33}]	*png*^{7}	‘cooked rice’
[ŋ̍^{14}]	*ng*^{5}	‘yellow’	[mŋ̍^{14}]	*mng*^{5}	‘door’

Vowels

Tai$^{5\text{-}7}$lo^{5} uses six vowel letters to write the oral vowels of TSM. They are listed in Table 10.

Table 10: Vowels in Tai$^{5\text{-}7}$lo^{5}

Phonological form	**Phonetic form**	**Tai$^{5\text{-}7}$lo^{5}**
/i/	[i, ɪə]	i
/e/	[e]	e
/a/	[a, ɑ]	a
/ɔ/	[ɔ]	oo
/o/	[o, ɤ̞]	o
/u/	[u]	u

Some of the vowels listed in Table 10 have systematic allophones: [ɑ] and [ɪə] are the allophones of /a/ and /i/ respectively before the final velar consonants /-k/ and /-ŋ/. Everywhere else, the

allophones [a] and [i] are used. There is regional variation between [ɤ] and [o], the former being used in Southern Taiwan (e.g. in the Districts of Tainan *Tai*$^{5-7}$*lam*$^{5-7}$ *tit*$^{8-4}$*hat*$^{4-8}$*tshi*^{7} 臺南直轄市 and Kaohsiung *Ko*$^{1-7}$*hiong*$^{5-7}$ *tit*$^{8-4}$*hat*$^{4-8}$*tshi*^{7} 高雄直轄市) and the latter being used in Northern Taiwan (e.g. in the Districts of Taipei *Tai*$^{5-7}$*pak*$^{4-8}$ *tit*$^{8-4}$*hat*$^{4-8}$*tshi*^{7} 臺北直轄市 and Taichung *Tai*$^{5-7}$*tiong*$^{1-7}$ *tit*$^{8-4}$*hat*$^{4-8}$*tshi*^{7} 臺中直轄市).

In addition to these six oral vowels, TSM has four nasal vowels. Table 11 lists these nasal vowels and provides minimal pairs with oral counterparts.

Table 11: Nasal and oral vowels in TSM

	Phonemic	**Tai$^{5-7}$lo^{5}**	**Meaning**
Nasal	/tĩ^{1}/	*tinn*^{1}	‘sweet’
	/sẽ^{3}/	*senn*^{3}	‘family name’
	/kã^{2}/	*kann*^{2}	‘dare’
	/kɔ̃^{5}/	*konn*^{5}	‘snore’
Oral	/ti^{1}/	*ti*^{1}	‘pig’
	/se^{3}/	*se*^{3}	‘small’
	/ka^{2}/	*ka*^{2}	‘holiday’
	/kɔ^{5}/	*koo*^{5}	‘paste’

In Tai$^{5-7}$lo^{5}, there are two ways of indicating nasal vowels. The first is to add <nn> immediately after the vowel. The form /sẽ^{3}/ ‘family name’ would thus in Tai$^{5-7}$lo^{5} be written <sènn>[5]. If the nasal vowel is followed by a consonant, <nn> is written following the vowel sign and preceding the consonant sign. The form /hĩã^{4}/ ‘carry’, phonetically [hj̃ãʔ$^{\neg 32}$], would thus in Tai$^{5-7}$lo^{5} be written <hiannh>.

If a nasal vowel is preceded by a voiced stop /b, l, g/, Tai$^{5-7}$lo^{5} writes the nasal allophones of the voiced stops <m, n, ng>, without indicating the nasality of the vowel in any other way. The form /bẽ^{7}/ ‘scold’ would in thus in Tai$^{5-7}$lo^{5} be written <mē> and the form /gẽ^{8}/ ‘wedge in’ would be written <ngėh>. In such cases, <m, n, ng> are used to mark the nasality of both the voiced stops /b, l, g/ and the vowels /i, e, a, ɔ/.

Adaptations

Below I will list the adaptations to the MoE Tai$^{5-7}$lo^{5} that I will apply in the transcription of my data.

5 For the diacritic <`> marking the third tone, see Table 12 below and the discussion there.

Tai$^{5-7}$lo^{5} uses diacritics to mark the eight tones of TSM. In light of the pervasiveness of tone sandhi in TSM (see above), I have opted to mark tones by way of numbers, as shown in Table 12 below.

Table 12: Transcription of Tones

No.	**MoE Tai$^{5-7}$lo^{5}**	**Adapted Tai$^{5-7}$lo^{5}**	**English**
1	*phang*	*phang^{1}*	'smell good'
2	*bé*	*be^{2}*	'buy'
3	*khì*	*khi^{3}*	'go'
4	*hmh*	*hmh^{4}*	'beat with something heavy'
5	*kuân*	*kuan^{5}*	'tall'
7	*hn̄g*	*hng^{7}*	'far'
8	*tsia̍h*	*tsiah^{8}*	'eat'
0	*--a*	*a^{0}*	'PERF'
9	*pe̍nn-pênn-pênn*	*penn$^{5-9}$penn$^{5-7}$penn^{5}*	'extremely flat'

For any morpheme displaying tone sandhi, the citation tone is followed by a hyphen and the sandhi tone. For example, the phrase *tsiah$^{8-3}$ png^{7}* 'eat' consists of the verb *tsiah^{8}* 'eat' and the noun *png^{7}* 'cooked rice'. When pronounced in isolation or before a pause, the word *tsiah^{8}* is pronounced in the eighth tone. However, because of the following word *png^{7}* 'cooked rice', the tone of *tsiah^{8}* sandhies to the third tone. In the present article, this will be written as <tsiah$^{8-3}$png^{7}>. Official Tai$^{5-7}$lo^{5}, as is used by the Ministry of Education (MoE), does not mark tone sandhi. The same phrase would be written as <tsia̍h pn̄g>.

Tai$^{5-7}$lo^{5} too, recognizes the existence of the Neutral tone. There, the Neutral tone is not marked by means of a diacritic, but by two hyphens preceding the syllable. The word for 'day after tomorrow' given above is thus written <ān--ji̍t> in MoE Tai$^{5-7}$lo^{5}. Note that the diacritic <'> which marks the citation tone of the morpheme is retained on syllables with a Neutral tone and that whenever it appears on the letter <i>, it replaces the dot, i.e. <--ji̍t> instead of <--jit>. In this article, however, I will use the double hyphen to indicate hesitation by the speaker or interruption by an interlocutor. I will use the numer 0 to indicate a neutral tone. A word such as *ān--ji̍t* in MoE Tai$^{5-7}$lo^{5}, I will transcribe as *au^{7}jit$^{8-0}$*.

Tai$^{5-7}$lo^{5} uses hyphens to separate syllables that belong to the same written word, as in <thâu-mn̂g> 'hair', and spaces to separate words from each other, as in <thâu kap mn̂g> 'head and hair'. In this article too, spaces are used to separate words from each other. I will use the following

provisional definition of a word: a morpheme or cluster of morphemes that can be used independently in an utterance without the need for other morphemes. No hyphens, however, will be used to separate syllables within the same word. This is because the superscript numbers which have been employed here to mark tones already clearly divide a word into syllables, so 'hair' will be written as <thau$^{5\text{-}7}$mng^{5}>.

The above described adaptations to the Tai$^{5\text{-}7}$lo^{5} transcription system used in this article do not apply to TSM words in English translations. This goes for personal names, as well as for cases where the speakers are talking about TSM words and expressions themselves – i.e. where they use metalanguage – as in example (4):

(4) A^{0} soo$^{2\text{-}1}$i$^{2\text{-}1}$ toh$^{8\text{-}3}$ **kong$^{2\text{-}1}$-tioh$^{8\text{-}3}$** hit$^{4\text{-}8}$ e$^{5\text{-}7}$ tshenn$^{1\text{-}7}$me$^{5\text{-}7}$ gu^{5}, sionn$^{7\text{-}3}$ **kong$^{2\text{-}1}$** lan$^{2\text{-}1}$
RLT therefore just **speak-get** that entity blind buffalo want.to **speak** 1INCL

le^{1} tshi$^{7\text{-}3}$ gu^{5}, an$^{2\text{-}1}$ne^{1} hit$^{4\text{-}8}$ e$^{5\text{-}7}$ gu^{5} i$^{1\text{-}7}$ toh$^{8\text{-}3}$ ah-- m$^{7\text{-}3}$ bat$^{4\text{-}8}$-- m$^{7\text{-}3}$ bat$^{4\text{-}8}$
PROG feed buffalo like.this that entity buffalo 3 just HST NEG familiar.with NEG familiar.with

ji^{7} ma^{0} <si^{7} a^{0}> honn^{0}.
character WARN be PERF XPC

'So when we come to speak of [the term] *tshenn-mê gû* 'blind buffalo', we mean to say that the buffalos we are raising, such a buffalo, he just uhm-- doesn't know-- doesn't know characters, right <oh, yes>. (21C; 10:50)

Here, the speaker is explaining why in TSM one is said to be a *tshenn$^{1\text{-}7}$me$^{5\text{-}7}$ gu^{5}* 'blind buffalo' when they are illiterate. The speaker is therefore talking about the expression *tshenn$^{1\text{-}7}$me$^{5\text{-}7}$ gu^{5}* itself, which is why, in the English translation, *tshenn$^{1\text{-}7}$me$^{5\text{-}7}$ gu^{5}* is written in unadapted Tai$^{5\text{-}7}$lo^{5} using diacritics and hyphens.

For the transcription of Mandarin, I use the *Hànyǔ Pīnyīn* 漢語拼音 romanization system. For an introduction to Hànyǔ Pīnyīn, the reader is referred to Pīnyīn (2010). For the transcription of other Sinitic languages, I follow the system used in my sources for these languages.

More generally, in the written presentation of both the Southern Mĭn and the Mandarin data, my aim is to remain as true to the original spoken data as possible. This means that when speakers hesitate, make mistakes, repeat themselves, code-switch, start a new sentence or get interrupted, these utterances have been transcribed, and an effort has been made to preserve these hesitations, repetitions and mistakes in the English translations as well.

In (5) we can find examples of hesitation, repetition and backchannel. The English translation tries to incorporate these as well. The translation is, therefore, not meant as an

idealization of what the speaker should have said, but aims to present what I feel the speaker would have said in English, including hesitations and rephrasings.

(5) Lan$^{2-1}$ sing$^{1-7}$-- lan$^{2-1}$ sing$^{1-7}$ lai$^{5-7}$ **kong$^{2-1}$** hit$^{4-8}$ e$^{5-7}$-- <si^{7}> **kong$^{2-1}$** thau$^{5-7}$a^{2} **kong^{2}** e$^{5-7}$
1INCL first 1INCL first come **speak** that entity be **speak** at.the.start **speak** SUB

tshenn$^{1-7}$me$^{5-7}$ gu^{5} la^{0} <hann^{0}> honn^{0} <si^{7}>.
blind buffalo WARN CFM XPC be

'Let's first-- let's first talk about that-- <yes> about *tshenn-mê gû* 'blind buffalo' (19C; 10:33)
we talked about at the beginning, <hm> right <yes>.'

Given that the aim is to record the data as accurately as possible, when a speaker pronounces a word non-canonically, this form is documented, regardless of any corresponding canonical dictionary form. I take Lôo (2011) as the standard for the canonical Taiwanese forms and Zhāng (2009) for the Mandarin forms. In these cases, instead of writing down a canonical form, my record of the way the word was spoken is confined only by the limits of Tai$^{5-7}$lo^{5} and Hànyǔ Pīnyīn. An example for Taiwanese can be found in (2), repeated below, and for Mandarin in (6):

(2) A^{0} soo$^{2-1}$i$^{2-1}$ kho$^{2-1}$ling^{5} *=ne*, eh-- koh$^{4-1}$ ui$^{7-3}$ thiann$^{1-7}$tsiong$^{3-2}$ ping$^{5-7}$iu^{2} e$^{7-3}$ kam$^{2-1}$kak$^{4-8}$
RLT therefore maybe =RLV HST every place listening.audience friend will feel

kong$^{2-1}$ m^{0} khak$^{4-8}$sit^{8} *=ne*, tsin$^{1-7}$tsiann^{2} **tsainn$^{1-7}$iann$^{2-1}$--** beh$^{4-1}$ **tsainn$^{1-7}$iann$^{2-1}$**
say AGR indeed =RLV really **know** FUT **know**

i$^{1-7}$ e$^{5-7}$ i$^{3-2}$su^{3}.
3 SUB meaning

'So it may be that, uhm-- our listeners will feel that, hm yes, they really know-- (3A; 05:40)
want to know its meaning.'

(6) Siunn$^{7-3}$beh$^{4-1}$ oh^{8} *=ne*, ha^{0}, lang^{5} long$^{2-1}$-- tsong$^{2-1}$si$^{7-3}$ hoo^{0}, eh-- be$^{7-3}$tang$^{3-2}$
want.to learn =RLV RLT person in.every.case always XPC HST NEG:be.able.to

kong$^{2-1}$ eh-- s-- ***nóu-chū*** *zhèi yàng =de shíjiān chū lái*.
speak HST HST **move-exit** this kind =SUB time exit come

'If you want to learn it, ow, people can always-- never, you know, eh-- (13A; 09:09)
say eh-- s-- muster up that sort of time.'

Canonically, the word for ‘know’ in TSM is *tsai*$^{1-7}$*iann*$^{2-1}$, with the first syllable pronounced orally (Lôo 2011: 551). In (2), the speaker twice says *tsainn*$^{1-7}$*iann*$^{2-1}$ ‘know’ with two nasal syllables. Presumably, this is because the nasality from the surrounding syllables has spread to *tsai*$^{1-7}$.

In (6), the speaker uses the word *nóuchū* ‘move out’. The standard pronunciation for this word would be *nuóchū*, with a closed back on-glide instead of the closed back off-glide the speaker produces (Zhāng 2009: 591).

The one thing that could not be transferred to my translations are instances of code switching between TSM and Mandarin. When reading the examples, the transcribed and glossed utterances should be considered primary. The English translations are only there to assist the reader in understanding these data.

1.4 Methodology

The description presented in this study has been composed on the basis of a corpus of 56 example sentences in which *kong*^{2} or its sandhi form *kong*$^{2-1}$ appears, and which I have transcribed using the Tai$^{5-7}$lo^{5} system described in § 1.3. The source for my corpus is the episode 毋通做台語的青盲牛：台語文書寫 *M*$^{7-3}$ *thang*$^{1-7}$ *tso*$^{3-2}$ *Tai*$^{5-7}$*gi*^{2} *e*$^{5-7}$ *tshenn*$^{1-7}$*me*$^{5-7}$*gu*^{5}*: Tai*$^{5-7}$*gi*$^{2-1}$*bun*^{5} *tsu*$^{1-7}$*sia*^{2} ‘Don’t be illiterate in Taiwanese: Writing Taiwanese’ of the radio show 做伙守台語 *Tso*$^{3-2}$*hue*^{2} *tsiu*$^{2-1}$ *Tai*$^{5-7}$*gi*^{2} ‘Looking after Taiwanese together’ by the 臺灣母語聯盟 Tai$^{5-7}$uan$^{5-7}$ Bu$^{2-1}$gi$^{2-1}$ Lian$^{5-7}$bing^{5} ‘Alliance for the Native Languages of Taiwan’. The whole episode spans 57:03 minutes. It was aired on Saturday, August 13, 2016 from 20h00 to 21h00 on 綠色和平電台 *Lik*$^{8-4}$*sik*^{4} *ho*$^{5-7}$*ping*^{5} *tian*$^{7-3}$*tai*^{5} ‘GreenPeace Broadcasting Station’.[6] In the episode used here, four Taiwanese talk about the different ways of writing Taiwanese and the problems people can encounter when they want to write Taiwanese. I have transcribed all instances of *kong*^{2} ~ *kong*$^{2-1}$ from 04:50 up to 17:50. All 56 sentences are listed chronologically in the Appendix.

In the main body of this thesis, most examples are presented in three different layers of analysis, as in (7).

(7)	Long$^{2-1}$	m$^{7-3}$	si$^{7-3}$	tsiann$^{3-2}$thong^{2}--	m--	**m--**	**gua$^{2-1}$--**	tui$^{3-2}$	gua$^{2-1}$	lai$^{5-7}$	kong^{2}	la^{0}	hoo^{0},
	in.every.case	NEG	be	orthodox	HST	**HST**	**1SG**	DIR	1SG	come	say	WARN	XPC

6 Even though their names are remarkably similar, this radio station is not affiliated with the environmental organization Greenpeace. More information on this radio station can be found on their website at www.greenpeace.com.tw.

long$^{2-1}$ sia^{2} tshut^{0} lai^{0}, long$^{2-1}$ m$^{7-3}$ si$^{7-3}$-- kam$^{2-1}$kak$^{4-8}$ m$^{7-3}$ si$^{7-3}$ Tai$^{5-7}$gi^{2} =le^{0},
in.every.case write exit come in.every.case NEG be feel NEG be Taiwanese =RLV

soo$^{2-1}$i$^{2-1}$ **<tioh^{8} a^{0}>** ***háisì*** *lǎosī* ai$^{3-2}$ kong$^{2-1}$ khah$^{4-2}$ tshing$^{1-7}$tsho^{2} <henn^{0}>
therefore **right** **PERF** **would.be.better** teacher have.to say relatively clear CFM

e$^{5-0}$ hoo^{0}.
SUB XPC

'It's all not the right-- uhm-- uhm-- I-- to me, you know, when you write in them they all don't give-- it feels like it's all not Taiwanese, so <right> it is still necessary that you make it a clearer <yes>, you know.' (18A; 10:24)

The first layer is a transcription of the spoken data. As discussed in the preceding section, I have transcribed tone sandhi whenever it occurred, e.g. *gua$^{2-1}$* '1SG'. I use a double hyphen <--> to indicate hesitation on the part of the speaker, e.g. *m*--. Code-switching Mandarin utterances have, within the limits of the Hànyǔ Pīnyīn romanization system, been transcribed as spoken, even when this diverged from the canonical form of words, e.g. *háisì* 'would be better' instead of *háishì*. Finally, interruptions and backchannel from other people have been transcribed in between angle brackets, e.g. <tioh^{8} a^{0}> 'right'.

The second layer is that of the glosses. Glosses are one-to-one translations of the morphemes they are paired with vertically. As opposed to the data of the first layer, there is no horizontal relation between different glosses of the second layer. Abbreviations used in the glosses can be found in the list Abbreviations & symbols above.

The third and final layer is an English translation of the Taiwanese sentence. As stated in § 1.3, the English translation aims to be good English on the one hand, while corresponding as best as possible to the Taiwanese sentence of layer one.

Whenever I needed to use footnotes in examples, I invariably appended these to the glosses and not to the Taiwanese words, in order to avoid confusion with tone numbers.

All example sentences have been numbered numerically in order of first appearance. For example, when an example is presented for the first time, it receives a number which is written between brackets in front of the first line. When an example is given for the second or third time, in order to avoid that the reader has to go back and forth through the pages, it is preceded by the number it was given when it was presented the first time. At the end of most examples, at the right of the English translation, the number of the example as presented in the Appendix has been given, followed by a semicolon and the starting time of the utterance in the audio file. For example, in (7), we have (18A; 10:24). This means that it is example number 18 in the Appendix and that its speaker is Speaker A; see the explanation at the start of the Appendix. The time notation 10:24 means that

the reader has to set the audio file “Audio Pronk 2017” at 10:24 in order to hear to the original spoken data. The “Audio Pronk 2017” can be found under the following link:
https://drive.google.com/open?id=0B477ABtMKQ8YUzNYRC1XUlB3YWc

Sometimes, example sentences do not have an appendix number following the translation. These examples are not from my corpus. The English and Dutch example sentences have been devised by me. Some of the Taiwanese examples have been devised by me as well, and were checked with a native speaker. Other Taiwanese examples have been taken from the literature. When this is the case, the source has been stated.

I consulted with native speakers whenever I was not certain about what was said in the spoken material. I have also consulted native speakers for grammaticality judgements. Whenever necessary to avoid influencing or biasing the consultants, I have formulated my questions in English instead of Taiwanese or Mandarin.

2 *Kong*^{2}

2.1 *Kong*^{2} as a transitive verb

2.1.1 Data & analysis

In the introduction, two examples of the use of word *kong*^{2} in TSM were given. The first of those, example (1), is repeated below:

(1) A^{0} ka$^{3-2}$ tsit$^{4-8}$ tsun^{7}, eh-- eh-- tau$^{3-2}$te^{2} =*ne*, lan$^{2-1}$ i$^{2-1}$tsing$^{1-7}$ u$^{7-3}$ **kong$^{2-1}$** =kue$^{3-2}$
RLT arrive this moment HST HST in.the.end =RLV 1INCL already^{7} EXST **say** =EXP

siann$^{2-1}$mih$^{8-3}$ hue^{7}?
what thing

'So and up to this moment eh-- eh-- what exactly have we already talked about?' (5A; 06:23)

In this example, *u*$^{7-3}$ *kong*$^{2-1}$*=kue*$^{3-2}$ translates to the English 'have talked about'. It thus corroborates what both Douglas (1899: 246) and Lôo (2011: 191) say about *kong*^{2}, viz. that the meaning of *kong*^{2} is 'say, talk' (Douglas) and 'speak, talk, utter, say' (Lôo).

Syntactically, *kong*^{2} 'say' is a transitive verb constructed with an object, *siann*$^{2-1}$*mih*$^{8-3}$ *hue*^{7} 'what thing'. *Kong*$^{2-1}$ is preceded by the auxiliary verb *u*$^{7-3}$ 'EXST', indicating perfective aspect. In TSM, auxiliaries usually come before the verb phrase they modify.[8] Finally, in (1), there is the enclitic *=kue*$^{3-2}$ 'EXP' which denotes the experiential aspect.

In TSM, transitive verbs are verbs that require a direct object. A direct object refers to the thing that is directly involved in the action, but never as the agent. The most common function of the direct object is that of patient, the undergoer of the action. In example (7), *tsiah*^{8} 'eat' is constructed with the direct object *bah*^{4} 'meat'.

(7) Gua$^{2-1}$ tsiah$^{8-3}$ bah^{4}.
1SG eat meat

'I am eating meat.'

7 Note that the speaker uses *i*$^{2-1}$*tsing*^{1} to mean 'already'. The regular Taiwanese pronunciation is *i*$^{2-1}$*king*^{1}. This is probably due to influence of the Mandarin form *yǐjīng* 'already'.

8 A possible exception to this rule could be the potential mood expressed through the morphemes *u*^{7} and *e*^{7} in the affirmative and *bo*^{5} and *be*^{7} in the negative, all of which come between a verb and its resultative complement (cf. Lin 2015: 253-258).

Here, *bah*^{4} ‘meat’ is presented as the thing that is being eaten by the speaker. *Bah*^{4} is constructed as the direct object of the transitive verb *tsiah*^{8} ‘eat’. In (8), *tsiah*^{8} is constructed with the object *png*^{7}:

(8) Gua$^{2\text{-}1}$ tsiah$^{8\text{-}3}$ png^{7}.
1SG eat cooked.rice

‘I am eating.’

In this case, we don’t find “cooked rice” in the translation, even though *png*^{7} literally means ‘cooked rice’, and by extension ‘meal’. In TSM, *tsiah*^{8} ‘eat’ is a transitive verb and as a rule, in this language, transitive verbs must be constructed with a direct object. In a language such as English, this is not necessarily the case. One can both use the sentence *I am eating a meal*. And the sentence *I am eating.* to describe a situation in which one is having lunch, for instance. In the former, the object *a meal*, referring to the patient or thing that is being eaten, is explicitly mentioned. In the latter, this patient is simply not mentioned and even though common sense tells us that whenever the action of eating takes place, there inevitably is something that is being eaten, in the sentence *I am eating.* the thing that is eaten is simply not denoted. In TSM, however, it is impossible not to denote the presence of a patient and both English *I am eating a meal.* and *I am eating.* could be translated as *Gua*$^{1\text{-}7}$ *tsiah*$^{8\text{-}3}$ *png*^{7}. seen in (8). If the patient is known from either the linguistic or extralinguistic context, however, this direct object does not have to be stated *per se*, as illustrated in sentence (9), which illustrates a use of the transitive verb *kong*^{2} similar to that in (1):

(9) A^{0} tshenn$^{1\text{-}7}$me$^{5\text{-}7}$ gu^{5} m$^{7\text{-}3}$ bat^{4}, e$^{7\text{-}3}$hiau$^{2\text{-}1}$ **kong^{2}**, be$^{7\text{-}3}$hiau$^{2\text{-}1}$ sia^{2}.
RLT blind buffalo NEG familiar.with know.how.to **say** NEG:know.how.to write

‘But an illiterate person doesn't understand; he can speak it, but he can't write it.’ (9A; 07:13)

In the verb phrase *e*$^{7\text{-}3}$*hiau*$^{2\text{-}1}$ *kong*^{2} ‘he can speak it’, the verb *kong*^{2} ‘say’ is modified by the auxiliary verb *e*$^{7\text{-}3}$*hiau*^{2} ‘know how to’. This auxiliary verb can usually be translated as ‘can’ in English, but is used in a smaller variety of contexts than the English verb *can*. *E*$^{7\text{-}3}$*hiau*^{2} is used to assert the ability of a certain skill, i.e. one can do something because one has learned it, for example, to ride a bike or to speak a language.

Kong^{2} ‘say’, in example (9), is not followed by a direct object. If we look at the English translation, however, we do find an object: *it*. There seems to be a discrepancy between the TSM clause and its translation. As I have mentioned, *kong*^{2} ‘say’ is a transitive verb and transitive verbs must be constructed with a direct object. In the preceding text, I also briefly mentioned that,

paradoxically, if the patient of an action is known from the linguistic or extralinguistic context, the direct object does not have to be present overtly. In TSM, just as in the other Sinitic languages (see for example Wiedenhof 2015: 123-124 for Mandarin), the absence of a direct object constructed with a transitive verb implies that something known from the context must be understood as the object of the transitive verb. If one is asked which languages he speaks, he could answer with (10):

(10) Gua$^{2\text{-}1}$ kan$^{1\text{-}7}$na$^{1\text{-}7}$ e$^{7\text{-}3}$hiau$^{2\text{-}1}$ **kong$^{2\text{-}1}$** Tai$^{5\text{-}7}$gi^{2} nia^{5}.
1SG only know.how.to **say** Taiwanese merely

'I can only speak Taiwanese.'

If, on the other hand, two people were talking about Taiwanese and one of them realized he does not know whether the other speaks it, he could say the following sentence:

(11) A^{0} li$^{2\text{-}1}$ e$^{7\text{-}3}$hiau$^{2\text{-}1}$ **kong^{2}** =le^{0}?
RLT 2SG know.how.to **say** =RLV

'So do you speak it?'

Here, because *kong^{2}* is a transitive verb, the listener has to look for a possible direct object. In this case, he will not find one. The absence of a direct object in the sentence enforces a definite direct object in terms of semantics, in the case of example (11) referring to 'Taiwanese'. For most English verbs, such a definite object would have to be present overtly, but this is not the case in TSM. I will call such a definite object a *covert object*. I use this term to refer to the absence of an overtly present definite direct object following a transitive verb. The absence of such an overt element bears meaning in its own right. In English too, there are cases of ommitted objects. These cases usually involve verbs that refer to cognitive states such as *know* or *understand*. An example is the sentence *I know.*, in which a direct object, i.e. 'it', is present semantically but not stated overtly.

Looking back at (9), we see that there is no direct object mentioned in the sentence. This means a definite object is implied, which is rendered in the English translation as "it". This sentence is part of a conversation about being literate in Taiwanese. The speaker of the sentence in (9) says that a *tsenn$^{1\text{-}7}$me$^{5\text{-}7}$ gu^{5}* 'illiterate person' (literally a 'blind buffalo') does not understand Taiwanese writing; he can speak it (i.e. Taiwanese), but he cannot write it.

In example (12), there are other instances of *kong^{2}* as a transitive verb.

(12) Lan$^{2\text{-}1}$ sing$^{1\text{-}7}$-- lan$^{2\text{-}1}$ sing$^{1\text{-}7}$ lai$^{5\text{-}7}$ **kong$^{2\text{-}1}$** hit$^{4\text{-}8}$ e$^{5\text{-}7}$-- <si^{7}> **kong$^{2\text{-}1}$** thau$^{5\text{-}7}$a^{2} **kong^{2}** e$^{5\text{-}7}$
1INCL first 1INCL first come **say** that entity be **say** at.the.start **say** SUB

tshenn$^{1-7}$me$^{5-7}$ gu^{5} la^{0} <hann^{0}> honn^{0} <si^{7}>.
blind buffalo WARN CFM XPC be

'Let's first-- let's first talk about that-- <yes> about *tshenn-mê gû* 'blind buffalo' we talked about at the beginning, <hm> right <yes>.' (19C; 10:33)

Here, we see three instances of *kong*$^{2-1}$ 'say'. The first *kong*$^{2-1}$ is followed by its direct object *hit*$^{4-8}$ *e*$^{5-7}$ 'that one'. After saying this, the speaker hesitates and rephrases *kong*$^{2-1}$ *hit*$^{4-8}$ *e*$^{5-7}$ 'talk about that' with the words *kong*$^{2-1}$ *thau*$^{5-7}$*a*$^{2-1}$ *kong*^{2} *e*$^{5-7}$ *tshenn*$^{1-7}$*me*$^{5-7}$ *gu*^{5} 'talk about 'blind buffalo'[9] we talked about at the beginning'. This includes both the second and the third instance of *kong*$^{2-1}$. Looking at its function within the sentence we find that the second instance of *kong*$^{2-1}$ is just a repetition of the first one: the speaker hesitates and rephrases. This second time, the speaker states more clearly what he meant with *hit*$^{4-8}$ *e*$^{5-7}$ 'that' the first time. This specification is *thau*$^{5-7}$*a*$^{2-1}$ *kong*^{2} *e*$^{5-7}$ *tshenn*$^{1-7}$*me*$^{5-7}$ *gu*^{5}, which means ''blind buffalo' we talked about at the beginning'. This entire specification is one noun phrase which consists of two parts: a head *tshenn*$^{1-7}$*me*$^{5-7}$ *gu*^{5} 'blind buffalo' and its modifier *thau*$^{1-7}$*a*$^{2-1}$ *kong*^{2} *e*$^{5-7}$ 'that which we talked about at the beginning'. We can represent this syntactic structure using square brackets: *kong*$^{2-1}$ [[*thau*$^{5-7}$*a*$^{2-1}$ *kong*$^{2-1}$ *e*$^{5-7}$] *tshenn*$^{1-7}$*me*$^{5-7}$ *gu*^{5}].

Following Wiedenhof (1995: 15-17), I take *head* to mean that part of a noun phrase which characterizes directly the entity to which the phrase as a whole refers. The part of a phrase which characterizes the entity to which the entire phrase refers indirectly, I call *modifier*. For example, in the compound *school bus*, the meaning 'school' corresponding with *school* and the meaning 'bus' corresponding with *bus* are used to refer to the same entity in the world (be it the real world or an imaginary one). This entity is called a *referent*. In this compound, however, *bus* characterizes the referent of the meaning 'school book' directly, whereas *school* does so only indirectly. The result is that the referent of the meaning 'school book' is construed as a type of book and not as a type of school.

In *thau*$^{5-7}$*a*$^{2-1}$ *kong*^{2} *e*$^{5-7}$ *tshenn*$^{1-7}$*me*$^{5-7}$ *gu*^{5}, the modifier is a phrase in which *kong*^{2} is a verb. This phrase is subordinated to *tshenn*$^{1-7}$*me*$^{5-7}$ *gu*^{5} 'blind buffalo' by means of the subordinating particle *e*$^{5-7}$ 'SUB'. If this particle had not been there, we would have had the sentence *Thau*$^{5-7}$*a*$^{2-1}$ *kong*$^{2-1}$ *tshenn*$^{1-7}$*me*$^{5-7}$ *gu*^{5}. 'We talked about 'blind buffalo' at the beginning.' instead of the noun phrase *thau*$^{5-7}$*a*$^{2-1}$ *kong*^{2} *e*$^{5-7}$ *tshenn*$^{1-7}$*me*$^{5-7}$ *gu*^{5}. Due to the presence of *e*$^{5-7}$, this sentence is a noun phrase. Just as *hit*$^{4-8}$ *e*$^{5-7}$ 'that' was the direct object of the first instance of *kong*$^{2-1}$, so the noun phrase *thau*$^{5-7}$*a*$^{2-1}$ *kong*^{2} *e*$^{5-7}$ *tshenn*$^{1-7}$*me*$^{5-7}$ *gu*^{5} ''blind buffalo' we talked about at the beginning' is

9 In this part, I have left the phrase *tshenn-mê gú*, which is cited as a term here, out of the translations for ease of reading.

the direct object of the second instance of *kong*$^{2\text{-}1}$. Summing up, we have *kong*^{2} as a transitive verb three times: once with the object *hit*$^{4\text{-}8}$ *e*$^{5\text{-}7}$ 'that', once with the object *thau*$^{5\text{-}7}$*a*$^{2\text{-}1}$ *kong*^{2} *e*$^{5\text{-}7}$ *tshenn*$^{1\text{-}7}$*me*$^{5\text{-}7}$*gu*^{5}. The third time, *kong*^{2} is subordinated by means of the particle *e*$^{5\text{-}7}$ 'SUB'. In this case there is no object mentioned, which means we are dealing with a covert object. This covert object is coreferential with the head of the noun phrase *tshenn*$^{1\text{-}7}$*me*$^{5\text{-}7}$*gu*^{5} 'blind buffalo', i.e. the ommitted object and *tshenn*$^{1\text{-}7}$*me*$^{5\text{-}7}$*gu*^{5} have the same referent: the TSM word *tshenn*$^{1\text{-}7}$*me*$^{5\text{-}7}$*gu*^{5}.

Kong^{2} 'say' can also introduce direct speech, as in (13):

(13) A^{0} i$^{1\text{-}7}$ loh$^{8\text{-}3}$ long$^{2\text{-}1}$ ka$^{7\text{-}3}$ gua$^{2\text{-}1}$ **kong$^{2\text{-}1}$**: "A^{0} gua$^{2\text{-}1}$ na$^{2\text{-}1}$ tshenn$^{1\text{-}7}$me$^{5\text{-}7}$
RLT 3 just in.every.case PRTC 1SG **say** RLT 1SG be.like blind

gu^{5} =*a* beh$^{4\text{-}1}$ an$^{2\text{-}1}$nua^{2} <hann^{0}>?"
buffalo =EC want.to how CFM

'And they always said to me: "So what if I'm like a *tshenn-mê gú* 'blind buffalo' <hm>?"' (22C; 10:59)

The speaker uses (13) to relate to others what his parents used to say to him in the past. His parents both worked in the field and were both illiterate, as is the case with most older peasants in Taiwan. They can therefore be referred to as *tshenn*$^{1\text{-}7}$*me*$^{5\text{-}7}$ *gu*^{5} 'blind buffalo' in TSM. There are two instances of *gua*$^{2\text{-}1}$ '1SG' in (13), but they have different referents. This is because one *gua*$^{2\text{-}1}$ falls within the scope of the direct speech whereas the other does not (I will say more on the term direct speech below).

The first *gua*$^{2\text{-}1}$ refers to the speaker of sentence (13) – i.e. speaker C in my corpus. In the sentence, this *gua*$^{2\text{-}1}$ acts as the goal – that to which the action is directed – introduced by the particle *ka*$^{7\text{-}3}$.[10] That is to say, the first *gua*$^{2\text{-}1}$ presents the speaker of sentence (13) as the hearer in the speech act he is reporting here: something has been said to him by his parents. What his parents said to him follows *kong*$^{2\text{-}1}$ 'say'. This is where the speaker of sentence (13) starts quoting his parents, who told him: "*A*^{0} *gua*$^{2\text{-}1}$ *na*$^{7\text{-}3}$ *tshenn*$^{1\text{-}7}$*me*$^{5\text{-}7}$ *gu*^{5} a *beh*$^{4\text{-}1}$ *an*$^{2\text{-}1}$*nua*^{2}?", which translates to '"So what if I'm like a *tshenn-mê gû* 'blind buffalo'?"'. Within this quotation, the *gua*$^{2\text{-}1}$, therefore, refers to the speaker of the quotation which is one or both of the parents of the speaker of sentence (13).

Apart from the context, a linguistic clue as to why this is a case of direct speech as opposed to indirect speech is that the first word of the quotation is the conjunction *a*^{0} which is used at the start of clauses and sentences to express both adversative ('but') and cumulative ('and')

10 For more information on *ka*$^{7\text{-}3}$, the reader is referred to Yang (2006) and Yáng (1991: 223-225).

coordinativity and is comparable to 'well' in English or *a* in Russian. A^0 can, however, not be used to introduce clauses that are the object of a verb, which is the case with indirect speech.

Indirect speech is a paraphrase of something someone has said. If person A tells person B what person C said to A, person A has two options. The first is to use the exact wording of person C and thus to quote him exactly; this is direct speech. The second option is to paraphrase person C; as opposed to a quotation, person A would not use the exact wording of person C, but rather A would convey in his own words to person B what he perceived to be the intended meaning of person C. In (14a), an example of direct speech in English is presented, and in (14b), an example of indirect speech in English is presented:

(14) a. Obama always says: "Yes, we can!".
b. Obama always says that we can.

Obama was the surname of the first black president of the United States of America, who was in office from January 20, 2009 to January 20, 2017. Main themes in his campaigns for presidency were hope and change. One of his most famous slogans was "Yes, we can!".

In (14a) Obama is quoted directly. In (14b), the speaker does not use the exact words of Obama in relating what he says, i.e. he does not quote him. This can be seen from the omission of the interjection *yes*, which is unusual in indirect speech – cf. a^0 in (13). Rather, he paraphrases Obama introducing the clause *we can* with the complementizer *that*. A complementizer is a morpheme that can be seen as the head of a dependent clause, but can also be thought of as introducing the dependent clause to the preceding verb (Matthews 2007: 68). Here, I will take the broader definition of complementizer as a morpheme introducing the dependent clause to the preceding verb. In this case, the clause *that we can* functions as the object of the verb *says*. Such verbs that take a clause as their object are generally termed matrix verbs (see a.o. Foolen et al. 2006; Lau 2013 and Diercks 2013). I use the term *dependent clause* to refer to the clause that functions as the object of such a matrix verb.

In English, the complementizer *that* is not obligatory, as (14c) shows:

(14) c. Obama always says we can.

Grammatically, the clause *we can* functions in this case directly as the object of the matrix verb *says*, and is not introduced by the complementizer *that*. Prosody does make it possible to distinguish the indirect speech of (14c) from the direct speech of (14d) below, but this distinction is in fact

optional. In written English, the intonation of a direct quotation is usually marked with a colon preceding the quotation and quotation marks <“ ”> surrounding the quotation:

(14) d. Obama always says: “We can.”.

According to Li (1986), the functional difference between direct and indirect speech in a language such as English is clear. In direct speech the speaker reporting the utterance acts as if he were the original speaker of the utterance. This means that both the form, the content and the non-verbal messages of the utterance are presented as being the same as in the original utterance. In indirect speech only the content is presented as being the same, whereas the form of the utterance as well as any non-verbal messages are presented as not being part of the original utterance (Li 1986: 38-39; cf. especially the schemes in (36) and (37) there). Following Li’s line of reasoning, if the utterance of (14c) is pronounced with contempt, the hearer knows this contempt was not part of the original utterance as made by Obama, but rather is a judgement of the reporting speaker on what Obama always says. If, on the contrary, quotation *“We can.”* in (14d) is made with contempt, the listener not familiar with Obama’s slogan is likely to assume that Obama himself always makes this utterance with contempt.

I find Li’s statement hard to assess. It seems reasonable to assume that if a speaker employs indirect speech to relate something someone else has said and in doing so makes the utterance with sarcasm or contempt, a hearer will interpret this sarcasm or contempt as a way for the speaker to express his own attitude towards the content of what he is relating, cf. (14c). The question is whether the opposite is also the case. Is it safe to say that if a speaker pronounces a direct quotation with sarcasm, the hearer will interpret this sarcasm as being part of the original message being related? If one looks at video footage of Donald Trump, now president of the United States, impersonating the disabled reporter Serge Kovaleski[11], one sees Trump at a rally in South Carolina November 24, 215, at that time still running for president, talking about the physically disabled reporter Serge Kovaleski using the third person singular pronoun *he* to refer to Kovaleski. In the middle of his report, Trump switches over to direct speech which can be deduced from the fact that now, Trump does not use the third person pronoun *he* to refer to the reporter, but the first person pronoun *I*. During this quotation, Trump is shaking his arms in front of his body and stuttering, speaking in a flustered way. This flusteredness indeed is presented as being part of the original

11 Clips of this can be found on YouTube.com, see for instance the following link: https://www.youtube.com/watch?v=PX9reO3QnUA. (downloaded, 16 May 2017)

utterance as made by Kovaleski, and not as an addition of Trump. After the quotation Trump switches his tone of voice and continues in an agitated, offended way. The fact that he uses a different tones of voice for the quotation and for his own words strongly steers the hearer to interpreting the flusteredness as being part of the original quotation – as indeed it has been presented. But suppose that Trump would have spoken in an agitated voice throughout both the quotation and his own words. Judging from Li (1986: 38-39), the hearer can only interpret this as 1) Trump is agitated, for he uses an agitated voice when telling this story, and 2) Kovaleski is also agitated, for the quotation is presented in an agitated voice as well. I am not sure whether this is true, but I also do not know for certain whether it is common for one to use one's own tone of voice throughout both one's own words and a quotation of someone else.

I will not attempt to reach a conclusion about this relationship between non-verbal messages and quotation here. I will, however, follow Li in saying that when a speaker uses indirect speech, and thus paraphrases someone else's words, any non-verbal messages accompanying the utterance will be interpreted by the hearer as reflecting the attitude of the speaker who is paraphrasing, not the attitude of the original speaker whose words are being paraphrased. I also follow Li in saying that in direct speech, the reporting speaker (i.e. the speaker relating the original utterance to a third party) acts as if he were the reported speaker (i.e. the one who made the original utterance). I will leave this issue as to whom any non-verbal messages are attributed asside.

On a structural level, the difference between direct and indirect speech seems to be best described in terms of grammatical integration; i.e. a message conveyed by means of indirect speech is grammatically speaking more integrated within the rest of the sentence than a quote using direct speech is (Coulmas 1986: 19-23).

(15) a. I see that he is eating olives.
b. I see him eating olives.
c. I see his eating olives.

In a sentence such as (15a), the dependent clause presenting what was seen is structurally less integrated into the clause of the verb *see* than in (15b). In the dependent clause *that he is eating olives* of (15a) reference is made to an agent, the one who is eating, which is denoted by means of the grammatical subject *he*. The action itself is denoted by the predicate *is eating olives*. In (15b), the part denoting the one eating olives, i.e. *him eating olives*, is the grammatical subject of a finite verb, but is in this case the object pronoun *him* which functions as the direct object of the verb *see*. Compared to (15a) and (15b), (15c) shows even more grammatical integration of the action 'seeing'

and the action ‘olive eating’. Here, the agent of the action is not expressed as either a free-standing subject or object pronoun, but as a possessive pronoun which functions as the modifier of the (verbal) noun *eating*. In a similar fashion, the following sentences show a varying degree of grammatical integration.

(16) a. Mary says: “I’m not at home.”
b. Mary says that she isn’t at home.
c. Mary says she isn’t at home.
d. That she isn’t at home is regrettable.
e. *She isn’t at home is regrettable

In sentence (16a), Mary is reported to have said the following words “I’m not at home.”. Structurally, the reported speech is less integrated into the clause *Mary says* than it is in (16b). We can see this especially clearly from the different pronouns being used: in (16a) the first person singular pronoun *I* is used which is coreferential with *Mary*. In (16b), the third person pronoun *she* is used which can be coreferential with *Mary*, but need not be. It would, however, not be possible to use the pronoun *I* to refer to the person Mary and be coreferential with the word *Mary*. If *I* had been used here, it would refer to the speaker of the utterance in (16b) and not to Mary. (16c) goes one step further than (16b), for the phrase *that she isn’t at home* can be the subject of a sentence as shown in (16d), whereas (16e) shows that the clause *she isn’t at home* cannot be the independent subject of a sentence. The clause *she isn’t at home* is thus grammatically more strongly integrated into the clause *Mary says* than its counterpart *that she isn’t at home* is in (16b).

In my opinion, the fact that direct speech is grammatically less integrated into the sentence of the matrix verb than indirect speech, is linked to the functional properties as described by Li (1986: 38-39) of direct speech which he argues conveys both content and form (and possibly non-verbal messages) as being part of the original utterance whereas he argues indirect speech to convey only the content as being part of the original utterance. This difference of grammatical integration can be seen from the difference between (14a) and (14b), repeated here.

(14) a. Obama always says: “Yes, we can!”
b. Obama always says that we can.
c. Obama always says we can.
d. Obama always says: “We can.”

In the quotation in (14a), we have the interjection *yes* as part of the reported speech. In the case of indirect speech in (14b), we do not have the interjection *yes*, for interjections are notoriously hard to integrate within the grammatical structure of a sentence. Here, we find the complementizer *that.* In English, the complementizer *that* is not, however, obligatorily used in a dependent clause, as was shown in (14c). Structurally, the difference is that in (14b) the clause *we can* is the object of the complementizer *that*, thus forming the dependent clause *that we can* which functions as the direct object of the matrix verb *says*. In (14c), however, the clause *we can* directly functions as the direct object of the matrix verb.

Whether or not we should consider the direct speech *Yes, we can.* of (14a) or *We can.* of (14d) to be the object of the verb *says* is a difficult problem. For sure, semantically *Yes, we can.* and *We can.* both fulfill the role of the patient, i.c. the thing that is said. However, as was argued in this paragraph, the level of grammatical integration of direct speech within the sentence is lower than that of indirect speech. I do not see any solution to this problem and for reasons of convenience will, in the remainder of this article, choose to view both direct and indirect speech as the object of the verb of speech it is constructed with. I take the main difference between the two to be the above described level of grammatical integration.

The following example shows *kong*$^{2\text{-}1}$ followed by indirect speech:

(17)

Ia^{7}--	ia^{7}--	ia$^{7\text{-}3}$	e$^{7\text{-}3}$sai$^{2\text{-}1}$	**kong$^{2\text{-}1}$**	han$^{3\text{-}2}$ji^{7}	e$^{7\text{-}3}$sai$^{2\text{-}1}$	theh$^{8\text{-}3}$	lai$^{5\text{-}7}$
as.well	as.well	as.well	be.allowed.to	**say**	Chinese.character	be.allowed.to	take	come

tso$^{3\text{-}2}$	tsok$^{4\text{-}8}$	tse$^{7\text{-}3}$	tsiong$^{2\text{-}1}$	gi$^{2\text{-}1}$gian^{5}	e$^{5\text{-}7}$	su$^{1\text{-}7}$sia$^{2\text{-}1}$	bun$^{5\text{-}7}$hian^{3}	<oo>	<tioh^{8}>.
do	very	many	kind.of	language	SUB	write	document	SRPS	right

‘You can also-- also-- also say that Chinese characters can be used to make written documents for a lot of different kinds of languages <ow> <right>.’ (55B; 17:16)

Here, just as in (14b & c) we see that the long clause *han*$^{3\text{-}2}$*ji*^{7} *e*$^{7\text{-}3}$*sai*$^{2\text{-}1}$ *theh*$^{8\text{-}3}$ *lai*$^{5\text{-}7}$ *tso*$^{3\text{-}2}$ *tsok*$^{4\text{-}8}$ *tse*$^{7\text{-}3}$ *tsiong*$^{2\text{-}1}$ *gi*$^{2\text{-}1}$*gian*^{5} *e*$^{5\text{-}7}$ *su*$^{1\text{-}7}$*sia*$^{2\text{-}1}$ *bun*$^{5\text{-}7}$*hian*^{3} ‘Chinese characters can be used to make written documents for a lot of different kinds of languages’ is constructed as the direct object of the verb phrase *ia*$^{7\text{-}3}$-- *ia*$^{7\text{-}3}$-- *ia*$^{7\text{-}3}$ *e*$^{7\text{-}3}$*sai*$^{2\text{-}1}$ *kong*$^{2\text{-}1}$ ‘you can also-- also-- also say’. There is no quotation in this sentence and the speaker does not make use of direct speech, which can be inferred from the context and can be heard from the prosody. Syntactically, however, the TSM sentence in (17) is more similar to the English sentence in (14c) than it is to (14b). This is because in (14b) there is the complementizer *that* introducing the clause *we can* as the direct object to the verb *says*. In (17), just as in (14c), there is no such complementizer between the verb *kong*$^{2\text{-}1}$ and the complement clause

han$^{3-2}$*ji*^{7} *e*$^{7-3}$*sai*$^{2-1}$ *theh*$^{8-3}$ *lai*$^{5-7}$ *tso*$^{3-2}$ *tsok*$^{4-8}$ *tse*$^{7-3}$ *tsiong*$^{2-1}$ *gi*$^{2-1}$*gian*^{5} *e*$^{5-7}$ *su*$^{1-7}$*sia*$^{2-1}$ *bun*$^{5-7}$*hian*^{3} 'Chinese characters can be used to make written documents for a lot of different kinds of languages'. Semantically, the sentence in (17) is also similar to (14c), for in the absence of context, both can be construed as direct or indirect speech.

In all the TSM examples we have seen so far, *kong*$^{2-1}$ 'say' is used as the only transitive verb in the sentence. Any other verbs found along with *kong*$^{2-1}$ turned out to be auxiliary verbs such as *u*$^{7-3}$ 'EXST' in (1), *e*$^{7-3}$*hiau*$^{2-1}$ 'know how to' in (9-11), *lai*$^{5-7}$ 'come, get to' in (12), and *e*$^{7-3}$*sai*$^{2-1}$ 'be allowed to' in (17).

In the Appendix on pp. 82-95, more examples of *kong*^{2} as a transitive verb can be found. In my corpus, out of a total of 83 instances of *kong*^{2} 52 are clear examples of *kong*^{2} 'say' as a transitive verb. Rounded off, this is 63%. The example sentences in Appendix I are arranged chronologically. In the main text, however, I have not presented them chronologically, cf. § 1.3. Table 5 lists all the examples in which *kong*^{2} is a transitive verb in the numerical order of the Appendix.

Table 5: Examples of *kong*^{2} as transitive verb

Sentence #	Instance #	Sentence #	Instance #	Sentence #	Instance #	Sentence #	Instance #
2A		16A	2	23C		45D	3
5A		16A	3	25C		45D	4
6B	1	17A		26C		46D	
6B	2	18A	1	28C		47D	2
7A	2	18A	2	29C	1	49D	2
8A	1	19C	1	29C	2	50A	
8A	2	19C	2	30C		51A	
9A		19C	3	31A		52A	1
11B		20C	1	32C	2	52A	2
13A		20C	3	36C		52A	3
14A	2	21C	1	38C		55B	
15A		21C	2	41D	1	56D	1
16A	1	22C		42D	1	56D	2

Table 5 has columns labeled "Sentence #" and "Instance #". In the columns labeled "Sentence #", the numbers of all sentences in which *kong*^{2} functions as a full verb are listed numerically. If one example sentence has more than one instance of *kong*^{2}, these instances are numbered chronologically. The first instance of *kong*^{2} has been assigned the number 1, the second number 2, the third number 3 and the fourth number 4. These numbers, if present, are listed in the columns labeled "Instance #" next to the sentence number they belong to.

To illustrate example (12) in the text above corresponds to example 19C in the Appendix so it is listed as 19C, which can be found in the third column of Table 5. There are three instances of *kong*^{2} within this example. These instances are all instances of *kong*^{2} as a full verb. 19C is therefore listed in Table 5 thrice with the instance numbers 1, 2 and 3 following in the cells directly to the right of those with 19C in them.

2.1.2 Summary

The prototypical function of *kong*^{2} is that of a transitive verb meaning 'say'. This is apparent from dictionaries (Lôo 2011: 191; Douglas 1899: 246) and can be corroborated by the example sentences in my corpus. *Kong*^{2} is a transitive verb that can be constructed with an explicit object such as *siann*$^{2\text{-}1}$*mih*$^{8\text{-}3}$ *hue*^{7} 'what thing' as in (1). If there is no overt object in the sentence, a definite object is covertly present, as in (12). The listener has to rely on context in order to know what known entity the speaker refers to when he does not explicitly mention the direct object of *kong*^{2}. *Kong*^{2} can also be constructed with direct speech in order to quote someone, as in (13); or it can be constructed as indirect speech, in which a so-called complement clause functions as the direct object of *kong*^{2}. An example of this has been presented in (17). Of the total of 83 instances of *kong*^{2} in my corpus, there were 52 clear-cut cases of *kong*^{2} being used as a transitive verb. Rounded off, this is 63 %.

2.2 *Kong*^{2} following transitive & intransitive verbs

2.2.1 Data & analysis

So far, I have discussed the function of *kong*^{2} ~ *kong*$^{2\text{-}1}$ as a transitive verb meaning 'say'. In this section, I will discuss examples such as (2) – repeated below – which are different from the examples discussed in § 2.1 in that it is not clear from the English translations if *kong*^{2} functions as a verb meaning 'say'.

(2) A^{0} soo$^{2\text{-}1}$i$^{2\text{-}1}$ kho$^{2\text{-}1}$ling^{5} =*ne*, eh-- koh$^{4\text{-}1}$ ui$^{7\text{-}3}$ thiann$^{1\text{-}7}$tsiong$^{3\text{-}2}$ ping$^{5\text{-}7}$iu^{2} e$^{7\text{-}3}$ **kam$^{2\text{-}1}$kak$^{4\text{-}8}$**
RLT therefore maybe =RLV HST every place listening.audience friend will **feel**

kong$^{2\text{-}1}$, m^{0} khak$^{4\text{-}8}$sit^{8} =*ne*, tsin$^{1\text{-}7}$tsiann^{2} tsainn$^{1\text{-}7}$iann$^{2\text{-}1}$-- beh$^{4\text{-}1}$ tsainn$^{1\text{-}7}$iann$^{2\text{-}1}$
say AGR indeed =RLV really know want.to know

i$^{1\text{-}7}$ e$^{5\text{-}7}$ i$^{3\text{-}2}$su^{3}.
3 SUB meaning

'So it may be that uhm our listeners will feel that, hm yes, they really know-- want to know its meaning.' (3A; 05:40)

In (2), we do indeed find *kong*$^{2\text{-}1}$, but we do not see a verb of speech mentioned in the translation. *Kong*$^{2\text{-}1}$ follows the verb phrase *e*$^{7\text{-}3}$ *kam*$^{2\text{-}1}$*kak*$^{4\text{-}8}$ 'will feel', which we do find in the translation. Let us take a look at a couple of similar examples in which *kong*$^{2\text{-}1}$ follows a transitive verb.

(18) A^{0} soo$^{2\text{-}1}$i$^{2\text{-}1}$ ji^{7} lan$^{2\text{-}1}$ ti$^{7\text{-}3}$ Tai$^{5\text{-}7}$uan^{5} n-- thong$^{1\text{-}7}$siong^{5} long$^{2\text{-}1}$ e$^{7\text{-}3}$ hong^{9}
RLT therefore character 1INCL be.at Taiwan HST usually in.every.case will give:person

goo$^{7\text{-}3}$kai$^{2\text{-}1}$ kong$^{2\text{-}1}$, a^{0} ji^{7}, toh$^{8\text{-}3}$ si$^{7\text{-}3}$ Khong$^{2\text{-}1}$tsu$^{2\text{-}1}$ ji^{7}, toh$^{8\text{-}3}$ si$^{7\text{-}3}$
misunderstand say RLT character just be Confucius character just be

han$^{3\text{-}2}$ji^{7}, kio$^{3\text{-}2}$ tso$^{3\text{-}2}$ ji^{7}, <hmm> honn^{0}.
Chinese.character call do character hmm XPC

'So *jī* 'characters', we in Taiwan n-- usually always misunderstand them like, well *jī* are *Khóng-tsú jī* 'Confucius' characters', *hàn-jī* 'Chinese characters' in other words, that is what are called *jī* <hmm>, right?' (34C; 12:45)

(19) A^{0} khi$^{5\text{-}7}$sit^{8} li$^{2\text{-}1}$ na$^{7\text{-}3}$ thak$^{8\text{-}4}$ hit$^{4\text{-}8}$ lo^{7}-- lan$^{2\text{-}1}$ jin$^{5\text{-}7}$lui^{7} hit$^{4\text{-}8}$ e$^{5\text{-}7}$ bun$^{5\text{-}7}$ji^{7} e$^{5\text{-}7}$ ian$^{2\text{-}1}$piat^{4}
RLT actually 2SG if study that sort 1INCL humanity that entity writing.system SUB evolve

li$^{2\text{-}1}$ toh$^{8\text{-}3}$ **tsai$^{1\text{-}7}$iann$^{2\text{-}1}$ kong$^{2\text{-}1}$** khi$^{5\text{-}7}$sit^{8} tsit$^{8\text{-}4}$ tsiong$^{2\text{-}1}$ ji^{7} si$^{7\text{-}3}$ khah$^{4\text{-}1}$-- khah$^{4\text{-}1}$
2SG just **know** **say** actually one kind.of character be relatively relatively

tsa^{2} tsin$^{3-2}$tsing^{5} u^{7} e$^{5-0}$.
early before EXST SUB

'And actually if you study uhm-- the evolution of mankind's writing systems, you know that actually one kind of *jī* 'writing' existed relatively-- relatively early on.' (35C; 12:55)

In both (18) and (19) we see *kong*$^{2-1}$ following a verb: *goo*$^{7-3}$*kai*$^{3-2}$ 'misunderstand' and *tsai*$^{1-7}$*iann*$^{2-1}$ 'know' respectively. In (18), speaker C is explaining that most Taiwanese do not understand the concept *ji*^{7} 'character' properly. He does this by first stating that the Taiwanese people *goo*$^{7-3}$*kai*$^{3-2}$ 'misunderstand' characters and he then immediately proceeds to quote a probably fictitious person. This means that speaker C presents the utterance by means of direct speech, which will be understood by the listeners as being a quotation. Rember from the discussion of (13) that the fact that the conjunction *a*^{0} is present also indicates we are dealing with a quotation. The utterance in this quotation, however, has probably never been made before. That is to say, speaker C is not quoting a real person he has met, rather, he presents the utterance as though it has been made by someone in particular; a fictitious person, in other words. This fictitious person explains what he thinks *ji*^{7} 'characters' are, viz. *toh*$^{8-3}$ *si*$^{7-3}$ *Khong*$^{2-1}$*tsu*$^{2-1}$ *ji*^{7}, *toh*$^{8-3}$ *si*$^{7-3}$ *han*$^{3-2}$*ji*^{7}, *kio*$^{3-2}$ *tso*$^{3-2}$ *ji*^{7} '[those] are Confucius' characters, Chinese characters in other words, that is what are called *jī*'. Note that by using direct speech, speaker C gives more credibility to his statement that some Taiwanese misunderstand characters in the way speaker C describes it. After all, he is quoting someone saying this. Speaker C's point is that this notion of what *ji*^{7} are is wrong, whereas his own notion is the right one, as is clear from his entire monologue. According to speaker C, *ji*^{7} are the things you use to represent spoken language: i.e. writing ("*A*^{0} *ji*^{7} *toh*$^{8-3}$ *si*$^{7-3}$ *kong*$^{2-1}$, *lan*$^{2-1}$ *ka*$^{7-3}$ *i*$^{1-7}$ *sia*^{2} *tshut*$^{4-0}$*lai*$^{5-0}$ *a*^{0} *lan*$^{2-1}$ *khi*$^{3-2}$ *ka*$^{7-3}$ *i*$^{1-7}$ *thak*^{8}, *an*$^{2-1}$*ne*$^{1-7}$ *khuann*$^{3-2}$*u*$^{7-3}$ *i*$^{3-2}$*su*^{3} *e*$^{5-0}$ *kio*$^{3-2}$*tso*$^{3-2}$ *ji*^{7} *ma*^{0}" 'And characters are, you know, we write them down and we read them, things in which we can see meaning like this are called *ji*^{7} right' TML 12:39). I want to stress that it is irrelevant for the current purpose what the supposedly correct interpretation of the concept *jī* 'character' is. What interests us here is the way speaker C conveys what he wants to say – the language of the most-likely-fictitious person being of course also speakers C's own words. In this case, speaker C uses *kong*$^{2-1}$ immediately preceding the quotation. We will look at the connection between *goo*$^{7-3}$*kai*$^{3-2}$ 'misunderstand', *kong*$^{2-1}$ and the quotation in more detail at a later stage.

Example (19) is the continuation of (18) and speaker C is still talking about *ji*^{7} 'characters'. Speaker C argues that if we look at writing systems from all over the world, we see that *tsit*$^{8-4}$ *tsiong*$^{2-1}$ *ji*^{7} 'one kind of *jī* 'writing'' keeps cropping up. Although in (18) and (19), it is not yet clear

what kind of writing system he is talking about, a couple of sentences later, we can deduce that he is talking about pictographic writing ("*long*$^{2\text{-}1}$ *ue*$^{7\text{-}3}$ *ang*$^{1\text{-}7}$*a*^{2} *an*$^{2\text{-}1}$*ne*^{1} *la*^{0}." 'it was all like drawing little people, you know' TML 13:10; "*long*$^{2\text{-}1}$ *khuann*$^{3\text{-}2}$*-tioh*$^{8\text{-}3}$ *sim*$^{2\text{-}1}$*mih*$^{8\text{-}3}$ *hue*^{3}*, a*^{0} *ka*$^{7\text{-}3}$ *ue*$^{7\text{-}3}$ *hit*$^{4\text{-}8}$ *hing*^{5} *tshut*$^{4\text{-}0}$*lai*$^{5\text{-}0}$." 'it was all whatever people saw and then drawing its shape' TML 13:11).

In (19), *kong*$^{2\text{-}1}$ follows the verb *tsai*$^{1\text{-}7}$*iann*$^{2\text{-}1}$ 'know'. *Tsai*$^{1\text{-}7}$*iann*^{2} is a verb denoting the presence of the knowledge of some fact or another. In English, the verb *know* can a.o. be used with a dependent clause expressing what one knows (e.g. *I know he isn't coming.*) or with a noun phrase, referring to a person or thing (e.g. *I know that girl.*). In TSM, if one explicitly states what is known, this usually is done using a clause, e.g.:

(20) Gua-n$^{2\text{-}1}$ boo^{2} tsai$^{1\text{-}7}$iann$^{2\text{-}1}$ i-n$^{1\text{-}7}$ a$^{1\text{-}7}$-hiann^{1} m$^{7\text{-}3}$ lai^{5}.
1SG-PL wife know 3-PL FAM-older.brother NEG come

'My wife knows her brother is not coming.'

In TSM, semantically speaking, the thing that is known is in TSM always an event and never an entity. Syntactically, this event can both be expressed as a dependent clause, as in (20), and as a noun phrase, e.g. *tsit*$^{8\text{-}4}$ *kiann*$^{7\text{-}3}$ *tai*$^{7\text{-}3}$*tsi*^{3} 'this matter'. In other words, unlike English *know*, the verb *tsai*$^{1\text{-}7}$*iann*^{2} 'know' cannot be used to express that one knows a person or a thing, i.e. that one is familiar with someone or something. In such cases, the verb *bat*^{4} 'familiar with' would be used in Taiwanese, as in (21) below:

(21) Tsit$^{4\text{-}8}$ e$^{5\text{-}7}$ lang^{5} li$^{2\text{-}1}$ kam$^{2\text{-}1}$ bat^{4} i$^{1\text{-}0}$.
this entity person 2SG QW familiar with 3

'This guy, do you know him?'

(19) A^{0} khi$^{5\text{-}7}$sit^{8} li$^{2\text{-}1}$ na$^{7\text{-}3}$ thak$^{8\text{-}4}$ hit$^{4\text{-}8}$ lo^{7}-- lan$^{2\text{-}1}$ jin$^{5\text{-}7}$lui^{7} hit$^{4\text{-}8}$ e$^{5\text{-}7}$ bun$^{5\text{-}7}$ji^{7} e$^{5\text{-}7}$ ian$^{2\text{-}1}$piat^{4}
RLT actually 2SG if study that sort 1INCL humanity that entity writing.system SUB evolve

li$^{2\text{-}1}$ toh$^{8\text{-}3}$ **tsai$^{1\text{-}7}$iann$^{2\text{-}1}$** **kong$^{2\text{-}1}$** khi$^{5\text{-}7}$sit^{8} tsit$^{8\text{-}4}$ tsiong$^{2\text{-}1}$ ji^{7} si$^{7\text{-}3}$ khah$^{4\text{-}1}$-- khah$^{4\text{-}1}$
2SG just **know** **say** actually one kind.of character be relatively relatively

tsa^{2} tsin$^{3\text{-}2}$tsing^{5} u^{7} e$^{5\text{-}0}$.
early before EXST SUB

'And actually if you study uhm-- the evolution of mankind's writing systems, you know that actually one kind of *jī* 'character' existed relatively-- relatively early on.' (35C; 12:55)

If we look back at (19), we see that following *kong*$^{2-1}$ there is the clause *khi*$^{5-7}$*sit*^{8} *tsit*$^{8-4}$ *tsiong*$^{2-1}$ *ji*^{7} *si*$^{7-3}$ *khah*$^{4-1}$-- *khah*$^{4-1}$ *tsa*^{2} *tsin*$^{3-2}$*tsing*^{5} *u*^{7} *e*$^{5-0}$ 'actually one kind of *jī* 'character' existed relatively--relatively early on'. Semantically, this clause is a description of the thing that is known. It could be expected then that, syntactically, this clause is dependent on the verb *tsai*$^{1-7}$*iann*$^{2-1}$, which would function as the matrix verb and with which the dependent clause is constructed as its object. However, that which is known is separated from the verb *tsai*$^{1-7}$*iann*$^{2-1}$ by *kong*$^{2-1}$. This sentence seems reminiscent of both the sentence in (20) and the case of indirect speech in (17) of § 2.1.1 – repeated below.

(17) Ia^{7}-- ia^{7}-- ia$^{7-3}$ e$^{7-3}$sai$^{2-1}$ **kong$^{2-1}$** han$^{3-2}$ji^{7} e$^{7-3}$sai$^{2-1}$ theh$^{8-3}$ lai$^{5-7}$
as.well as.well as.well be.allowed.to **say** Chinese.character be.allowed.to take come

tso$^{3-2}$ tsok$^{4-8}$ tse$^{7-3}$ tsiong$^{2-1}$ gi$^{2-1}$gian^{5} e$^{5-7}$ su$^{1-7}$sia$^{2-1}$ bun$^{5-7}$hian^{3} <oo> <tioh^{8}>.
do very many kind.of language SUB write document SRPS right

'You can also-- also-- also say that Chinese characters can be used to make written documents for a lot of different kinds of languages <ow> <right>.' (55B; 17:16)

Syntactically, the question is whether we should take the clause *khi*$^{5-7}$*sit*^{8} *tsit*$^{8-4}$ *tsiong*$^{2-1}$ *ji*^{7} *si*$^{7-3}$ *khah*$^{4-1}$-- *khah*$^{4-1}$ *tsa*^{2} *tsin*$^{3-2}$*tsing*^{5} *u*^{7} *e*$^{5-0}$ 'actually one kind of *jī* 'character' existed relatively--relatively early on' as the object of the verb *kong*$^{2-1}$ 'say', to which it is closer; or of the verb *tsai*$^{1-7}$*iann*$^{2-1}$ 'know', with which it seems to be closer aligned semantically, as can be seen from the English translation. I will present my views on the relationship between syntax and semantics in the discussion of examples (14a) and (14b).

In (22), the same structure as in (19) is shown.

(22) A^{0} soo$^{2-1}$i$^{2-1}$ kong$^{2-1}$, ti$^{7-3}$ tsia^{1} *=ne*, *tòuguò* *=ne*, Soo$^{1-7}$ Se$^{3-2}$hiong^{5} *lăosī* hoo^{0}, ua$^{2-1}$ ma$^{7-3}$
RLT therefore say be.at here =RLV via =RLV PNSoo PNSè-hiông teacher XPC 1SG also

hi$^{1-7}$bang$^{7-3}$ **kong$^{2-1}$** i$^{1-7}$ e$^{7-3}$tang$^{3-2}$ sio$^{2-1}$ kan^{9} kai$^{2-1}$sueh^{4} tsit$^{8-0}$e$^{7-0}$.
hope **say** 3 be.able.to a.little OBJ:1INCL[12] explain a.while

'So, here, through, you know, Mr. Soo Sè-hiông, I also hope that he can explain it a bit to us.' (4A; 06:07)

Here, we see *kong*$^{2-1}$ twice: once at the start following *soo*$^{2-1}$*i*$^{2-1}$ 'therefore', and once following the verb *hi*$^{1-7}$*bang*$^{7-3}$ 'hope' and preceding the clause *i*$^{1-7}$ *e*$^{7-3}$*tang*$^{3-2}$ *sio*$^{2-1}$ *kan*^{9} *kai*$^{2-1}$*sueh*^{4} *tsit*$^{8-0}$*e*$^{7-0}$ 'he can

12 The word *kan*^{9} is a fusion of *ka*^{7} 'OBJ' and *lan*^{2} '1INCL'.

explain it a bit to us'. We will get back to the first one in § 2.3; the second one is of interest here, as it displays a structure similar to (19). We have the verb *hi*$^{1-7}$*bang*$^{7-3}$ 'hope', which expresses the desire for something to be the case, as is shown in (23).

(23) Li-n$^{2-1}$ lau$^{7-3}$bo^{2} ma$^{7-3}$ hi$^{1-7}$bang$^{7-3}$ li-n$^{2-1}$ lau$^{7-3}$pe^{7} e$^{7-3}$tang$^{3-2}$ koh$^{4-1}$ tshue$^{7-3}$-tioh$^{8-3}$ thau$^{5-7}$loo^{7}.
2SG-PL mother also hope 2SG-PL father be.able.to again search-get job

'Your mother too is hoping your father will be able to find a job again.'

In (23), we see that, semantically, the clause *lin*$^{2-1}$ *lau*$^{7-3}$*pe*^{7} *e*$^{7-3}$*tang*$^{3-2}$ *koh*$^{4-1}$ *tshue*$^{7-3}$*tioh*$^{8-3}$ *thau*$^{5-7}$*loo*^{7} 'you father will be able to find a job again' expresses that which is hoped. Syntactically, it directly follows the verb *hi*$^{1-7}$*bang*$^{7-3}$ 'hope' and is constructed as its dependent clause.

As is the case with *tsai*$^{1-7}$*iann*^{2}, whenever the desire or that which is hoped is explicitly mentioned, it usually is constructed as a dependent clause to the matrix verb *hi*$^{1-7}$*bang*^{7} 'hope'. In (22), this can be found in the dependent clause *i*$^{1-7}$ *e*$^{7-3}$*tang*$^{3-2}$ *sio*$^{2-1}$ *kan*^{9} *kai*$^{2-1}$*sueh*^{4} *tsit*$^{8-0}$*e*$^{7-0}$ 'he can explain it a bit to us', which semantically fills the role of that which the *ua*$^{2-1}$ 'I' hopes to be the case. Again, in (22), we find *kong*$^{2-1}$ 'say' quite literally separating the verb *hi*$^{1-7}$*bang*$^{7-3}$ 'hope' from the thing that is hoped, viz. *i*$^{1-7}$ *e*$^{7-3}$*tang*$^{3-2}$ *sio*$^{2-1}$ *kan*^{9} *kai*$^{2-1}$*sueh*^{4} *tsit*$^{8-0}$*e*$^{7-0}$ 'he can explain it a bit to us'. Syntactically, the clause *i*$^{1-7}$ *e*$^{7-3}$*tang*$^{3-2}$ *sio*$^{2-1}$ *kan*^{9} *kai*$^{2-1}$*sueh*^{4} *tsit*$^{8-0}$*e*$^{7-0}$ seems to be constructed as the object of *kong*$^{2-1}$, which would seem to make *kong*$^{2-1}$ a transitive verb here. Semantically, however, this clause is not what is presented as being said, it is presented as what is being hoped. Note that because of the nature of the verb *kong*^{2} 'say', there might arise some confusion here. An act of speaking, i.e. 'I also hope to say that he can explain it a bit to us', is not part of the interpretation of (22). Rather, the speaker conveys the meaning as has been presented in the translation of (22), viz. 'I also hope that he can explain it a bit to us.'.

Even though in its context, the sentence in (22) is interpreted by native speakers as meaning 'I hope that he can explain it a bit to us', without context (22) could also be interpreted as 'I also hope to say that he can explain it to us.'. Native speakers are especially reluctant to interpret the collocation *hi*$^{1-7}$*bang*$^{7-3}$ *kong*$^{2-1}$ followed by a clause as 'hoping to say that [dependent clause]'. The widely preferred interpretation is 'hoping that [dependent clause]'. Still, the fact that it is in fact possible for native speakers to interpret such a collocation as 'hoping to say that ...' means that this is not due to the meaning of the sentence, but due to the interpretation of the sentence in a certain context. However, the fact that native speakers are so reluctant to interpret the collocation *hi*$^{1-7}$*bang*$^{7-3}$ *kong*$^{2-1}$ as 'hoping to say that...' and that the interpretation 'hoping that [dependent clause]' is widely preferred are important indications of the degree of grammaticalization of this

collocation and of *kong*$^{2-1}$ in particular. Another important indication of the degree of grammaticalization is that for native speakers it is impossible to interpret a collocation such as *ua*$^{2-1}$ *ma*$^{7-3}$ *hi1*$^{-7}$*bang*$^{7-3}$ *kong*$^{2-1}$ as 'I also hope that X say(s)' where X stands for either a first person plural or a second or third person singular or plural agent. In other words: it is impossible for *kong*$^{2-1}$ and the verb preceding it to have different subjects when these subjects are not stated explicitly.

If we look at example (2), repeated below, we see that it does not allow for an interpretation in terms of 'say' being part of the meaning.

(2) A^{0} soo$^{2-1}$i$^{2-1}$ kho$^{2-1}$ling^{5} =*ne*, eh-- koh$^{4-1}$ ui$^{7-3}$ thiann$^{1-7}$tsiong$^{3-2}$ ping$^{5-7}$iu^{2} e$^{7-3}$ **kam$^{2-1}$kak$^{4-8}$**
RLT therefore maybe =RLV HST every place listening.audience friend will **feel**

kong$^{2-1}$, m^{0} khak$^{4-8}$sit^{8} =*ne*, tsin$^{1-7}$tsiann^{2} tsainn$^{1-7}$iann$^{2-1}$-- beh$^{4-1}$ tsainn$^{1-7}$iann$^{2-1}$
speak AGR indeed =RLV really know want.to know

i$^{1-7}$ e$^{5-7}$ i$^{3-2}$su^{3}.
3 SUB meaning

'So it may be that uhm-- our listeners will feel that, hm yes, they really know-- want to know its meaning.' (3A; 05:40)

Here, we are not dealing with a combination of *hi*$^{1-7}$*bang*$^{7-3}$ 'hope' with *kong*$^{2-1}$ 'say' and a dependent clause, but with the verb *kam*$^{2-1}$*kak*$^{8-4}$ 'feel' with *kong*$^{2-1}$ and a dependent clause. Again, according to native speakers, it is not possible to interpret this sentence literally, as 'So it may be that uhm-- our listeners will have the feeling to say "hm yes, we really want to know-- want to know its meaning."'.

I believe that if a certain linguistic utterance can only be used to convey one notion, it has a fixed meaning which is somehow linguistically coded in the utterance. This meaning can be coded either lexically, morphologically or syntactically. Let us look at example (26a) and (26b) to illustrate this last point.

(26) a. Peter loves Mary.
b. Mary loves Peter.

Both (26a) and (26b) describe a situation in which a person loves another person and in both sentences, Peter and Mary are referred to as playing a role in this situation. To a native speaker of English, (26a) can only be used to describe a situation in which Peter is the one who is in love with Mary, whereas in (26b) Mary is the one in love with Peter. There is no morphological or lexical

difference between these two sentences, and still they are used to describe two different situations. The only structural difference between (26a) and (26b) is the difference in word order. To me, this is a clear example where syntax, in this case word order, encodes meaning: in both sentences, the noun following the verb *loves* is associated with the one being loved and the noun preceding the verb is associated with the one who is the lover.

In the case of (22), this means that the semantic connection between *hi*$^{1-7}$*bang*$^{7-3}$ 'hope' and *i*$^{1-7}$ *e*$^{7-3}$*tang*$^{3-2}$ *sio*$^{2-1}$ *kan*^{9} *kai*$^{2-1}$*sueh*^{4} *tsit*$^{8-0}$*e*$^{7-0}$ 'he can explain it a bit to us' must be due to a syntactic connection of some sort. The most obvious syntactic analysis sees *i*$^{1-7}$ *e*$^{7-3}$*tang*$^{3-2}$ *sio*$^{2-1}$ *kan*^{9} *kai*$^{2-1}$*sueh*^{4} *tsit*$^{8-0}$*e*$^{7-0}$ 'he can explain it a bit to us' as the dependent clause of the verb *hi*$^{1-7}$*bang*$^{7-3}$ 'hope'. It has already been noted that in contrast to sentences such as (20) and (23), where the dependent clause immediately follows the verb, this is not the case in (22). My analysis of the combination of verb + *kong*$^{2-1}$ + dependent clause is that *kong*$^{2-1}$ is used to introduce the dependent clause of the verb preceding *kong*$^{2-1}$, similar to *that* preceding dependent clauses in English, cf. (14b) in § 2.1.1. In other words, *kong*$^{2-1}$ is analysed as a complementizer. In both (22) and (23), we see the verb *hi*$^{1-7}$*bang*$^{7-3}$ 'hope' and a clause that is syntactically dependent on it and semantically refers to the thing that is hoped, but only in (22) these two are connected by the complementizer *kong*$^{2-1}$. The grammaticality of both constructions indicates that the use of *kong*$^{2-1}$ as a complementizer between a verb and its dependent clause is optional. In this respect, TSM seems similar to English, as illustrated in examples (14b) and (14c). It is also similar to these examples in that the absence of the complementizer *kong*$^{2-1}$ in TSM or *that* in English can lead to ambiguity as to the presentation of the reported speech, viz. whether it is a case of direct or indirect speech. There is also a difference. The English sentences in (14b) and (14c) have the matrix verb *say*, whereas the TSM sentence has *kong*$^{2-1}$, originally a verb meaning 'say', as its complementizer. A similar analysis can be used for (2), repeated below.

(2) A^{0} soo$^{2-1}$i$^{2-1}$ kho$^{2-1}$ling^{5} =*ne*, eh-- koh$^{4-1}$ ui$^{7-3}$ thiann$^{1-7}$tsiong$^{3-2}$ ping$^{5-7}$iu^{2} e$^{7-3}$ **kam$^{2-1}$kak$^{4-8}$**
RLT therefore maybe =RLV HST every place listening.audience friend will **feel**

kong$^{2-1}$, m^{0} khak$^{4-8}$sit^{8} =*ne*, tsin$^{1-7}$tsiann^{2} tsainn$^{1-7}$iann$^{2-1}$-- beh$^{4-1}$ tsainn$^{1-7}$iann$^{2-1}$
speak AGR indeed =RLV really know want.to know

i$^{1-7}$ e$^{5-7}$ i$^{3-2}$su^{3}.
3 SUB meaning

'So it may be that uhm our listeners will feel that, hm yes, they really know-- want to know its meaning.' (3A; 05:40)

Here, the clause *tsin*$^{1\text{-}7}$*tsiann*^{2} *tsainn*$^{1\text{-}7}$*iann*$^{2\text{-}1}$-- *beh*$^{4\text{-}1}$ *tsainn*$^{1\text{-}7}$*iann*$^{2\text{-}1}$ *i*$^{1\text{-}7}$ *e*$^{5\text{-}7}$ *i*$^{3\text{-}2}$*su*^{3} 'they really know--want to know its meaning' is syntactically dependent on the verb *kam*$^{2\text{-}1}$*kak*$^{4\text{-}8}$ 'feel' and is introduced by the complementizer *kong*$^{2\text{-}1}$. The phrase *m*^{0} *khak*$^{4\text{-}8}$*sit*^{8} *ne* 'hm yes' is interjected into the sentence and signals that the speaker, after having swiftly rethought what she wanted to say, again arrives at the conclusion that she believes what is expressed with the dependent clause to be the case.

The functions of TSM *kong*$^{2\text{-}1}$ as a complementizer have been discussed by Gāo (2007: 101-104), Tseng (2008: 39), Lau (2013: 66) and others. It is remarkable, however, that none of these authors mention the tone sandhi which *kong*$^{2\text{-}1}$ invariably exhibits when it functions as a complementizer. In its function as a transitive verb, we come across both the unsandhied form *kong*^{2}, e.g. (9), or in its sandhied form *kong*$^{2\text{-}1}$, e.g. (12). In its function as a complementizer, however, only the sandhied form *kong*$^{2\text{-}1}$ is attested.

Note that following this sandhi form *kong*$^{2\text{-}1}$ and preceding the dependent clause, the speaker can optionally pause to think or catch his breath (cf. (30) below). The fact that the above cited works do not mention tone sandhi in the context of the of *kong*$^{2\text{-}1}$ as a complementizer may be due to the way their data are encoded. Gāo (2007) gives all Taiwanese examples in Chinese characters. This script does not indicate tone. Tseng (2008: 2) states that he uses a phonetic transcription for his TSM examples that is based on the phonetic transcription endorsed by the Taiwanese Ministry of Education, but adds that he "will neglect the tonal system because of the problem of tone sandhi". In the presentation of his examples, Tseng not only uses a romanization system, he also uses characters to transcribe the examples – something which he also did not mention in the introduction. Contrary to what he seemed to imply, in his transcriptions he does not completely neglect the TSM tones. He marks them with the numbers 1-8, skipping number 6 – as is conventional in Southern Mǐn tonal numbering, cf. § 1.2. He does neglect, however, both the neutral tone and tone sandhi. It is not clear to me why he has opted to do so, nor is it clear to me what "problem of tone sandhi" he is referring to. In the remainder of his 118-pages-long thesis, he mentions tone sandhi just once (*ibid.*: 44), but not in the context of *kong*$^{2\text{-}1}$ as a complementizer. Lau (2013) presents his examples both in Chinese characters and in the Tâi-lô romanization system. He does not transcribe tone sandhi, however, nor does he mention tone sandhi anywhere in his article.

I think this is a good illustration of the necessity of good data and good transcription of that data. Most of the time, a lot of linguistically relevant information is lost in the process of transcription. Prosody is usually the first victim, and my *ad hoc* use of punctuation cannot do justice to its intricacies. In the case of tonal languages such as TSM, if one does not transcribe tones or the interaction of tones in certain morpho-syntactic contexts (viz. tone sandhi) much more linguistically

relevant information is lost. In the case of Taiwanese, this information is both of a lexical nature (i.e. tones are used to distinguish different between lexical items) and of a morpho-syntactic nature (i.e. tone sandhi is a cue for indicating syntactic connection between certain elements).

There is more to this use of *kong*$^{2-1}$ than the function of complementizer in the sense of English *that*. In (18), repeated below, we see that following *kong*$^{2-1}$ there is a quotation, i.e. direct speech.

(18) A^{0} soo$^{2-1}$i$^{2-1}$ ji^{7} lan$^{2-1}$ ti$^{7-3}$ Tai$^{5-7}$uan^{5} n-- thong$^{1-7}$siong^{5} long$^{2-1}$ e$^{7-3}$ hong^{9}
RLT therefore character 1INCL be.at Taiwan HST usually in.every.case will give:person

goo$^{7-3}$kai$^{2-1}$ kong$^{2-1}$, a^{0} ji^{7}, toh$^{8-3}$ si$^{7-3}$ Khong$^{2-1}$tsu$^{2-1}$ ji^{7}, toh$^{8-3}$ si$^{7-3}$
misunderstand say RLT character just be Confucius character just be

han$^{3-2}$ji^{7}, kio$^{3-2}$ tso$^{3-2}$ ji^{7}, <hmm> honn^{0}.
Chinese.character call do character hmm XPC

'So *jī* 'characters', we in Taiwan n-- usually always misunderstand them like, well *jī* are *Khóng-tsú jī* 'Confucius' characters', *hàn-jī* 'Chinese characters' in other words, that is what are called *jī* <hmm>, right?' (34C; 12:45)

One might say that this is yet another example of *kong*$^{2-1}$ 'say' as a transitive verb which, as we have discussed in § 2.1.1, can be followed by both direct and indirect speech. In this case, it is followed by direct speech: *a^{0} ji^{7}, toh$^{8-3}$ si$^{7-3}$ Khong$^{2-1}$tsu$^{2-1}$ ji^{7}, toh$^{8-3}$ si$^{7-3}$ han$^{3-2}$ji^{7}, kio$^{3-2}$tso$^{3-2}$ ji^{7} 'well jī are Khóng-tsú jī 'Confucius' characters', hàn-jī 'Chinese characters' in other words, that is what are called jī'*.[13] In (18), there also is the verb *goo$^{7-3}$kai$^{2-1}$* 'misunderstand'. This is a transitive verb and as such reference to a difinite direct object is made, which in this case is co-referential with the topic of the sentence, viz. *ji^{7}* 'characters'. In the translation, we see the word *them* which is the English equivalent of the definite direct object present in the meaning of the TSM in (18). It is interesting to note, however, that an analysis of (18) as being two consecutive and independent sentences (i.e. a sentence ending with the transitive verb *goo$^{7-3}$kai$^{2-1}$* 'misunderstand', cf. (26); and a sentence consisting of *kong*$^{2-1}$ 'say' followed by a quotation, cf. (27)) seems problematic for a morpho-syntactic reason:

*(26) A^{0} soo$^{2-1}$i$^{2-1}$ ji^{7} lan$^{2-1}$ ti$^{7-3}$ Tai$^{5-7}$uan^{5} n-- thong$^{1-7}$siong^{5} long$^{2-1}$ e$^{7-3}$ hong^{9}
RLT therefore character 1INCL be.at Taiwan HST usually in.every.case will give:person

13 Note that in this case, we cannot not quite be certain where the direct speech ends and Speaker C takes over again. It is certain that the first part (*a^{0} ji^{7}, toh$^{8-3}$ si$^{7-3}$ Khong$^{2-1}$tsu$^{2-1}$ ji^{7}* 'well jī are *Khóng-tsú* jī 'Confucius' characters'') is direct speech; the second part (*toh$^{8-3}$ si$^{7-3}$ han$^{3-2}$ji^{7}* 'so *hàn-jī* 'Chinese characters'') and third part (*kio$^{3-2}$tso$^{3-2}$ ji^{7}* 'that is what are called *jī* 'characters'') may either fall within or outside the range of the direct speech. This does not impact the argument presented here.

goo$^{7-3}$kai$^{2-1}$.
misunderstand

(Intended meaning: 'So *jī* 'characters', we in Taiwan n-- usually always misunderstand them.')

(27) **Kong$^{2-1}$,** a^{0} ji^{7}, toh$^{8-3}$ si$^{7-3}$ Khong$^{2-1}$tsu$^{2-1}$ ji^{7}, toh$^{8-3}$ si$^{7-3}$ han$^{3-2}$ji^{7}, kio$^{3-2}$
say RLT character just be Confucius character just be Chinese.character call

ji^{7}, <hmm> honn^{0}.
character hmm XPC

They say: "Well *jī* are *Khóng-tsú jī* 'Confucius' characters', so *hàn-jī* 'Chinese characters', that is what are called *jī*." <hmm>, right?'

If (18) really were to be analyzed as consisting of two sentences as shown in (26) and (27), the first of these would be ungrammatical. This is because for (26) to be a finished sentence, the second syllable of *goo$^{7-3}$kai$^{2-1}$* 'misunderstand' has to preserve its original tone, viz. *goo$^{7-3}$kai^{2}*.

In § 1.1.2 under the sections of tone, I discussed tone sandhi and noted that the final syllable of a sentence or phrase does not exhibit tone sandhi. Since *goo$^{7-3}$kai$^{2-1}$* 'misunderstand' in (18) does exhibit tone sandhi, *goo$^{1-7}$kai$^{2-1}$* cannot be the final syllable in a sentence as presented in (26) and hence (18) cannot be analysed as two independent sentences. Or, phrased positively, since the verb *goo$^{7-3}$kai$^{2-1}$* exhibits tone sandhi on its second syllable, a mandatory, syntactic connection with the following *kong$^{2-1}$* and the quotation is established. Such instances of tone sandhi are also seen with e.g. *kam$^{2-1}$kak$^{4-8}$* 'feel' in (2) and *hi$^{1-7}$bang$^{7-3}$* 'hope' in (22). Semantically, the quotation in (18) illustrates the way people in Taiwan might phrase their mistaken concept of *ji^{7}* 'character' according to speaker C.

Because of this tone sandhi, signaling a syntactic grouping of the verb *goo$^{7-3}$kai$^{2-1}$* and *kong$^{2-1}$* plus the quotation, sentences such as (18) can be seen as the counterpart of the previously discussed construction of verb + *kong$^{2-1}$* + dependent clause which we find in sentences such as (22) – the former being constructed with direct speech (i.e. the quotation) and the latter with indirect speech (i.e. the dependent clause).

Before finishing this section, I need to get back to example (19).

(19) A^{0} khi$^{5-7}$sit^{8} li$^{2-1}$ na$^{7-3}$ thak$^{8-4}$ hit$^{4-8}$ lo^{7}-- lan$^{2-1}$ jin$^{5-7}$lui^{7} hit$^{4-8}$ e$^{5-7}$ bun$^{5-7}$ji^{7} e$^{5-7}$ ian$^{2-1}$piat^{4}
RLT actually 2SG if study that sort 1INCL humanity that entity writing.system SUB evolve

li$^{2-1}$ toh$^{8-3}$ **tsai$^{1-7}$iann$^{2-1}$** **kong$^{2-1}$** khi$^{5-7}$sit^{8} tsit$^{8-4}$ tsiong$^{2-1}$ ji^{7} si$^{7-3}$ khah$^{4-1}$-- khah$^{4-1}$
2SG just **know** **say** actually one kind.of character be relatively relatively

tsa^{2} tsin$^{3-2}$tsing^{5} u^{7} e$^{5-0}$.
early before EXST SUB

‘And actually if you study uhm-- the evolution of mankind's writing systems, you know that actually one kind of *jī* ‘writing’ existed relatively-- relatively early on.’ (35C; 12:55)

In my previous discussion of (19), I left off with the dilemma how to interpret the second half of the sentence syntactically. By now, it is clear that in (19), the clause *khi*$^{5-7}$*sit*^{8} *tsit*$^{8-4}$ *tsiong*$^{2-1}$ *ji*^{7} *si*$^{7-3}$ *khah*$^{4-1}$-- *khah*$^{4-1}$ *tsa*^{2} *tsin*$^{3-2}$*tsing*^{5} *u*^{7} *e*$^{5-0}$ ‘actually one kind of *jī* ‘writing’ existed relatively-- relatively early on’ functions as the dependent clause of the verb *tsai*$^{1-7}$*iann*$^{2-1}$ ‘know’, *kong*$^{2-1}$ functions as a complementizer introducing the dependent clause to the matrix verb. This analysis is based on the interpretation of *khi*$^{5-7}$*sit*^{8} *tsit*$^{8-4}$ *tsiong*$^{2-1}$ *ji*^{7} *si*$^{7-3}$ *khah*$^{4-1}$-- *khah*$^{4-1}$ *tsa*^{2} *tsin*$^{3-2}$*tsing*^{5} *u*^{7} *e*$^{5-0}$ as not being a quotation. However, in this case, one might argue that the clause is in fact a quotation and thus direct speech. The quotation would then be an exemplification of what the *li*$^{2-1}$ ‘you’ knows when he studies “the evolution of mankind’s writing systems”. In other words, what is known would be expressed using direct speech instead of indirect speech. This reading could be translated in English as ‘...then you know: “Actually, one kind of character existed relatively-- relatively early on.”’

In their ambiguity, sentences such as the one in (19), establish a link between cases such as (18) in which direct speech is involved and cases such as (22) in which indirect speech is involved. This construction can be rendered abstractly as in (28).

(28) V + *kong*$^{2-1}$ + CLAUSE

Here, V can be any of a certain set of verbs and CLAUSE is a clause that denote that which is V’d. This formulation captures structures such as the one found in (18) as well as those such as the one found in (22). In TSM, these can be considered instances of one and the same construction. I use the term construction to denote a linguistic pattern of which the meaning is different from the combined meaning of its constituents. In (29), we see an instantiation of the construction shown in (28): *tsai*$^{1-7}$*iann*$^{2-1}$(V) + *kong*$^{2-1}$ + *khi*$^{5-7}$*sit*^{8} *tsit*$^{8-4}$ *tsiong*$^{2-1}$ *ji*^{7} *si*$^{7-3}$ *khah*$^{4-1}$-- *khah*$^{4-1}$ *tsa*^{2} *tsin*$^{3-2}$*tsing*^{5} *u*^{7} *e*$^{5-0}$ (CLAUSE) ‘you know that actually one kind of *jī* ‘writing’ existed relatively-- relatively early on’. Since *kong*$^{2-1}$ ‘say’, here, does not denote the meaning ‘you know that you say that...’ or ‘you know to say that...’ (more on this in the next subsection), the meaning of the constituents does not add up to the meaning of V + *kong*$^{2-1}$ + CLAUSE, which is therefore viewed as a construction.

Types of verbs constructed with the complementizer kong$^{2\text{-}1}$

The above formulation of "any of a certain set of verbs" calls for semantic specification. In this subsection, I aim to classify all verbs constructed with the complementizer *kong*$^{2\text{-}1}$ in my corpus. Roughly speaking, the instances of matrix verbs constructed with the complementizer *kong*$^{2\text{-}1}$ can be classified on morpho-syntactic grounds and on semantic grounds.

Out of a total of 83 instances of *kong*^{2} ~ *kong*$^{2\text{-}1}$ in my corpus, there are 17 cases of *kong*$^{2\text{-}1}$ appearing as the complementizer to a verb, which rounded off is 20%. Of these 17 cases, 14 are plain verbs without suffixes (17% of the total of 83 instances) and 3 are verbs with a suffix (4% of the total). Of the plain verbs, we already saw examples in the previous section – cf. (2), (18), (19), (22), etc. above. The three instances in which a suffix had been added to the matrix verb are listed below:

(29) Bo$^{5\text{-}7}$ **siunn$^{7\text{-}3}$-tioh$^{8\text{-}3}$** **kong$^{2\text{-}1}$** Tai$^{5\text{-}7}$gi^{2} tsia^{1} u$^{7\text{-}3}$ hak$^{8\text{-}4}$bun^{5} ho^{0}.
NEG:EXST **think-get** **say** Taiwanese that.much EXST erudition XPC

'I never would have thought that there was so much to speaking Taiwanese, you know.' (12A; 09:06)

(30) <*Qíng* *Sū*> U$^{7\text{-}3}$ <*lǎosī* a--> tiann$^{7\text{-}3}$tiann$^{7\text{-}3}$ te^{1} khuann$^{3\text{-}2}$ tian$^{7\text{-}3}$iann^{2} e$^{5\text{-}7}$
invite PNSū EXST teacher HST often PROG look movie SUB

lang^{5} honn^{0}, <henn^{0}> kho$^{2\text{-}1}$ling$^{5\text{-}7}$ e$^{7\text{-}3}$ h-- i^{1} **tsu$^{2\text{-}1}$i$^{3\text{-}2}$-tioh$^{8\text{-}3}$** **kong$^{2\text{-}1}$**, hit$^{4\text{-}8}$ e$^{5\text{-}7}$
person XPC CFM possible will HST 3 **notice-get** **say** that entity

Hiong$^{1\text{-}7}$kang$^{2\text{-}1}$phinn^{3} honn^{0} <henn^{0}>, i-n$^{1\text{-}7}$ e$^{5\text{-}7}$ ue^{7} kah$^{4\text{-}8}$ i-n e$^{5\text{-}7}$ ji^{7}
Hong.Kongese.movie XPC CFM 3-PL SUB speech and 3-PL SUB character

he^{1} long$^{2\text{-}1}$ ham$^{7\text{-}3}$ gua-n$^{2\text{-}1}$ *sǒwèi* *=de* *Guóyǔ* bo$^{5\text{-}7}$ kang$^{7\text{-}3}$
that:entity in.every.case with 1-PL so.called =SUB National.language NEG.EXST same

khuan^{2} e$^{5\text{-}0}$ <tioh^{8} la^{0}> <tioh^{8} tioh^{8}>.
type SUB right WARN right right

'<I want to ask Mr.> There are <Sū a--> the people that often are watching films, right, <yes> maybe they will h-- they notice that, those Hong Kong movies, you know <yes>, their language and Chinese characters those are all different from our so-called *Guóyǔ* 'National language' <indeed> <right, right>.' (53B; 16:57)

Here, we see the verbs *siunn*$^{7\text{-}3}$ 'think' and *tsu*$^{2\text{-}1}$*i*$^{3\text{-}2}$ 'notice' both suffixed with *-tioh*$^{8\text{-}3}$ 'get', which can be used to denote the action of the preceding verb has led to a result. In (29) the speaker uses

this suffix to convey that it never would have occured to her that "there was so much to speaking Taiwanese". That the action *siunn*^{7} 'think' has led to a result (i.e. *siunn*$^{7-3}$*-tioh*^{8}) is what is negated by the negative existential *bo*$^{5-7}$ 'NEG:EXST', and not merely the act of *siunn*^{7} 'think'.

In (30), the same suffix *-tioh*$^{8-3}$ 'get' is attached to the verb *tsu*$^{2-1}$*i*$^{3-2}$ 'notice'. This verb regularly occurs with this suffix, matching the semantics of the verb *tsu*$^{2-1}$*i*$^{3-2}$ 'notice', which itself indicates that an action has led to a result. In other words, the act of noticing something involves a specific point in time where you can say "Now one notices something.".

Both *siunn*$^{7-3}$*-tioh*$^{8-3}$ and *tsu*$^{2-1}$*i*$^{3-2}$*-tioh*$^{8-3}$ are followed by the complementizer *kong*$^{2-1}$ and a dependent clause. This is clear both from the meaning of (29) and (30) and from the morpho-syntax of these verb phrases: in both cases, the final syllable of the verb phrase, i.e. the suffix *-tioh*$^{8-3}$, exhibits tone sandhi. Because of this tone sandhi, the only interpretation can be that the verb phrases *siunn*$^{7-3}$*-tioh*$^{8-3}$ and *tsu*$^{2-1}$*i*$^{3-2}$*-tioh*$^{8-3}$ are connected with the immediately following *kong*$^{2-1}$, which, as argued in the discussion of (26), corresponds with a semantic connection.

(31) A^{0}-- A^{0} jin$^{5-7}$lui^{7} e$^{5-7}$ kiong$^{7-3}$iu$^{2-1}$ tsu$^{1-7}$san^{2}, a^{0} lan$^{2-1}$ na$^{7-3}$ long$^{2-1}$ h-- ka$^{7-3}$-- ka$^{7-3}$
RLT RLT humanity SUB shared property RLT 1INCL if in.every.case HST PRTC PRTC

pang$^{3-2}$-tiau$^{7-3}$ **kong$^{2-1}$** he^{1} long$^{2-1}$ Se$^{1-7}$hong^{1} e$^{5-0}$ a^{0} koh$^{4-1}$ lan$^{2-1}$ bo$^{5-7}$ ai^{3}...
release-away **say** that:entity in.every.case West SUB RLT again 1INCL NEG.EXST want

'But-- but mankind's shared property, but if we all th-- do away with it like "those are all from the West, and we don't want it"...' (40C; 14:11)

In (31) too, we see a verb, *pang*$^{3-2}$ 'release', with a suffix, *-tiau*$^{7-3}$ 'away', followed by the complementizer *kong*$^{2-1}$. The suffix *-tiau*$^{7-3}$ is used to denote that the patient of the verb *-tiau*$^{7-3}$ attaches to not only undergoes the action denoted by the verb, but also is removed from its original place. This can be used to describe physical as well as psychological removal from a certain place. In (31), we are dealing with the latter case: the speaker states that he thinks that when it comes to writing Taiwanese, it is foolish to disregard the Roman alphabet merely because it comes from the West and is supposedly not indigenous of Taiwanese culture, as opposed to Chinese characters. Here, the verb *pang*$^{3-2}$ 'release' on its own cannot convey the meaning that the Roman alphabet is done away with by some people. In order to convey this meaning, speaker C attaches the suffix *-tiau*$^{7-3}$ to the verb *pang*$^{3-2}$.

As with (29) and (30), in (31), the verb phrase *pang*$^{3-2}$*-tiau*$^{7-3}$ exhibits tone sandhi on its final syllable, which means the only possible interpretation is that of connecting it with *kong*$^{2-1}$ and the dependent clause, with *kong*$^{2-1}$ functioning as a complementizer.

Apart from morpho-syntactic structure, in the discussion above, I also mentioned a second way of classifying the matrix verbs that are used with *kong*$^{2\text{-}1}$ as a complementizer in my corpus. Before giving a semantic classification of the attested matrix verbs, I would like to discuss collocations of certain types of verb followed by *kong*$^{2\text{-}1}$ not found in my corpus.

For example, from personal experience, I know that *kong*$^{2\text{-}1}$ can also follow verbs of speech such as *mng*^{7} 'ask' or *thong*$^{1\text{-}7}$*ti*^{1} 'inform' as the following examples taken from Gão (2007; transcriptions and translations are my own based on the pronunciation of native speakers) show:

(32) Ong^{5} e$^{5\text{-}0}$ ka^{7} **thong**$^{1\text{-}7}$**ti**$^{1\text{-}7}$ **kong**$^{2\text{-}1}$ Li^{2} e$^{5\text{-}0}$ it$^{4\text{-}8}$ting$^{7\text{-}3}$ si$^{7\text{-}3}$ beh$^{4\text{-}1}$ be$^{2\text{-}1}$ siann^{2}.
PNÔng SUB OBJ **notify** **say** PNLí SUB definitely be want.to buy what

'Ông said notified him that Lí wanted to buy something for sure.' (Gão 2007: 73)

(33) I$^{1\text{-}7}$ le^{1} **mng**$^{7\text{-}3}$ **kong**$^{2\text{-}1}$ li$^{2\text{-}1}$ kam^{1} tsiah$^{8\text{-}3}$ png^{7} a^{0}.
3 PROG **ask** **say** 2SG QW eat cooked.rice PERF

'She is asking if you have eaten.' (Gão 2007: 87)

As to the question why there are no examples of verbs of speech in my corpus, I can offer only the following suggestion, but no definitive answer. The TSM verb corresponding to English *say* and *talk* is the verb *kong*^{2} itself, as has been discussed in § 2.1. It seems that the combination **kong*$^{2\text{-}1}$ *kong*$^{2\text{-}1}$ (intended meaning) 'to say that...' is not used in TSM. There are two possibile analyses: the first is that the structure *kong*$^{2\text{-}1}$ *kong*$^{2\text{-}1}$ (i.e. V + C) is uttered as *kong*$^{2\text{-}1}$ through haplology. Haplology is the change of successive syllables which are similar in form being reduced to one (Matthews 2007: 170). The second possibility is that the formation of the construction *kong*$^{2\text{-}1}$ 'say' and *kong*$^{2\text{-}1}$ 'COMPL' is impossible in the first place and that the structure of *kong*$^{2\text{-}1}$ is not V + C, but just V. I do not know the answer to this conundrum. It seems that verbs of speech other than *kong*^{2} 'say' are used significantly less often. In the sources of my corpus, I found not a single instance of a verb of speech other than *kong*^{2} (e.g. *kang*^{2} 'discuss', *kai*$^{3\text{-}2}$*siau*^{7} 'introduce', *mng*^{7} 'ask') being used as the matrix verb of a dependent clause.

Looking at Table 6, we may notice two things. The first is that all matrix verbs constructed with the complementizer *kong*$^{2\text{-}1}$ are verbs denoting psychological events such as *siunn*$^{7\text{-}3}$ 'think', *liau*$^{2\text{-}1}$*kai*$^{2\text{-}1}$ 'understand' and *tsu*$^{2\text{-}1}$*i*$^{3\text{-}2}$*-tioh*$^{8\text{-}3}$ 'notice'. I take the term psychological action to mean any action that takes place only in a person's mind instead of taking place in the physical world. An action such as thinking, for example, takes place in one's mind, whereas an action such as hitting, though being directed by the mind, takes place in the world outside of it. We

Table 6: Semantic types of verbs

PSYCHOLOGICAL VERBS				OTHER		
Cognitive verbs		**Perceptive verbs**				**Count**
Form	**Meaning**	**Form**	**Meaning**	**Form**	**Meaning**	
hi[1-7]*bang*[7-3]	'hope'					1
siunn[7-3]	'think'					1
liau[2-1]*kai*[2-1]	'understand'					1
goo[7-3]*kai*[2-1]	'misunderstand'					1
tsai[1-7]*iann*[2-1]	'know'					1
tsainn[1-7]*iann*[2-1]	'know'					1
kam[2-1]*kak*[4-8]	'feel'					3
kam[2-1]*ma*[2-1]	'feel'					1
		siunn[7-3]*-tioh*[8-3]	'come to think about'			1
		huat[4-8]*hian*[7-3]	'discover'			1
		pai[5-7]*thik*[8-4]	'exclude'			1
		pang[3-2]*-tiau*[7-3]	'do away with'			1
		tsu[2-1]*i*[3-2]*-tioh*[8-3]	'notice'			1
				tshiunn[7-3]	'resemble'	2
					Total:	17

can classify all but *tshiunn*[7-3] as either cognitive verb or perceptive verb. By *cognitive verb*, I mean a verb denoting a psychological state. *Tsai*[1-7]*iann*[2-1] 'know', for example, denotes a state in which one is aware of something. By *perceptive verb*, I mean a verb denoting that a psychological state is reached. *Huat*[4-8]*hian*[7-3] 'discover', for example, denotes the process which leads to the state that one knows something. There are four verbs which might be considered not to be psychological verbs. These are *kam*[2-1]*kak*[4-8] and its variant *kam*[2-1]*ma*[2-1] 'feel', *pai*[5-7]*thik*[8-4] 'exclude' and *pang*[3-2]*-tiau*[7-3] 'do away with'.

Kam[2-1]*kak*[4-8] and *kam*[2-1]*ma*[2-1] can both be used to indicate that someone feels something physically, for example with his hands. By extension, however, both these verbs can be used to indicate psychological feeling. The same is the case with the English verb 'feel'. In my corpus all three instances of *kam*[2-1]*kak*[4-8] with the complementizer *kong*[2-1] and the single instance of *kam*[2-1]*ma*[2-1] with the complementizer *kong*[2-1] are used to denote psychological feeling, instead of

physical feeling – cf. example 1A, 3A, the first instance of *kong*$^{2\text{-}1}$ in 7A and the third instance of *kong*$^{2\text{-}1}$ in 29C of Appendix 1. *Kam*$^{2\text{-}1}$*kak*$^{4\text{-}8}$ and *kam*$^{2\text{-}1}$*ma*$^{2\text{-}1}$ in Table 6 are therefore properly categorized as cognitive verbs.

The final two verbs of which categorization as a psychological verb might be contested are *pai*$^{5\text{-}7}$*thik*$^{8\text{-}4}$ 'exclude' and *pang*$^{3\text{-}2}$*-tiau*$^{7\text{-}3}$ 'do away with'. To start with the latter, in the discussion of the morpho-syntactic structure of the verb *pang*$^{3\text{-}2}$*-tiau*$^{7\text{-}3}$, it was noticed that *pang*$^{3\text{-}2}$*-tiau*$^{7\text{-}3}$ consists of the verb *pang*$^{3\text{-}2}$ 'release' suffixed with *-tiau*$^{7\text{-}3}$ 'away'. In itself, this verb *pang*^{3} 'release' is not a psychological verb – though it can be used as one – but in (31), suffixed with *-tiau*$^{7\text{-}3}$ 'away', it is used to describe the act of disregarding or dismissing the Roman alphabet. The same holds true for *pai*$^{5\text{-}7}$*thik*$^{8\text{-}4}$ 'exclude'. In itself, the verb *pai*$^{5\text{-}7}$*thik*^{8} 'exclude' is not a psychological verb – one can physically exclude or push away other people – but in my corpus (cf. 39C of the Appendix), it is used to describe the act of dismissing the Roman alphabet as some Taiwanese do, according to speaker C. Thus, in these two cases, *pai*$^{5\text{-}7}$*thik*$^{8\text{-}4}$ 'exclude' and *pang*$^{3\text{-}2}$*-tiau*$^{7\text{-}3}$ 'do away with' are used as psychological verbs.

Other authors writing about the types of matrix verb that can occur with the complementizer *kong*$^{2\text{-}1}$ have made different categorizations. For example, Tseng (2008: 75-80) categorizes the verbs *siunn*^{7} 'think', *tsai*$^{1\text{-}7}$*iann*^{2} 'know' and *kam*$^{2\text{-}1}$*kak*^{4} 'feel' as "cognition predicates". He defines "cognition" as "the proces of how to get knowledge or apply knowledge" (*ibid.*: 75). I assume that Tseng would consider the forms *siunn*$^{7\text{-}3}$*-tioh*$^{8\text{-}3}$, *tsainn*$^{1\text{-}7}$*iann*$^{2\text{-}1}$ and *kam*$^{2\text{-}1}$*ma*$^{2\text{-}1}$ as belonging to this group as well. On the other hand, Tseng categorizes the verbs *liau*$^{2\text{-}1}$*kai*^{2} 'understand' and *huat*$^{4\text{-}8}$*hian*^{7} 'discover' as "assertives", which he cryptically defines as "the speaker's belief to [sic] the factualness of the proposition, including truth or false condition" (*ibid.*: 78). The verbs *hi*$^{1\text{-}7}$*bang*^{7} 'hope' and *goo*$^{7\text{-}3}$*hue*^{7} 'misunderstand' – which is a near synonym of *goo*$^{7\text{-}3}$*kai*^{2} (MoE 2011b) – he classifies as "expressives" which according to Tseng (*ibid.*: 79) are used to "express the psychological state of the speaker". The "assertives" and the "expressives" are subgroups within what he calls "'[s]peech act' predicates" (*ibid.*: 77). His discussion of the differences between the different groups of matrix verbs is unclear to me, and I do not see why *goo*$^{7\text{-}3}$*hue*^{7} 'misunderstand' and *liau*$^{2\text{-}1}$*kai*^{2} 'understand' should not fall in the same category. I further find it unclear why a verb such as *siunn*^{7} 'think' has not been considered as an "assertive" or an "expressive". After all, if one thinks that something is the case, to me this does seem a valid instance of "the speaker's belief of the factualness of the proposition" – which would make *siunn*^{7} 'think' an assertive. The act of thinking also seems to "express the psychological state of the speaker", which would make *siunn*^{7} 'think' an expressive in Tseng's grouping.

In the preceding paragraphs, I described all verbs which I classify as cognitive or perceptive verbs in Table 6. The only morpheme which occurred preceding *kong*$^{2-1}$ + CLAUSE that is neither a cognitive verb nor a perceptive verb is the morpheme *tshiunn*$^{7-3}$ 'resemble'. I have two instances of *tshiunn*$^{7-3}$ used with *kong*$^{2-1}$ in my corpus, given in (34) and (35):

(34) A^{0} **tshiunn$^{7-3}$** **kong$^{2-1}$** Hua$^{5-7}$gi^{2} kong$^{2-1}$ *dì* *shàng* *=de* *dì* <hm> honn^{o}.
RLT **resemble** **say** Mandarin say earth above =SUB earth BKCH XPC

'Just like in Mandarin we say *dì* 'acreage' as in *dì shàng* 'on the acreage', <hm> right.' (47D; 15:24)

(35) **Tshiunn$^{7-3}$** **kong$^{2-1}$** tsit$^{4-8}$ e$^{5-7}$ thinn^{1} *=a* te^{7} la^{0} suann^{1} la tsui^{2}
resemble **say** this entity heaven =EC earth WARN mountain WARN water

la^{0} honn^{0} tsia^{1} e$^{5-7}$ han$^{3-2}$ji^{7} long$^{2-1}$ si$^{7-3}$ Tai$^{5-7}$gi^{2} ham$^{7-3}$ Hua$^{5-7}$gi^{2}
WARN XCP here SUB Chinese.character in.every.case be Taiwanese with Mandarin

i-n$^{1-7}$ iong^{7} e$^{5-7}$ ji^{7}, long$^{2-1}$ si$^{7-3}$ kang$^{7-3}$ khuan^{2} e^{0}.
3-PL use SUB character in.every.case be same type SUB

'Just like these *thinn* 'heaven', *tē* 'earth', *suann* 'mountain', *tsuí* 'water', you know, their Chinese characters are all characters used by Taiwanese and Mandarin, they're all the same.' (48D; 15:28)

(34) and (35) essentially follow the same pattern: we have the verb *tshiunn*$^{7-3}$ 'resemble' followed by the complementizer *kong*$^{2-1}$ and a clause. Literally, we can take this as 'it resembles [the case] that...' as in (34): 'It resembles [the case] that in Mandarin we say *dì* 'acreage' in *dì shàng* 'on the acreage', <hm> right.' > 'Just like in Mandarin we say *dì* 'acreage' in *dì shàng* 'on the acreage', <hm> right.'. The same goes for (35).

I have chosen not to assign *tshiunn*$^{7-3}$ 'resemble' to the same group as verbs such as *siunn*^{7} 'think', *hi*$^{1-7}$*bang*^{7} 'hope' and *goo*$^{7-3}$*kai*^{2} 'misunderstand' and mark it as a cognitive verb because, even though resembling is only apparent to a sentient being – that is, only in one's mind can a cloud resemble Ché Guevara – there is still a big difference with verbs such as *siunn*^{7}, *hi*$^{1-7}$*bang*^{7} or *goo*$^{7-3}$*kai*^{2}: the latter are used to say something about the one who is thinking, hoping or misunderstanding, whereas with *tshiunn*^{7} 'resemble', the fact that something resembles something else to someone is used as a qualification of the thing that resembles something else, and not as a qualification of the person to whom something resembles something else. Compare the following sentences:

(36) a. Paul hopes that his father will come to his wedding.

b. Paul resembles his father.

In (36a), it is in Paul's mind that the hoping "takes place" – the cognitive state of hoping is present in Paul's mind. In (36b), on the other hand, the psychological part of the fact that Paul resembles his father is not present in Paul's mind, but in someone else's – in this case, probably in the mind of the speaker of (36b). In the same way, *tshiunn*^{7} is different from my perceptive verbs: the latter are used to describe the person in whose mind a psychological state is reached, whereas *tshiunn*^{7}, as argued above, is not.

Tshiunn^{7} is thus different from the other verbs in Table 6 and I have therefore neither classified it as a cognitive verb nor as a perceptive verb. However, seeing that in my corpus only *tshiunn*$^{7\text{-}3}$ is not part of the other two categories, it is meaningless to come up with a third lable just for *tshiunn*$^{7\text{-}3}$. I have therefore labeled it as "other".

When we take another look at Table 6, a second phenomenon can be noticed: out of a total of 14 different verbs used in a V + *kong*$^{2\text{-}1}$ + CLAUSE-construction, there are 12 which occur only once, one (*tshiunn*$^{7\text{-}3}$ 'resemble') occurs twice and one (*kam*$^{2\text{-}1}$*kak*$^{4\text{-}8}$ 'feel') occurs three times. I will briefly come back to *tshiunn*$^{7\text{-}3}$ in § 2.3. Finally, in light of the high number of verbs that occur only once (12 out of 14), I think more research with larger corpora is needed to see which verbs occur more than others in the V + *kong*$^{2\text{-}1}$ + CLAUSE-construction.

2.2.2 Summary

In the preceding section we saw that 17 out of 83 instances of *kong*^{2} ~ *kong*$^{2\text{-}1}$ are ones in which *kong*$^{2\text{-}1}$ is used in the construction V + *kong*$^{2\text{-}1}$ + CLAUSE – this is 20% of all instances of *kong*^{2} ~ *kong*$^{2\text{-}1}$.

We also saw that in this construction, *kong*$^{2\text{-}1}$ is always attested in its sandhied form, as opposed to its function as a transitive verb – which was discussed in § 2.1.1 – where it is attested both in its unsandhied and in its sandhied form. The CLAUSE in this construction can be indirect speech, as for example in (22), or direct speech, as for example in (18). Sometimes, it seems that both an interpretation of the clause as direct speech and as indirect speech is possible, as in (19). I argue that regardless of whether we are dealing with direct or indirect speech, the construction is the same in TSM.

Finally, the matrix verb written V in the construction V + *kong*$^{2\text{-}1}$ + CLAUSE is for 10 out of 17 cases a so-called cognitive verb, such as *siunn*^{7} 'think', *hi*$^{1\text{-}7}$*bang*^{7} 'hope' or *liau*$^{2\text{-}1}$*kai*^{2}

'understand'. In 5 out of 17 cases, the matrix verb is a so-called perceptive verb. The remaining two times involved the morpheme *tshiunn*^{7} 'resemble'. Cases such as (29) and (30) (*siunn*$^{7\text{-}3}$*-tioh*$^{8\text{-}3}$ 'thought' and *tsu*$^{2\text{-}1}$*i*$^{3\text{-}2}$*-tioh*$^{8\text{-}3}$ 'notice' respectively) and (31) (*pang*$^{3\text{-}2}$*-tiau*$^{7\text{-}3}$ 'do away with') show that the verb can be a morphologically complex word, although most of the verbs, 14 out of 17 (or 17 % of all 83 instances of *kong*^{2} ~ *kong*$^{2\text{-}1}$), are morphologically simplex.

2.3 Morphological & syntactic phenomena

In the previous sections, two functions of *kong*^{2} ~ *kong*$^{2\text{-}1}$ in my corpus were discussed. Together, these accounted for 69 instances out of a total of 83, which is roughly 83%. I presented examples in which *kong*^{2} functioned as a transitive verb or as a complementizer between a matrix verb and its dependent clause. In this section, I will discuss some cases that seem to behave differently. To wrap these up, I will discuss grammaticalization in the context of *kong*^{2} ~ *kong*$^{2\text{-}1}$.

2.3.1 Syntactic ambiguity

Th distinction between the functions of *kong*^{2} ~ *kong*$^{2\text{-}1}$ as a transitive verb (§ 2.1) and as a complementizer (§ 2.2) has been established. There are, however, cases in which *kong*$^{2\text{-}1}$ can be interpreted as having either function, as (37) shows:

(37) A^{0} soo$^{2\text{-}1}$i$^{2\text{-}1}$ **siunn$^{7\text{-}3}$** **kong$^{2\text{-}1}$** =*a*, u$^{7\text{-}3}$ e$^{5\text{-}7}$ khah$^{4\text{-}1}$ kan$^{2\text{-}1}$tan^{1} e$^{5\text{-}7}$ tsit$^{4\text{-}8}$ e$^{5\text{-}7}$
RLT therefore **think** **say** =EC EXST SUB relatively simple SUB this entity

han$^{3\text{-}2}$ji$^{7\text{-}3}$ =*a*, **siunn$^{7\text{-}3}$** **kong$^{2\text{-}1}$,** Hua$^{5\text{-}7}$gi^{2} **kong$^{2\text{-}1}$** *tiāndì* =*de* *tiān*
Chinese.character =EC **think** **say** Mandarin **say** heaven.and.earth =SUB heaven

tioh^{8} <hmm> <hann^{0}> bo$^{5\text{-}0}$ honn^{0}, a^{0} lan$^{2\text{-}1}$ Tai$^{5\text{-}7}$gi^{2} ma$^{7\text{-}3}$ **kong$^{2\text{-}1}$** <thinn^{1}>
right BKCH CFM NEG.EXST XPC RLT 1INCL Taiwanese.language also **say** heaven

thinn$^{1\text{-}7}$ kah$^{4\text{-}1}$ thian^{1} =ma^{0} <henn^{0}> <tioh^{8}> honn^{0}?
heaven and heaven =DASS CFM right XPC

'So I want to say, some of the simpler characters, I want to say, in Mandarin we say the *tiān* 'heaven' of *tiāndì* 'heaven and earth', <hmm> right <yes>, in Taiwanese we also say <*thinn*> *thinn* 'heaven' and *thian* 'heaven', you know <yes> <right> right?' (45D; 15:11)

In this example, the word *kong*$^{2\text{-}1}$ appears four times. The analyses of the third and fourth instances of *kong*$^{2\text{-}1}$ are straightforward: both times, *kong*$^{2\text{-}1}$ 'say' functions as a transitive verb after which an object follows – *tiāndì de tiān* 'the *tiān* 'heaven' of *tiāndì* 'heaven and earth'' and *thinn*$^{1\text{-}7}$ *kah*$^{4\text{-}1}$ *thian*^{1} '*thinn* 'heaven' and *thian* 'heaven'' respectively. In addition to this, the fourth instance of *kong*$^{2\text{-}1}$ is also preceded by the nonmovable adverb *ma*^{7} ~ *ma*$^{7\text{-}3}$ 'also'. Nonmovable adverbs are adverbs that always immediately precede the predicate (Li & Thompson 2009: 322). Here *ma*$^{7\text{-}3}$ 'also' precedes *kong*$^{2\text{-}1}$ 'say', *kong*$^{2\text{-}1}$ must be the predicate and hence functions as a transitive verb.

Shifting our attention to the first two instances of *kong*$^{2\text{-}1}$ in (37), we find that determining the function of *kong*$^{2\text{-}1}$ is much less straightforward than with the latter two instances. Here, *kong*$^{2\text{-}1}$

is preceded by *siunn*$^{7-3}$. As we saw in (29), *siunn*$^{7-3}$ can be a verb with the meaning 'think'. It can also be an auxiliary verb with the meaning 'want to', as in *siunn*$^{7-3}$ *hioh*$^{4-2}$*khun*^{3} 'want to rest' (Khóo 1998: 425). Besides Khóo, Lôo (2003: 102) also mentions 想 [*siunn*^{7}] is an auxiliary verb in TSM; Lin (2015: 283) states that *siunn*^{7} by itself can only mean 'think'. According to him *siunn*^{7} can only be used in collocation with *beh*^{4} 'FUT' or *ai*^{3} 'have to; love' to arrive at the meaning 'want to', i.e. *siunn*$^{7-3}$*beh*^{4} 'want to' or *siunn*$^{7-3}$*ai*^{3} 'want to'. I asked two native speakers, one in his thirties and the other in her seventies, whether they thought *siunn*$^{7-3}$ *khi*$^{3-1}$ *Tai*$^{5-7}$*uan*^{5} (Khóo 1998: 424) was regular Taiwanese. Both found it strange and said they would say *siunn*$^{7-3}$*beh*$^{4-1}$ *khi*$^{3-1}$ *Tai*$^{5-7}$*uan*^{5} 'want to go to Taiwan'. Mr. Khóo was born in 1935, the District of Kaohsiong (1998: 作者簡介). His native language is Taiwanese. Since apparently, for some Taiwanese speakers *siunn*^{7} can be an auxiliary meaning 'want to', in this section, I will discuss the resulting ambiguity between *siunn*$^{7-3}$ *kong*$^{2-1}$ 'want to say' and *siunn*$^{7-3}$ *kong*$^{2-1}$ 'think that'. The reader should keep in mind that for some, *siunn*$^{7-3}$ can only function as a transitive verb meaning 'think'.

Auxiliary verbs only take other verbs as their complement (Khóo 1998: 424). This distinguishes them from transitive verbs, which take nouns as their complement. The collocation *siunn*$^{7-3}$ *kong*$^{2-1}$ can thus mean 'want to say', when *siunn*$^{7-3}$ is interpreted as the auxiliary meaning 'want to' and *kong*$^{2-1}$ is interpreted as the transitive verb 'say'.

Following the second *siunn*$^{7-3}$ *kong*$^{2-1}$ in (37), we see the clause *hua*$^{5-7}$*gi*^{2} *kong*$^{2-1}$ tiāndì de tiān *tioh*^{8} *bo*$^{5-0}$ 'in Mandarin we say *tiān* 'heaven' of *tiāndì* 'heaven and earth', right'. This clause can be interpreted as dependent clause of the *siunn*$^{7-3}$ 'think', giving 'I think that in Mandarin we say *tiān* 'heaven' of *tiāndì* 'heaven and earth', right'. Under this interpretation, *siunn*$^{7-3}$ is the matrix verb and *kong*$^{2-1}$ functions as the complementizer, explicitly linking the dependent clause *hua*$^{5-7}$*gi*^{2} *kong*$^{2-1}$ tiāndì de tiān *tioh*^{8} *bo*$^{5-0}$ with the matrix verb. The same analysis holds for the first sequence of *siunn*$^{7-3}$ *kong*$^{2-1}$ in (37). In this case *siunn*$^{7-3}$ *kong*$^{2-1}$ is followed by 1) the clitic =*a* 'EC' which indicates the expectance of the speaker that the preceding utterance will get a follow-up by either himself or one of his listeners, and 2) a pause. The clitic =*a* 'EC' can be used at any place where a speaker can pause. As noted above on p. 61, speakers have the option to pause between a matrix verb followed by *kong*$^{2-1}$ on the one hand and the corresponding dependent clause on the other hand. The fact that the clitic =*a* 'EC' with a pause is following the first instance of *siunn*$^{7-3}$ *kong*$^{2-1}$ is therefore no impediment to an analysis of *siunn*$^{7-3}$ as the matrix verb 'think' and *kong*$^{2-1}$ as a complementizer. To sum up, we have now established that the two instances of the phrase *siunn*$^{7-3}$ *kong*$^{2-1}$ in (37) are ambiguous. Both can be interpreted as an auxiliary verb followed by a transitive verb, i.e. *siunn*$^{7-3}$ *kong*$^{2-1}$ 'I want to say...'; and as a transitive verb followed by a complementizer, i.e. *siunn*$^{7-3}$ *kong*$^{2-1}$ 'I think that...'.

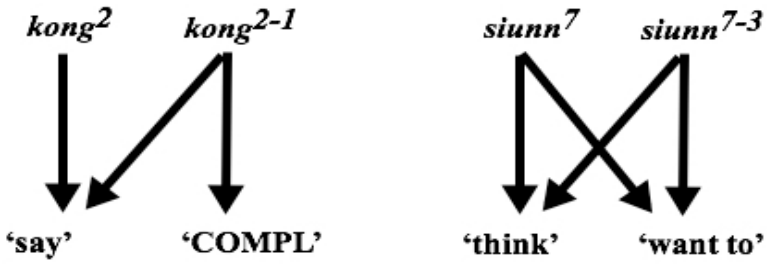
What is the nature of this ambiguity? The ambiguity arises from a different interpretation of the syntactic structure of this phrase and accordingly from a different interpretation of the morphemes involved. Are we dealing with the transitive verb *siunn*^{7} ~ *siunn*$^{7-3}$ 'think' or with the auxiliary *siunn*^{7} ~ *siunn*$^{7-3}$ 'want to'? And are we dealing with the transitive verb *kong*^{2} ~ *kong*$^{2-1}$ 'say' or with the complementizer *kong*$^{2-1}$? The form *siunn*$^{7-3}$ *kong*$^{2-1}$ can thus denote two different meanings.

The relation between the two meanings of *siunn*^{7} 'think' and 'want to' will be discussed in more detail below. The relation between *kong*^{2} ~ *kong*$^{2-1}$ 'say' and *kong*$^{2-1}$ 'COMPL' will be discussed under the heading "Grammaticalization". Linguists usually use the term *polysemy* to describe one linguistic form which denotes two or more related meanings. When one form denotes two unrelated meanings the term *homonymy* is used (Wiedenhof 1995: 17). An English example of homonymy can be found in the triplet *to, two* and *too*, all of which are semantically unrelated but have the same form /tu/, even though the spelling of these three words is different. An example of polysemy in English is the form *head*, which can refer to that part of a human or animal where the eyes, ears, nose and mouth are situated or to a person who is in charge of an organization or group of people. *Kong*$^{2-1}$ the verb and *kong*$^{2-1}$ the complementizer have related meanings; this will be discussed under "Grammaticalization". It thus seems we can treat this as a case of polysemy.

We must bare in mind, however, that whether the two meanings of one form are a case of polysemy or homonymy is something that may vary from speaker to speaker. To some speakers of English, for example, *head* 'person in charge of an organization' and *head* 'part of the body where the ears and eyes are situated' may semantically be completely unrelated and just happen to have the same form. For these speakers then, *head* 'person in charge of an organization' and *head* 'part of the body where the ears and eyes are situated' are homonymous. The same might be true for *kong*$^{2-1}$: for some speakers, the meanings may be related and thus polysemous, whereas for others the meanings may be completely unrelated and thus homonymous.

Note that as opposed to its sandhi form *kong*$^{2-1}$, the form *kong*^{2} is neither polysemous nor homonymous since it can only have the meaning 'say'. This is different for the form *siunn*^{7} ~ *siunn*$^{7-3}$, for which both the form with the original tone *siunn*^{7} and the form with the sandhi tone *siunn*$^{7-3}$ exhibit the same polysemy. Yuen Ren Chao calls relations such as the one between *kong*^{2} 'say' and *kong*$^{2-1}$ 'say; COMPL' "skewed relations" (1959: 11-12), which I have graphically presented in (38a). This is opposed to the symmetrical relation of *siunn*^{7} and *siunn*$^{7-3}$ 'think; want', cf. (38b).

(38) a. b.



The examples given in the preceding paragraph are all cases of polysemy dealing with lexical items – words, in other words. The definition of *polysemy* given above ran as follows: "one linguistic form which denotes two or more related meanings". This definition does not specify what kind of form we are dealing with, i.e. the form does not need to involve lexical items *per se*. In § 2.2.1, I argued that meaning can also be coded in the syntax. This is how I accounted for the difference in meaning between (26a) and (26b), repeated below:

(26) a. Peter loves Mary.
b. Mary loves Peter.

Both sentences are made up of the same lexical items, but their syntax is different. I will call two different yet related meanings that arise from two syntactic structures, but are presented by a single phonological form, syntactic polysemy. An example of syntactic polysemy in my native language, the Hollandish variety of Dutch, is given in (39).

(39) Péter schop Marie.
PNPeter kick PNMary

a) 'It is Peter who Mary kicks.'
b) 'It is Peter who kicks Mary.'

The accute accent <ˊ> on the first <e> in <Peter> represents stress. This sentence, including its prosody, can be used to describe two situations: a) the situation where Peter is the one kicking Mary, and b) the situation where Peter is the one kicked by Mary. The syntactic analysis of this single form is either ŚVO or (preposed) Ó + VS.[14] This is a case of syntactic polysemy because 1)

14 Here too, the acute accent is used to indicate stress.

there are two different yet related meanings, viz. ‘It is Peter who Mary kicks.’ and ‘It is Peter who kicks Mary.’; 2) these meanings arise from different syntactic structures, ÓVS and ŚVO respectively; and 3) they are presented by one phonological form, i.e. *Péter schop Marie.*.

We can now argue that *siunn*$^{7\text{-}3}$ *kong*$^{2\text{-}1}$ is a case of syntactic polysemy: the meaning ‘I want to say’ has the syntactic structure AUX + V, where AUX stands for auxiliary verb and V for verb; the meaning ‘I think that...’ has the syntactic structure V + C, where C stands for complementizer. Both these structures have the same phonological form, i.c. *siunn*$^{7\text{-}3}$*kong*$^{2\text{-}1}$.

On the other hand, seen from a lexical perspective, we could also argue that the polysemy of *siunn*$^{7\text{-}3}$ *kong*$^{2\text{-}1}$ is a case of lexical polysemy: the meaning ‘I want to say’ is denoted by the lexical items *siunn*$^{7\text{-}3}$ ‘want to’ and *kong*$^{2\text{-}1}$ ‘say’; whereas the meaning ‘I think that...’ is denoted by the lexical items *siunn*$^{7\text{-}3}$ ‘think’ and *kong*$^{2\text{-}1}$ ‘COMPL’.

I think that syntactic and lexical polysemy do not have to exclude each other, since they are different layers of analysis – nor do they presuppose each other. In the following pages, I will present examples for lexical polysemy without syntactic polysemy (40), syntactic polysemy without lexical polysemy (41) and lexical polysemy with syntactic polysemy (42).

The Dutch word *aarde* has two meanings which to me, being a native speaker of Dutch, are polysemous: ‘soil; Earth[15]’. The meaning ‘Earth’, here, can be analyzed as an extension of ‘soil’, ‘soil’ being the thing plants grow in and the thing we build houses on. By extension, in Dutch, one refers to the planet we live on as *de Aarde*, i.e. literally “the Soil”. If we use the word *aarde* to form sentence (40), the polysemy is lexical and not syntactic in nature.

(40) K=loop over de=aarde.
a) 1SG=walk over C.SG.DEF=soil
b) 1SG=walk over C.SG.DEF=Earth

a) ‘I’m walking across the soil.’
b) ‘I’m walking across the Earth.’

Both a) ‘I’m walking across the soil.’ and b) ‘I’m walking across the Earth.’ are possible meanings for *Kloop over deaarde.*. This means that, given the right context, this utterance can be interpreted in both ways. The only difference between the two, however, is the location, i.e. the place across which the speaker of (40) is walking to – either the soil or the Earth. The syntactic structure of (40) for either meaning is the same.

15 Note that in written Dutch, the form meaning ‘Earth’ is written with a capital letter: <Aarde>.

In (40), we saw a case of lexical polysemy without syntactic polysemy. Wiedenhof (1995: 31-32) gives the following Mandarin example, which he cites from Zhū (1985: 170, transcription, glosses and translations are Wiedenhof's, based on the meanings of (41) described by Zhū):

(41) 反对 的 是 少数 人.
fǎnduì de shi shǎoshù rén
oppose SUB be minority person

a) 'Those opposing it are in a minority.'
b) 'Those opposed to are in a minority.'

According to Wiedenhof, the two meanings of (41) "derive from the syntactic association of *fǎnduì* 'oppose' with *de* 'SUB'." (p. 32). No part of the phonological form of the utterance is interpreted as a different lexical item in meaning b) and in meaning in a). We can therefore conclude that, in this case, polysemy arises only from the syntactic association of *fǎnduì* and *de*, and not from different lexical items.

We have now seen examples of lexical polysemy without syntactic polysemy (40) and *vice versa* (41). Now we will look at an utterance in which both lexical polysemy and syntactic polysemy are present. An example presents itself in the name of a shop dealing in wool in the British town of Cambridge: The Sheep Shop. Let us take the name of this shop not as a name, but as a regular utterance as presented in (42) (the acute accent has been added to indicate stress):

(42) No, the shéep shop.

Two meanings can correspond to this phonological form. The first meaning denotes something like 'On the contrary, a place identifiable from context where one can buy things and which is somehow characterized by a certain kind of woolly mammal (*ovis aries*) as opposed to being characterized by something else'. This meaning can be used in reaction to someone asking whether the shop on the corner was called "the goat shop". *No, the shéep shop.* The second meaning denotes: 'a multitude of woolly mammals (*ovis aries*) identifiable from context that are the ones that go out to buy things on a regular basis'. Analyzing this utterance from a lexical perspective, we see that the first two lexical items, *the* and *sheep*, are the same for both meanings corresponding with the utterance in (42); it is the final part of this utterance that causes the different meanings. *Shop* can both be a noun meaning 'place where one can buy things' and a verb meaning 'buy things'. Analyzing this utterance from a syntactic perspective, we see that for the first meaning, we are dealing with the noun phrase *the sheep shop* of which *shop* is the head – it is a type of shop. Both the definite article *the* and the noun

sheep modify the noun *shop*, resulting in the structure Modifier + Head. In the second meaning, we are dealing with the nominal phrase *the sheep* and the verb *shop* which are construed to form the construction SV in which *the sheep* are presented as the ones carrying out the action denoted by the verb *shop*.

Note that the acute accent indicating stress on the word *sheep* is crucial for the same form and thus for the establishment of polysemy. In English, compound nouns receive lexical stress on their first element, viz. *sheep*, but in this case, clausal stress is also placed on *sheep* so as to arrive at the meaning 'No, it is the shéep shop, not the goat shop.' In an SV-construction, the S does not receive clausal stress unless the construction is uttered in a situation which contrasts the S to another possible S. In this case, one could for example paraphrase this meaning as 'It is the sheep who shop (and not any other animals).'.

In the utterance *The shéep shop.*, one can only separate lexical polysemy from syntactic polysemy in one's analysis. In this linguistic form itself, lexical polysemy cannot exist without syntactic polysemy – they are two sides of the same coin. This does not mean that it is needless to distinguish between the two concepts of lexical and syntactic polysemy. Examples (26) and (39) to (41) discussed above showed that the two can indeed occur separately. In this case, however, they cannot. The same goes for our instances of *siunn*$^{7-3}$ *kong*$^{2-1}$ in (37), repeated here:

(37) A^{0} soo$^{2-1}$i$^{2-1}$ **siunn$^{7-3}$ kong$^{2-1}$** =*a*, u$^{7-3}$ e$^{5-7}$ khah$^{4-1}$ kan$^{2-1}$tan^{1} e$^{5-7}$ tsit$^{4-8}$ e$^{5-7}$
RLT therefore **want.to say** =EC EXST SUB relatively simple SUB this entity

han$^{3-2}$ji$^{7-3}$ =*a*, **siunn$^{7-3}$ kong$^{2-1}$,** Hua$^{5-7}$gi^{2} kong$^{2-1}$ *tiāndì* =*de tiān*
Chinese.character =EC **want.to say** Mandarin say heaven.and.earth =SUB heaven

tioh^{8} <hmm> <hann^{0}> bo$^{5-0}$ honn^{0}, a^{0} lan$^{2-1}$ Tai$^{5-7}$gi^{2} ma$^{7-3}$ kong$^{2-1}$ <thinn^{1}>
right BKCH CFM NEG.EXST XPC RLT 1INCL Taiwanese.language also say heaven

thinn$^{1-7}$ kah$^{4-1}$ thian^{1} =ma^{0} <henn^{0}> <tioh^{8}> honn^{0}?
heaven and heaven =DASS CFM right XPC

'So I want to say, some of the simpler characters, I want to say, in Mandarin we say the *tiān* 'heaven' of *tiāndì* 'heaven and earth', <hmm> right <yes>, in Taiwanese we also say <*thinn*> *thinn* 'heaven' and *thian* 'heaven', you know <yes> <right> right?' (45D; 15:11)

The meaning 'I want to say' has a syntactic structure of AUX + V, whereas the meaning 'I think that...' has the syntactic structure of V + C. In this case, however, different syntactic structures presuppose different lexical items. The structure AUX + V presupposes the lexical items *siunn*$^{7-3}$ 'want to' and *kong*$^{2-1}$ 'say', whereas the structure V + C presupposes the lexical items *siunn*$^{7-3}$ 'think' and *kong*$^{2-1}$ 'COMPL'. We are thus dealing with lexical and with syntactic polysemy at the

same time. Whether this is a case of polysemy or homonymy is bound to vary from speaker to speaker. Before concluding this part about the syntactic ambiguity of *siunn*$^{7\text{-}3}$ *kong*$^{2\text{-}1}$, I want to add that it is possible and indeed plausible to treat *siunn*^{7} ~ *siunn*$^{7\text{-}3}$ 'think; want to' as a case of polysemy for two reasons. The first is that in some other Sinitic languages the verb 'think' has the same form as the auxiliary 'want to'. Examples are Moiyen Hakka *xiong*^{3} 'think; want to', Teo-swa Mĭn *xiõ*^{6} 'think; want to' (Ōuyáng et al. 2005: 338) and Mandarin *xiǎng* 'think; want to' (Zhāng 2009: 505). There are other Sinitic languages which employ different lexical items for 'think' and 'want to', e.g. Cantonese *lám* 'think' (Matthews & Yip 1994: 7) and *séung* 'want to' (*ibid.*: 235-236). The second reason is that for Mandarin, the diachronic development of the verb 想 *xiǎng* 'think' into the auxiliary *xiǎng* 'want to' can be attested in written documents from the Yuán 元 dynasty (1271-1368) onwards (Peyraube & Li 2012: 162-163). The meaning of *xiǎng* changed from 'think of' > 'think of (doing something)' > 'want to'. A similar development may be assumed for *xiong*^{3} in Moiyen, *xiõ*^{6} in Teo-swa and *siunn*^{7} in TSM. Looking at these arguments, I think it is safe to say that it is indeed a possibility to treat *siunn*^{7} ~ *siunn*$^{7\text{-}3}$ 'think; want to' as a case of polysemy.

2.3.2 Si$^{7\text{-}3}$ *kong*$^{2\text{-}1}$

In (43), we see the phrase *toh*$^{8\text{-}3}$ *si*$^{7\text{-}3}$ *kong*$^{2\text{-}1}$. *Toh*^{8} is an adverb meaning 'just' and *si*^{7} is a copula 'be'.

(43) A^{0} gua$^{2\text{-}1}$ koh$^{4\text{-}2}$jin^{5} e$^{5\text{-}7}$ kuan$^{1\text{-}7}$he^{7}-- **toh$^{8\text{-}3}$ si$^{7\text{-}3}$ kong$^{2\text{-}1}$** gua$^{2\text{-}1}$ tsin$^{1\text{-}7}$-- tsin$^{1\text{-}7}$
RLT 1SG personal SUB relation **just be say** 1SG real real

huann$^{1\text{-}7}$hi^{2} i$^{1\text{-}7}$ tik$^{8\text{-}4}$piat$^{8\text{-}4}$ koh$^{4\text{-}8}$ si$^{7\text{-}3}$ gua$^{2\text{-}1}$ hit$^{4\text{-}8}$ e$^{5\text{-}7}$ gian$^{2\text{-}1}$kiu$^{3\text{-}2}$soo$^{2\text{-}1}$ kai$^{1\text{-}7}$tuann^{7} e$^{5\text{-}7}$
happy 3 special again be 1SG that entity graduate.school phase SUB

tsi$^{2\text{-}1}$to$^{7\text{-}3}$ kau$^{3\text{-}2}$siu^{7}.
supervise professor

'And my own tie with him-- I mean, I am very-- very happy he is especially also my supervising professor during my time at graduate school.' (10B; 07:57)

The context of (43) is that speaker B is asked to introduce speaker C. This is something speaker B is happy to do because speaker C used to be his supervising professor when speaker B was at graduate school. First he says *a*^{0} *gua*$^{2\text{-}1}$ *koh*$^{4\text{-}2}$*jin*^{5} *e*$^{5\text{-}7}$ *kuan*$^{1\text{-}7}$*he*^{7}-- 'and my own tie [with him]--', but having said this, he is not so sure how he should continue and he uses the phrase *toh*$^{8\text{-}3}$ *si*$^{7\text{-}3}$ *kong*$^{2\text{-}1}$ to indicate he is rephrasing.

(44) A^{0} tshenn$^{1-7}$me$^{5-7}$ gu^{5}, lan$^{2-1}$ Tai$^{5-7}$uan$^{5-7}$ lang^{5} le^{1} kong$^{2-1}$ tshenn$^{1-7}$me$^{5-7}$ gu^{5}
RLT blind buffalo 1INCL Taiwan person PROG say blind buffalo

toh$^{8-3}$ si$^{7-3}$ kong$^{2-1}$, ah-- a^{0} gua$^{2-1}$ toh$^{8-3}$ e$^{7-3}$hiau$^{2-1}$ kong$^{2-1}$ Tai$^{5-7}$gi^{2} honn^{0} <si^{7}>,
just **be** **say** HST RLT 1SG just know.how.to say Taiwanese XPC be

a^{0} m^{7-}koh$^{4-1}$ =he^{0}, ji^{7} honn^{0}, m$^{7-3}$ bat^{4} gua$^{2-0}$ la^{0} <hann0> honn^{0} <tioh^{8}>.
RLT but =RLV character XPC NEG familiar.with 1SG WARN CFM XPC correct

'And *tshenn-mê gû* 'blind buffalo', if we Taiwanese are saying *tshenn-mê gû*, then this means uhm-- "I can speak Taiwanese as you know <yes>, but, the characters, you know, they are not familiar with me" <hm>, right <right>.' (20C; 10:38)

In (44), we see another example of *toh$^{8-3}$ si$^{7-3}$ kong$^{2-1}$* – here, in contrast to (43), the speaker pauses for a moment after *toh$^{8-3}$ si$^{7-3}$ kong$^{2-1}$*, before continuing talking. *Toh$^{8-3}$ si$^{7-3}$ kong$^{2-1}$* is used here to provide additional information, in this case to explain what Taiwanese people mean when they use the phrase *tshenn$^{1-7}$me$^{5-7}$ gu^{5}* 'blind buffalo': it means one can speak Taiwanese, but cannot read or write it. In English, *toh$^{8-3}$ si$^{7-3}$ kong$^{2-1}$* has been translated with 'then this means', *it* referring to the phrase *tshenn$^{1-7}$me$^{5-7}$ gu^{5}*. There are however, other interpretations possible as well: *toh$^{8-3}$ si$^{7-3}$ kong$^{2-1}$* in (44), could also refer to the Taiwanese, i.e. 'if we Taiwanese are saying *tshenn-mê gû*, then we mean'. A third possibility is that *toh$^{8-3}$ si$^{7-3}$ kong$^{2-1}$* refers to speaker C, i.e. 'I mean', just as in (43). The difference between (44) and (43), however, is that in (43), the only possibility is 'I mean'. Compared with *toh$^{8-3}$ si$^{7-3}$ kong$^{2-1}$* in (43) then, *toh$^{8-3}$ si$^{7-3}$ kong$^{2-1}$* in (44) is only half-way in the process of grammaticalization. For more on grammaticalization, see below (§ 2.3.5).

(45) **Si$^{7-3}$ kong$^{2-1}$** tsia^{1} e$^{5-7}$ ji^{7} toh$^{8-3}$ si$^{7-3}$ ti$^{7-3}$ =le^{1} hia^{1}, toh$^{8-3}$ tshiunn$^{7-3}$
be **say** here SUB character just be be.at =DUR there just resemble

tu$^{2-1}$tsiah$^{4-2}$ lau$^{7-3}$su^{1} kong^{2} e$^{5-0}$ i$^{1-7}$ e$^{5-7}$ lik$^{8-4}$su^{2} tsiok$^{4-8}$ tng^{5} e$^{5-0}$ <tioh^{8}> <hm> honn^{0}.
just.now teacher say SUB 3 SUB history very long SUB right BKCH XPC

'I mean, these characters are there, just like Mr. Lí said just know, their history is very long <yes> <hm> right.' (49D; 15:49)

Example (45) shows an instance of *si$^{7-3}$ kong$^{2-1}$*. In this case, *toh$^{8-3}$* is not present. The function of *si$^{7-3}$ kong$^{2-1}$* in (45), however, is similar to that in (43) and (44). In the context preceding the utterance in (45), speaker D was talking about the fact that Chinese characters do not especially belong to Mandarin. In (45), she tries to elaborate on that by saying that they are just there and that they go back a long time. After the utterance in (45), she pursues the same topic, saying that every Sinitic language has its own tradition of reading characters out loud. *Si$^{7-3}$ kong$^{2-1}$* is thus used to

specify or elaborate on what speaker D said in the first part of her narrative. The fact that the usage of *si*$^{7-3}$ *kong*$^{2-1}$ in (45) is similar to the usage of *toh*$^{8-3}$ *si*$^{7-3}$ *kong*$^{2-1}$ in (43) and (44), indicates that *toh*$^{8-3}$ is not part of fixed phrase. The difference between *toh*$^{8-3}$ *si*$^{7-3}$ *kong*$^{2-1}$ and *si*$^{7-3}$ *kong*$^{2-1}$ is thus one between 'that is just to say' and 'that is to say'.

In all three examples, *si*$^{7-3}$ *kong*$^{2-1}$ has a fixed form: both syllables exhibit tone sandhi. From a comparison of (43) with (44), it is clear that presence or absence of a pause following *si*$^{7-3}$ *kong*$^{2-1}$ does not influence the tonal contour of the syllable *kong*$^{2-1}$: the tonal contour is always that of a high level tone. This is an indication for the degree of lexicalization of the phrase *si*$^{7-3}$ *kong*$^{2-1}$.

Semantically, *si*$^{7-3}$ *kong*$^{2-1}$ is used when the speaker is rephrasing or elaborating on what he was saying. It seems to convey the meaning 'mean' in English, the appropriate subject to this verb being interpretable from the context. For example, in (43) and (45) it is 'I', whereas in (44), there are multiple interpretations possible: e.g. 'this', 'we Taiwanese' and 'I'. The meaning probably evolved from 'X BE saying' (44) > 'it is that X BE saying' > 'that is to say' > '[I] mean' (43) and (45). Here, X represents any subject and BE represents any form of the verb *be* congruent with X. The lexicalization of forms of verbs meaning 'be' and 'say' can also be seen in other languages. Examples are Spanish *es decir*, litterally 'it is to say', French *c'est à dire* 'that is to say' and the English form *that is to say* itself. All three of these are used to specify or rephrase something which has just been said, which is exactly what *si*$^{7-3}$ *kong*$^{2-1}$ seems to be used for in (43) to (45).

In my corpus, I found 9 clear cases of the phrase *si*$^{7-3}$ *kong*$^{2-1}$. For other examples, see 24C, 27C, 33C, 41D, 43D and 54B of the Appendix. In addition, I found the rather special form *ah*$^{8-3}$*si*$^{7-3}$ *kong*$^{2-1}$, shown in (46).

(46)

A^{0}	lan$^{2-1}$	Tai$^{5-7}$gi^{2}	ham$^{7-3}$	tsit$^{4-8}$ma^{2}	te$^{5-7}$	tsit$^{4-8}$	e$^{5-7}$	Hua$^{5-7}$gi^{2}	**ah**$^{8-3}$	**si**$^{7-3}$
RLT	1INCL	Taiwanese.language	with	now	SUB[16]	this	entity	Mandarin	**or**	**be**

kong$^{2-1}$	Kheh$^{4-2}$gi^{2}	long$^{2-1}$	si$^{7-3}$	tsia^{1}	e$^{5-7}$	gi$^{2-1}$gian$^{5-7}$	lai$^{7-3}$te^{2}	e$^{5-7}$	ki$^{5-7}$tiong^{1}	e$^{5-7}$
say	Hakka.language	in.every.case	be	here	SUB	language	inside	SUB	among	SUB

tsia^{1}	e$^{5-7}$	gi$^{2-1}$gian$^{5-7}$	lai$^{7-3}$te^{2}	e$^{5-7}$	ki$^{5-7}$tiong^{1}	e$^{5-7}$	kui$^{2-1}$	tsiong^{2}.
here	SUB	language	inside	SUB	among	SUB	how.many	kind.of

'And our Taiwanese together with this Mandarin or Hakka, for example, are all a couple of different varieties among those within this language.' (44D; 14:56)

Ah$^{8-3}$ means 'or' in TSM. In (46), as is often the case, *ah*$^{8-3}$ is followed by *si*^{7} 'be' (Lin 2015: 429-431), a comparable case is *hái ~ hár ~ háishi ~ háishì* 'still; or' in Mandarin (Wiedenhof 2015:

16 My informants think that this is a speech error, suggesting that the Taiwanese word *e*^{5} 'SUB' was influenced by the Mandarin clitic *=de* 'SUB' which has the same function.

311-312). *Ah*$^{8-3}$*si*$^{7-3}$ 'or' is folowed here by *kong*$^{2-1}$. *Ah*$^{8-3}$*si*$^{7-3}$ 'or' is used to expand with an extra option (i.e. a third language); *kong*$^{2-1}$ seems to be used to convey that this extra option is possibly *Kheh*$^{4-2}$*gi*^{2} 'Hakka'. Although it occurs just once in my corpus, I know from experience that the collocation *ah*$^{8-3}$*si*$^{7-3}$ *kong*$^{2-1}$ occurs quite often in TSM. What is interesting is that *ah*$^{8-3}$*si*^{7} consists of *ah*^{8} 'or' with the copula *si*^{7} 'be'. *Ah*^{8} has a strong affiliation with *si*^{7}. It may be that *ah*$^{8-3}$*si*$^{7-3}$ *kong*$^{2-1}$ originally was a collocation of *ah*$^{8-3}$*si*$^{7-3}$ and *si*$^{7-3}$ *kong*$^{2-1}$, the elaborating function of *si*$^{7-3}$ *kong*$^{2-1}$ matching well with the expanding function of *ah*$^{8-3}$*si*$^{7-3}$ 'or'. One of the two instances of *si*$^{7-3}$ from the resulting form *ah*$^{8-3}$*si*$^{7-3}$ *si*$^{7-3}$ *kong*$^{2-1}$ 'or, for example' could then have been lost resulting in a case of haplology. Another possibility is that the *si*$^{7-3}$ in *ah*$^{8-3}$*si*$^{7-3}$ is still felt to be the copula *si*$^{7-3}$ to such a degree that it is unnecessary to add an extra *si*^{7} and only the addition of *kong*$^{2-1}$ suffices.

Because I have only one example of *ah*$^{8-3}$*si*$^{7-3}$ *kong*$^{2-1}$ 'or, for example' in my corpus, I cannot draw any hard an fast conclusions, but in this case as well, *si*$^{7-3}$ *kong*$^{2-1}$ seems to be used to elaborate on something, just as we saw in (43) to (45).

(47)

A^{0}--	a^{0}	soo$^{2-1}$i$^{2-1}$	tsit$^{4-8}$	e$^{5-7}$	tshenn$^{1-7}$me$^{5-7}$	gu^{5},	ki$^{5-7}$-sit^{8}	ma$^{7-3}$	**bo$^{5-7}$**	**it$^{4-8}$ting$^{7-3}$**
RLT	RLT	therefore	this	entity	blind	buffalo	actually	also	NEG.EXST	**definitely**

kong$^{2-1}$	u$^{7-3}$	sim$^{2-1}$mih$^{8-3}$	phainn$^{2-1}$	i$^{3-2}$su^{3}	a^{0}	<si^{7}>.
say	EXST	what	bad	meaning	PERF	be

'And-- and so this *tshenn-mê gû*, actually, you know, it isn't necessarily the case that it has a bad meaning <yes>.' (26C; 11:25)

In (47), we see the collocation of *bo*$^{5-7}$ *it*$^{4-2}$*ting*$^{7-3}$ 'not necessarily' with *kong*$^{2-1}$. In the context preceding the utterance in (47), speaker C has been explaining that the Taiwanese cultural society of his parents' day, was one of agriculture where it was common to be illiterate. He also stated that people use the word *gu*^{5} 'buffalo' in all kinds of comparisons. He then makes the utterance transcribed in (47), concluding that the phrase *tshenn*$^{1-7}$*me*$^{5-7}$ *gu*^{5} 'blind buffalo' does not have negative connotations per se. *Kong*$^{2-1}$ in *bo*$^{5-7}$ *it*$^{4-8}$*ting*$^{7-3}$ *kong*$^{2-1}$ 'it is not necessarily the case that' seems to convey a similar meaning to *si*$^{7-3}$*kong*$^{2-1}$, i.e. in both collocations a meaning or intention is involved. The meaning may have developed as follows: 'X BE not necessarily saying' > 'it is not necessarily the case that X BE saying' > 'it is not necessarily the case that X means' > 'X not necessarily means'. In the context of the utterance in (47), which is about the supposed meaning of the phrase *tshenn*$^{1-7}$*me*$^{5-7}$ *gu*^{5}, the above described meaning of *kong*$^{2-1}$ added to *bo*$^{5-7}$ *it*$^{4-8}$*ting*$^{7-3}$ 'not necessarily' fits in nicely.

As was the case with *ah*$^{8-3}$*si*$^{7-3}$ *kong*$^{2-1}$ in (46), I have only one example of *bo*$^{5-7}$ *it*$^{4-8}$*ting*$^{7-3}$ *kong*$^{2-1}$ 'it isn't the case that' and the conclusions I draw here are therefore only tentative.

2.3.3 Soo$^{2-1}$i$^{2-1}$ kong$^{2-1}$

Apart from the construction *hi$^{1-7}$bang$^{7-3}$ kong$^{2-1}$* discussed in § 2.2 in which *kong$^{2-1}$* functions as a complementizer between *hi$^{1-7}$bang$^{7-3}$* 'hope' and the following dependent clause, we see the collocation *soo$^{2-1}$i$^{2-1}$ kong$^{2-1}$*.

(22) A^{0} **soo$^{2-1}$i$^{2-1}$ kong$^{2-1}$,** ti$^{7-3}$ tsia^{1} *=ne*, *tòuguò =ne*, Soo$^{1-7}$ Se$^{3-2}$hiong^{5} *lăosī* hoo^{0}, ua$^{2-1}$ ma$^{7-3}$
RLT **therefore say** be.at here =RLV via =RLV PNSoo PNSè-hiông teacher XPC 1SG also

hi$^{1-7}$bang$^{7-3}$ kong$^{2-1}$ i$^{1-7}$ e$^{7-3}$tang$^{3-2}$ sio$^{2-1}$ kan^{9} kai$^{2-1}$sueh^{4} tsit$^{8-0}$e$^{7-0}$.
hope say 3 be.able.to a.little OBJ:1INCL[17] explain a.while

'So, here, through, you know, Mr. Soo Sè-hiông, I also hope that he can explain it a bit to us.' (4A; 06:07)

Preceding the utterance shown in (22), speaker A, the host of the show, was introducing herself and her co-host, speaker B. While trying to thank speaker B for being the co-host, she has trouble getting her words out. In particular, when she wants to say the Taiwanese word for 'host', as in to host a show, *tsu$^{2-1}$tshi^{5}*, she gets stuck. She then quickly uses the Taiwan Mandarin word *zǔcí* 'host' which did spring to mind. Following this, she utters the phrase *a^{0} soo$^{2-1}$i$^{2-1}$ kong$^{2-1}$* and then goes on talking about what she wants him to do, i.e. explain a bit to the audience what it was they discussed in the last episode of the show. The trouble speaker A has with getting out the proper Taiwanese word for 'host', interrupts her narrative. It seems that she uses *soo$^{2-1}$i$^{2-1}$ kong$^{2-1}$* to get back on track after the interruption. I will therefore refer to this use of a linguistic form as a back-on-tracker.

(48) A^{0} **soo$^{2-1}$i$^{2-1}$ kong$^{2-1}$** ji^{7}, toh$^{8-3}$ si$^{7-3}$ ai$^{3-2}$ sing$^{1-7}$ kong$^{2-1}$ sim$^{2-1}$mih^{8} kio$^{3-2}$ tso$^{3-2}$
RLT **therefore say** character just be have.to first say what call do

ji^{7} la^{0}.
character WARN

'So if we talk about *jī* 'characters', we have to first say what we call *jī*, you know.' (37C; 12:35)

In (48), we see another instance of *soo$^{2-1}$i$^{2-1}$ kong$^{2-1}$*. Preceding uttering (48), speaker C was explaining that it was really common to be illiterate in rural Taiwan in his parents' generation and that the phrase *tshenn$^{1-7}$me$^{5-7}$ gu^{5}*, which means 'blind buffalo' but is used to refer to an illiterate person, was not considered a derogative term since it was so common to be illiterate and people

17 The word *kan^{9}* is a fusion of *ka^{7}* 'OBJ' and *lan^{2}* '1INCL'.

used the phrase to describe themselves. Speaker A, then, jokes that the buffalo must have thought something like "It is not my fault at all. How would they know I am illiterate.". They all have a laugh. This joke interrupts the narrative of speaker C, who after having laughed, says *a*^{0} *soo*$^{2-1}$*i*$^{2-1}$ *kong*$^{2-1}$ *ji*^{7} and then continues his narrative. The use of *soo*$^{2-1}$*i*$^{2-1}$ *kong*$^{2-1}$ in (22) and (48) seems to be similar: in both cases the speaker uses the back-on-tracker *soo*$^{2-1}$*i*$^{2-1}$ *kong*$^{2-1}$ to return to the topic they were discussing before the interruption. There is a difference though: in (22), *soo*$^{2-1}$*i*$^{2-1}$ *kong*$^{2-1}$ is followed by a pause, whereas in (48), it is followed by the word *ji*^{7}. There, speaker C uses the word *ji*^{7} 'characters' to introduce the topic of the next part of his narrative – which is about what *ji*^{7} actually are. Following this topic, there is another pause – notice in this respect that *ji*^{7} preserves its original tone and does not exhibit tone sandhi.

In both (22) and (48), *kong*$^{2-1}$ is attested with its high level sandhi tone. In (48), *kong*$^{2-1}$ is followed by the word *ji*^{7} which grammatically functions as the object of *kong*$^{2-1}$. This gives rise to tone sandhi of *ji*^{7}. In (22), however, *kong*$^{2-1}$ exhibits tone sandhi even though it is followed by a pause. This might be an indication of the degree of lexicalization of the phrase *soo*$^{2-1}$*i*$^{2-1}$ *kong*$^{2-1}$. If we take a closer look at the sentence in (48), however, we see that the remaining part of the sentence starts with the word *toh*$^{8-3}$ 'just, then'. This word is used at the start of a clause following another clause and denotes that the proposition described in the second clause is relevant when the proposition described in the first clause is the case. In other words, *toh*$^{8-3}$ directs the hearer to interpret the first clause as a temporal background to the second clause: 'So when we talk about *ji*^{7} 'characters', we have to first say what we call *ji*^{7}.' This can also be seen from the English translation which uses *if* in the first clause to arrive at the same end. It seems, then, that *soo*$^{2-1}$*i*$^{2-1}$ *kong*$^{2-1}$ in this case should not be analyzed as a lexicalized phrase, but just as the co-occurrence of *soo*$^{2-1}$*i*^{2} 'so, therefore' with *kong*^{2} 'say'.

At a closer look, the fact that *kong*$^{2-1}$ 'say' appears after *soo*$^{2-1}$*i*$^{2-1}$ 'so' in both (22) and (48) does not appear to be the same. Nontheless, these two examples might give us some insight in the lexicalization process of *soo*$^{2-1}$*i*$^{2-1}$ *kong*$^{2-1}$, represented schematically in (49).

(48) *soo*$^{2-1}$*i*$^{2-1}$ *kong*$^{2-1}$ + O > *soo*$^{2-1}$*i*$^{2-1}$ *kong*$^{2-1}$

It started out as a simple collocation of *soo*$^{2-1}$*i*^{2} and *kong*$^{2-1}$ + O, where O is the grammatical object o f *kong*$^{2-1}$, denoting its patient, i.e. that which is talked about. This phrase *soo*$^{2-1}$*i*$^{2-1}$ *kong*$^{2-1}$ eventually came to be used without the following O, the high level tone of *kong*$^{2-1}$ remaining as a modern trace of the fact that historically, there was an object following *kong*$^{2-1}$. This remaining high level sandhi tone is something we saw in the discussion of *si*$^{7-3}$ *kong*$^{2-1}$ as well. There it was noted

that such a fossilized sandhi tone is an indication of the degree of lexicalization. The same goes for the fossilized sandhi tone on *kong*$^{2-1}$ in *soo*$^{2-1}$*i*$^{2-1}$ *kong*$^{2-1}$ 'so'.

Since I only have two examples of *soo*$^{2-1}$*i*$^{2-1}$ *kong*$^{2-1}$, the above offered analysis can only be tentative. In order to reach a conclusion on the use and lexicalization of *soo*$^{2-1}$*i*$^{2-1}$ *kong*$^{2-1}$, more research with larger corpora is needed.

2.3.4 Tshiunn$^{7-3}$ *kong*$^{2-1}$*:*

Finally, I will return to the use of *kong*$^{2-1}$ is its collocation with *tshiunn*^{7}, forming *tshiunn*$^{7-3}$*kong*$^{2-1}$. I found this collocation twice. The examples can be seen below.

(50) A^{0} **tshiunn$^{7-3}$** **kong$^{2-1}$** Hua$^{5-7}$gi^{2} kong$^{2-1}$ *dì* *shàng* *=de* *dì* <hm> honn^{o}.
RLT **resemble** **say** Mandarin say earth above =SUB earth BKCH XPC

'Just like in Mandarin we say *dì* 'acreage' as in *dì shàng* 'on the acreage', <hm> right.' (47D; 15:24)

(35) **Tshiunn$^{7-3}$** **kong$^{2-1}$** tsit$^{4-8}$ e$^{5-7}$ thinn^{1} *=a* te^{7} la^{0} suann^{1} la tsui^{2}
resemble **say** this entity heaven =EC earth WARN mountain WARN water

la^{0} honn^{0}, tsia^{1} e$^{5-7}$ han$^{3-2}$ji^{7} long$^{2-1}$ si$^{7-3}$ Tai$^{5-7}$gi^{2} ham$^{7-3}$ Hua$^{5-7}$gi^{2}
WARN XCP here SUB Chinese.character in.every.case be Taiwanese with Mandarin

i-n$^{1-7}$ iong^{7} e$^{5-7}$ ji^{7}, long$^{2-1}$ si$^{7-3}$ kang$^{7-3}$ khuan^{2} e^{0}.
3-PL use SUB character in.every.case be same type SUB

'Just like these *thinn* 'heaven', *tē* 'earth', *suann* 'mountain', *tsuí* 'water', you know, their Chinese characters are all characters used by Taiwanese and Mandarin, they're all the same.' (48D; 15:28)

The sentences in (50) and (35), which follows four seconds later, are part of an explanation by speaker D as to why people who think that Chinese characters belong with Mandarin only, according to her, are mistaken. In her explanation, she gives a couple of examples of Mandarin words that are written with certain characters and which have cognates in Taiwanese which are written with the same characters. The utterance in (50) is the first part of one of these examples. She starts with *tshiunn*$^{7-3}$ *kong*$^{2-1}$ and then follows with the Mandarin word *dì* 'earth', which, as is customary, she describes using a short phrase so as to ensure everyone knows which character with the pronunciation *dì* she is talking about. The phrase she uses to describe *dì* is *dì shàng=de dì* 'the

dì of *dì shàng* ‘on the ground’’. After having established everyone knows which character she is talking about, she gives the Taiwanese pronunciation, i.c. *te*^{7} ‘earth’.

She then proceeds with the utterance in (35). Here, something similar happens. At the end of her argument, the speaker sums up a couple of examples she has already mentioned, along with new examples of TSM words whose characters are also used to write Mandarin cognates. This summing-up is preceded by the phrase *tshiunn*$^{7\text{-}3}$ *kong*$^{2\text{-}1}$. The two morphemes present in *tshiunn*$^{7\text{-}3}$ *kong*$^{2\text{-}1}$ are *tshiunn*^{7} ‘resemble’ and *kong*^{2} ‘say’. The combination of the two would mean something like ‘just like saying’. From the fact that in the sentences in (50) and (35) *kong*$^{2\text{-}1}$ in the phrase *tshiunn*$^{7\text{-}3}$ *kong*$^{2\text{-}1}$ seemingly does convey the meaning ‘say’, it could be inferred that the phrase *tshiunn*$^{7\text{-}3}$ *kong*$^{2\text{-}1}$, just as with *soo*$^{2\text{-}1}$*i*$^{2\text{-}1}$ *kong*$^{2\text{-}1}$ and *si*$^{7\text{-}3}$ *kong*$^{2\text{-}1}$ above, has undergone some degree of lexicalization. However, in contrast to *soo*$^{2\text{-}1}$*i*$^{2\text{-}1}$ *kong*$^{2\text{-}1}$ and *si*$^{7\text{-}3}$ *kong*$^{2\text{-}1}$, the fact that here too, *kong*$^{2\text{-}1}$ is pronounced in a high level tone cannot be used as an argument for its supposed lexicalization. This is because in both (50) and (35), *tshiunn*$^{7\text{-}3}$ *kong*$^{2\text{-}1}$ is not followed by a pause.

In order to understand the function and level of lexicalization of the phrase *tshiunn*$^{7\text{-}3}$ *kong*$^{2\text{-}1}$ better, more research with larger corpora is needed, so as to ensure more examples are found.

2.3.5 Grammaticalization & lexicalization

Although throughout this study, my aim has been to give a synchronic description of the use of the forms *kong*^{2} ~ *kong*$^{2\text{-}1}$ in present-day TSM, I have occasionally made some remarks about the possible origin of certain forms. To wrap up the analytic part of this study, let me discuss some aspects of the grammaticalization and lexicalization of *kong*^{2} ~ *kong*$^{2\text{-}1}$. I will present definitions for some key concepts and summarize what I have noticed on the basis of the materials collected for this study.

Grammaticalization has been defined in a variety of ways and there is quite some debate about the details of what it is supposed to entail. Most definitions involve the process of how lexical items attain grammatical meaning (Brinton & Traugott 2005: 23). I take a lexical item to be a linguistic form that is paired with a fixed meaning and as such is stored in the brain by way of rote memorization. All lexical items in the brain are together referred to as the lexicon. In Matthews’ *Dictionary of linguistics* (2007: 164), grammatical meaning is defined as “any aspect of meaning described as part of the syntax or morphology of a language as distinct from its lexicon”. A classical example of grammaticalization in English is the process of how *going to*, expressing movement to a place, has come to be used as a future marker. It grammaticalized along this pathway: ‘going to [a place]’ > ‘going away [to do sth.]’ > ‘[do sth.] in the immediate future’. The form of *going to* has

changed as well. Where the meaning ‘going to [a place]’ only allows the form *going to* [ˈgoɪŋtʊ], as a future marker the form can range from *going to* [ˈgoɪŋtʊ] to *gonna* [ˈgɒnɐ], and all the different forms in between.

According to Hopper & Traugott (2003: 106), there is a tendency for grammaticalization to start with prototypical members of categories such as nouns and verbs. Subsequently, these become less prototypical members of their categories. We can thus say that *kong*^{2} probably started out as a transitive verb meaning ‘say’. This use was discussed in § 2.1; an example of this was shown in (9):

(9) A^{0} tshenn$^{1-7}$me$^{5-7}$ gu^{5} m$^{7-3}$ bat^{4}, e$^{7-3}$hiau$^{2-1}$ **kong^{2}**, be$^{7-3}$hiau$^{2-1}$ sia^{2}.
RLT blind buffalo NEG familiar.with know.how.to **say** NEG:know.how.to write

‘But an illiterate person doesn't understand; he can speak it, but he can't write it.’ (9A; 07:13)

Followed by an object or dependent clause, which functions as an object, *kong*^{2} took on the form *kong*$^{2-1}$. An example of *kong*$^{2-1}$ followed by a noun was shown in (1), and an example in which *kong*$^{2-1}$ was followed by a dependent clause in (17):

(1) A^{0} ka$^{3-2}$ tsit$^{4-8}$ tsun^{7}, eh-- eh-- tau$^{3-2}$te^{2} =*ne*, lan$^{2-1}$ i$^{2-1}$tsing$^{1-7}$ u$^{7-3}$ **kong$^{2-1}$** =kue$^{3-2}$
RLT arrive this moment HST HST in.the.end =RLV 1INCL already[18] EXST **say** =EXP

siann$^{2-1}$mih$^{8-3}$ hue^{7}?
what thing

‘So and up to this moment eh-- eh-- what exactly have we already talked about?’ (5A; 06:23)

(17) Ia^{7}-- ia^{7}-- ia$^{7-3}$ e$^{7-3}$sai$^{2-1}$ **kong$^{2-1}$** han$^{3-2}$ji^{7} e$^{7-3}$sai$^{2-1}$ theh$^{8-3}$ lai$^{5-7}$
as.well as.well as.well be.allowed.to **say** Chinese.character be.allowed.to take come

tso$^{3-2}$ tsok$^{4-8}$ tse$^{7-3}$ tsiong$^{2-1}$ gi$^{2-1}$gian^{5} e$^{5-7}$ su$^{1-7}$sia$^{2-1}$ bun$^{5-7}$hian^{3} <oo> <tioh^{8}>.
do very many kind.of language SUB write document SRPS right

‘You can also-- also-- also say that Chinese characters can be used to make written documents for a lot of different kinds of languages <ow> <right>.’ (55B; 17:16)

Kong$^{2-1}$ then came to be used in conjunction with other verbs, at first probably with verbs of speech such as *me*^{7} ‘scold’ or *mng*^{7} ‘ask’ (33).

18 Note that the speaker uses *i*$^{2-1}$*tsing*^{1} to mean ‘already’. The regular Taiwanese pronunciation is *i*$^{2-1}$*king*^{1}. This is probably due to influence of the Mandarin form *yǐjīng* ‘already’.

(33) I[1-7] le[1] **mng[7-3]** **kong[2-1]** li[2-1] kam[1] tsiah[8-3] png[7] a[0].
3 PROG **ask** **say** 2SG QW eat cooked.rice PERF

'She is asking if you have eaten.' (Gāo 2007: 87)

Because *kong*[2-1] is a verb of speech with a very general meaning, 'say', it became reanalyzed as a complementizer in such constructions: a marker introducing a clause to a matrix verb. Then *kong*[2-1] started to be used with other verbs as well, such as *tsai*[1-7]*iann*[2] 'know' (19), *siunn*[7] 'think' (37) and other verbs denoting cognitive states – in the case of *siunn*[7-3] *kong*[2-1] this lead to syntactic polysemy for some TSM speakers.

(19) A[0] khi[5-7]sit[8] li[2-1] na[7-3] thak[8-4] hit[4-8] lo[7]-- lan[2-1] jin[5-7]lui[7] hit[4-8] e[5-7] bun[5-7]ji[7] e[5-7] ian[2-1]piat[4]
RLT actually 2SG if study that sort 1INCL humanity that entity writing.system SUB evolve

li[2-1] toh[8-3] **tsai[1-7]iann[2-1]** **kong[2-1]** khi[5-7]sit[8] tsit[8-4] tsiong[2-1] ji[7] si[7-3] khah[4-1]-- khah[4-1]
2SG just **know** **say** actually one kind.of character be relatively relatively

tsa[2] tsin[3-2]tsing[5] u[7] e[5-0].
early before EXST SUB

'And actually if you study uhm-- the evolution of mankind's writing systems, you know that actually one kind of *jī* 'writing' existed relatively-- relatively early on.' (35C; 12:55)

(37) A[0] soo[2-1]i[2-1] **siunn[7-3]** **kong[2-1]** =*a*, u[7-3] e[5-7] khah[4-1] kan[2-1]tan[1] e[5-7] tsit[4-8] e[5-7]
RLT therefore **want.to** **say** =EC EXST SUB relatively simple SUB this entity

han[3-2]ji[7-3] =*a*, **siunn[7-3]** **kong[2-1],** Hua[5-7]gi[2] kong[2-1] *tiāndì* =*de* *tiān*
Chinese.character =EC **want.to** **say** Mandarin say heaven.and.earth =SUB heaven

tioh[8] <hmm> <hann[0]> bo[5-0] honn[0], a[0] lan[2-1] Tai[5-7]gi[2] ma[7-3] kong[2-1] <thinn[1]>
right BKCH CFM NEG.EXST XPC RLT 1INCL Taiwanese.language also say heaven

thinn[1-7] kah[4-1] thian[1] =ma[0] <henn[0]> <tioh[8]> honn[0]?
heaven and heaven =DASS CFM right XPC

'So I want to say, some of the simpler characters, I want to say, in Mandarin we say the *tiān* 'heaven' of *tiāndì* 'heaven and earth', <hmm> right <yes>, in Taiwanese we also say <*thinn*> *thinn* 'heaven' and *thian* 'heaven', you know <yes> <right> right?' (45D; 15:11)

Presumably *kong*[2-1] then started to be used with verbs denoting concrete actions, but which were used abstractly such as *pai*[5-7]*thik*[8] 'physically exclude' used in the sense of 'psychologically exclude' (51) and *pang*[3-2]*-tiau*[7] 'dispose of' used in the sense of 'do away with psychologically' (31).

(51) A^{0} Lo$^{5-7}$ma$^{2-1}$ ji^{7} u$^{7-3}$ e$^{5-7}$ lang^{5} long$^{2-1}$ e$^{7-3}$ ka$^{7-3}$ i$^{1-7}$-- ka$^{7-3}$ i$^{1-7}$ **pai$^{5-7}$thik$^{8-4}$**
RLT Rome character EXST SUB person in.every.case will PRTC 3 PRTC 3 **exclude**

kong$^{2-1}$ he^{1} Se$^{1-7}$hong^{1} e^{5}.
say that:entity west SUB

'But some people will always ... them saying "Those are from the West."' (39C; 13:58)

(31) A^{0}-- A^{0} jin$^{5-7}$lui^{7} e$^{5-7}$ kiong$^{7-3}$iu$^{2-1}$ tsu$^{1-7}$san^{2}, a^{0} lan$^{2-1}$ na$^{7-3}$ long$^{2-1}$ h-- ka$^{7-3}$-- ka$^{7-3}$
RLT RLT humanity SUB shared property RLT 1INCL if in.every.case HST PRTC PRTC

pang$^{3-2}$-tiau$^{7-3}$ **kong$^{2-1}$** he^{1} long$^{2-1}$ Se$^{1-7}$hong^{1} e$^{5-0}$ a^{0} koh$^{4-1}$ lan$^{2-1}$ bo$^{5-7}$ ai^{3}...
release-away **say** that:entity in.every.case West SUB RLT again 1INCL NEG.EXST want

'But-- but mankind's shared property, but if we all th-- do away with it like "those are all from the West, and we don't want it"...' (40C; 14:11)

In this process of grammaticalization, the transitive verb *kong^{2} ~ kong$^{2-1}$* 'say', which has two forms, became the complementizer *kong$^{2-1}$* with only one form.

Apart from this process of grammaticalization into a complementizer, combinations of the the transitive verb *kong^{2} ~ kong$^{2-1}$* with the words *soo$^{2-1}$i^{2}* 'so, therefore', *si^{7}* 'be' and *tshiunn^{7}* 'resemble' have probably lexicalized. I define lexicalization as the process of becoming a lexical item.

3 Conclusions

The process of writing this thesis has lead to some observations and findings which fall outside the scope of describing *kong*^{2} ~ *kong*$^{2\text{-}1}$ because they are of a more general nature. I will present them here, along with a summary of the description of *kong*$^{2\text{-}1}$ presented in § 2.

The first is that in the description of a language, whichever language that may be, it is essential to account for one's data. It is important to gather data that is best suited for one's research, but it is equally important to account for the data to the reader. Accounting for linguistic data involves three components: provicing the motivation for 1) the selection of the type of data one uses, 2) the process of gathering data, and 3) the presentation of the data. These three components correspond to three basic questions: 1) Why does this type of data suit the research? 2) How are the data gathered? And 3) How are the data presented? Together, these three components make up a complete, comprehensive account of the raw materials that serve as input for the analysis.

Providing a motivation for the selection of the type of data one opted for is important for two reasons: first, it forces the author to think about *what type* of data they use, and not just about *what* data they use. To stop and think for a moment about the type of data used will help prevent drawing conclusions on the basis of data from which no such conclusions can be drawn. Second, giving a motivation for choosing the type of data one opted for is important for the reader as well. This way, the reader too is forced to think about whether the author can justifiably arrive at the conclusions the latter has drawn on the basis of the data used; for in this respect, the reader has a similar responsibility to that of the author: to try to gain knowledge by drawing correct conclusions.

Accounting for the way in which data was gathered is important for similar reasons. Choosing *the type of data* that is well suited to the research does not mean *the actual data* one gathers and uses is good. Let me give an example drawn from own experiences. When, I decided to attempt a corpus-based description of the use of the TSM word *hoo*^{7} 'give', I meant to focus on the relation between what from the perspective of languages such as Mandarin and Dutch were perceived as a passive-like function and a causal-like function of *hoo*^{7}. In order to do this, I used spoken data and transcribed the sentences in which I heard the morpheme *hoo*^{7} and which seemed relevant for my research. In other words, I did not transcribe all instances of *hoo*^{7}, but reviewed my spoken corpus and selected what I thought was relevant. I now realize it was a gross error to select data beforehand; for in the stage of data-gathering, one cannot possibly say which data is relevant for the study and which is not. It might be that, at the stage of data-gathering, some construction or example seems irrelevant, whereas later on, it does prove to be important for the analysis. Having realized this, I subsequently redid the whole data-gathering process, this time including all instances

o f *hoo*^{7} or its sandhied form *hoo*$^{7\text{-}3}$ no matter how unrelevant they seemed at first sight. Not pre-selecting data may seem cumbersome and it does not guarantee better findings, but it enhances the possibility of arriving at better conclusions and it definitely does present one with a theoretically more justifiable research method – thus giving more credibility to the study.

The third component in responsible accounting for one's data is their representation. The fact that writing serves as a major medium of transmitting and disseminating research, has a huge impact on the representation of one's data. A large amount of valuable information is usually lost when spoken data is represented in written form. As has been put forward by others (cf. Jespersen 1924: 17-18; Chao 2011[1968]: 44; Wiedenhof 1995: 7-8), this goes for most of the suprasegmental part of the data, especially prosodic information such as pitch, duration, volume and intonational contour. For reasons of conveniance and readability, most authors limit their presentation of the data to that part of the data that they deem useful for their study. Some authors take a fairly broad approach to the presentation of their data, indicating such suprasegmentals as tone, tone sandhi, duration or stress, while others do not even present vowels and consonants, basically limiting their data to lexical items and the order in which they are arranged chronologically.[19] The present study is an exception to this in that it has representated such suprasegmentals as tone and tone sandhi systematically. Intonational contours have to some extent been presented using punctuation marks such as comma's and periods, but the main purpose of this has been to increase the readability of long examples and ease the burden of having to read spoken language for the reader not accustomed to this. Futhermore, I have specifically checked data with multiple native speakers, and I have made the spoken data used in this study online available so that the reader may check my transcriptions for themselves.

Being acurate in the presentation of one's data helps to prevent that important details that manifest themselves in aspects of the data that are less easily encoded in writing get lost. In the present study, the transcription of tone sandhi has turned out to be relevant since in some syntactic contexts, for example when functioning as a complementizer, *kong*$^{2\text{-}1}$ can only appear in its sandhied form. As has been noted in § 2.2.1, not all authors investigating the different functions of *kong*$^{2\text{-}1}$ have transcribed tone sandhi, or even tone per se.

Surveying the literature on *kong*^{2} ~ *kong*$^{2\text{-}1}$, I noticed that accounting for one's data is by no means something every author deems necessary. Some dive into the problem they are discussing right away and do not discuss where their data comes from at all, or only briefly mention it in

19 This is the case when one uses a non-alphabetic script, such as the Chinese character script, for the presentation of one's examples.

passing; others take for granted that the reader understands their way of transcription or merely mention the name of a transcription system without accounting for any other aspects of data representation, such as the choice to use glosses or not. In accounting for their data, the difference between different authors is striking; some being really extensive in their accounts while others keep it to a minimum or even skip it at all.

Regarding the literature of *kong*^{2} ~ *kong*$^{2\text{-}1}$, one remarkable phenomenon I noticed was that not as single one of the works I consulted uses data which documents code-switching or non-canonical language. I use *non-canonical* here to mean linguistic forms that are not found in dictionaries. *Code-switching* is used to describe the switching of one language variety to another either within different contexts or within the same sentence. Virtually all Taiwanese speakers code-switch (Yáng 1991: 1-2, there called "code-mixing"), as is also apparent from my data. In light of this, it is striking that no trace of Mandarin can be found in the examples of the literature on *kong*^{2} ~ *kong*$^{2\text{-}1}$. An average reader not familiar with TSM may not notice the absence of code-switching when perusing such studies. But the absence of a single instance of code-switching in the TSM data used in a study bears directly on the reliability and representativeness of these data. Examples of code-switching in my corpus can be found in (22) and (30). Also note the frequent use of the Mandarin clitic =*ne* 'RLV', which is used 16 times in my corpus of 56 sentences whereas the Taiwanese form =*le* ~ =*ne* 'RLV' occurs only twice. Even though the transcribed form *ne* looks similar, the vowel <e> in Hànyǔ Pīnyīn for Mandarin *ne* denotes [ə], whereas in Tâi-lô for TSM *ne* it denotes [e].

Apart from code-switching, examples of non-canonical language I have documented and transcribed are *tsainn*$^{1\text{-}7}$*iann*$^{2\text{-}1}$ 'know' instead of the canonical form *tsai*$^{1\text{-}7}$*iann*$^{2\text{-}1}$ in TSM, and *zōumò* 'weekend' and *bābājié* 'Father's Day' for *zhōumò* and *bàbàjié* in Mandarin respectively. There are also interesting cases in which a Mandarin form seems to have influenced the corresponding TSM form, such as *i*$^{2\text{-}1}$*tsing*$^{1\text{-}7}$ 'already' in (1). The canonical forms of this are *i*$^{2\text{-}1}$*king*$^{1\text{-}7}$ in TSM and *yǐjīng* in Mandarin – the TSM phoneme /ts/ taking on the form [tɕ] preceding a high vowel /i/, which is the same form the Mandarin phoneme /j/ usually has.

Because of the importance of precise transcription of data, it is crucial that the reader can understand the TSM transcription system used – i.e. Tâi-lô with some adaptations – and in order to appreciate this transcription, it is in turn necessary that the reader understand the sound system of Taiwanese. In § 1.2.1 therefore, I ventured to give a fair description of the phonology of TSM, including some of the phonetics involved, while drawing attention to any problematic points.

In § 2.1, I discussed *kong*^{2} ~ *kong*$^{2\text{-}1}$ 'say' as a transitive verb. There, it was classified as a transitive verb because of the covert object that is semantically present even when no overt object

is. An overt object following *kong*$^{2\text{-}1}$ can be a noun or a dependent clause. Roughly 62% – 52 out of a total of 83 instances – of all the instances of *kong*^{2} ~ *kong*$^{2\text{-}1}$ in my corpus are instances of *kong*^{2} ~ *kong*$^{2\text{-}1}$ functioning as a transitive verb. This function seems to be the only one for which both the original form *kong*^{2} and the sandhied form *kong*$^{2\text{-}1}$ are allowed. Functionally, the transitive verb *kong*$^{2\text{-}1}$ 'say' can be used to introduce direct as well as indirect speech.

In the following section, § 2.2, I directed my attention to the 17 instances (or 20%) of *kong*$^{2\text{-}1}$ following a verb and preceding a clause. I classified this function of *kong*$^{2\text{-}1}$ as a complementizer, an element used to introduce a dependent clause indicating indirect speech to its matrix verb. It was noted that *kong*$^{2\text{-}1}$ in this function only occurres with a level high tone, i.e. the sandhied form of *kong*^{2}. Apart from instances in which *kong*$^{2\text{-}1}$ is used to introduce a clause presenting indirect speech to the preceding matrix verb, e.g. in (22), it can also be used between a verb and a clause presenting a quotation, e.g. in (18). I therefore propose a construction of the form V + *kong*$^{2\text{-}1}$ + CLAUSE, in which CLAUSE can either be a dependent clause representing indirect speech, or a quotation representing direct speech.

At the end of section § 2.2, I examined the types of verb that could be V in the construction presented above. 10 out of 17 cases involved a so-called cognitive verb, a verb denoting a psychological state such as *tsainn*$^{1\text{-}7}$*iann*$^{2\text{-}1}$ 'know', whereas 5 cases involved a so-called perceptive verb, a verb denoting that a certain psychological state has been reached such as *siunn*$^{7\text{-}3}$*-tioh*$^{8\text{-}3}$ 'come to think about' or *huat*$^{4\text{-}8}$*hian*$^{7\text{-}3}$ 'discover'. It was noted that most of the matrix verbs occurred just once in my corpus. The conclusions I can draw are therefore limited. More research with larger corpora is needed in order to know which verbs occur more frequently as a matrix verb with *kong*$^{2\text{-}1}$ as a complementizer. That way, we may also notice things that were not found in the present study due its scope. Regarding the morphological structure of the matrix verbs, we saw that even though 14 out of 17 instances (or 17% of all 83 instances of *kong*^{2} ~ *kong*$^{2\text{-}1}$) were simplex verbs, the remaining three – viz. *siunn*$^{7\text{-}3}$*-tioh*$^{8\text{-}3}$ 'come to think about', *tsu*$^{2\text{-}1}$*i*$^{3\text{-}2}$*-tioh*$^{8\text{-}3}$ 'notice' and *pang*$^{3\text{-}2}$*-tiau*$^{7\text{-}3}$ 'do away with' – including a suffix denoting the result of the action denoted by the preceding verb.

In the final section of the main text, attention was drawn to problems of syntactic ambiguity for the interpretation of the collocation *siunn*$^{7\text{-}3}$ *kong*$^{2\text{-}1}$ 'want to say' and *siunn*$^{7\text{-}3}$ *kong*$^{2\text{-}1}$ 'think that...'. I argued that we were dealing with a case of syntactic polysemy and lexical polysemy since the same phonological form can be interpreted as having the either syntactic structure AUX + V with the lexical items *siunn*^{7} 'want' and *kong*$^{2\text{-}1}$ 'say', resulting in *siunn*$^{7\text{-}3}$ *kong*$^{2\text{-}1}$ 'want to say'; or as having the structure V + C with the lexical items *siunn*^{7} 'think' and *kong*$^{2\text{-}1}$ 'COMPL' resulting in *siunn*$^{7\text{-}3}$ *kong*$^{2\text{-}1}$ 'think that'.

In this section, I also accounted for fixed collocations with *kong*$^{2-1}$. These collocations were *si*$^{7-3}$ *kong*$^{2-1}$ 'I mean'; the back-on-tracker *soo*$^{2-1}$*i*$^{2-1}$ *kong*$^{2-1}$ 'so', used in a conversation to return to a certain topic, in this respect resembling English 'anyway'; and finally *tshiunn*$^{7-3}$ *kong*$^{2-1}$ 'just like'. I proposed that these three collocations have been lexicalized to some degree, since they have undergone a semantic shift, i.e., they are not just the sum of the meanings they are composed of.

Finally, I briefly discussed the notions of grammaticalization and lexicalization and the problem of defining these two notions without getting lost in the web of terminology that surrounds both these terms. I propose the following grammaticalization process: *kong*^{2} ~ *kong*$^{2-1}$ started out as a transitive verb meaning 'say', taking on the form *kong*$^{2-1}$ when followed by a noun or dependent clause functioning as its object. It then came to be used following verbs of speech such as *mng*^{7} 'ask' and *me*^{7} 'scold'. Following such verbs of speech, *kong*$^{2-1}$ 'say', being the most general of all verbs of speech semantically, came to be reanalyzed as a complementizer. *Kong*$^{2-1}$ then started to be used with other verbs, such as the verbs *siunn*^{7} 'think' and *hi*$^{1-7}$*bang*^{7} 'hope' denoting cognitive states, and later with verbs denoting concrete actions but which were used abstractly such as *pai*$^{5-7}$*thik*^{8} 'physically exclude' used as 'psychologically exclude' and *pang*$^{3-2}$*-tiau*^{7} 'dispose of' used as 'do away with psychologically'. In the process of grammaticalization, not only the meaning of *kong*^{2} ~ *kong*$^{2-1}$ changed, but also its form: as a verb meaning 'say', it has the forms *kong*^{2} and *kong*$^{2-1}$, but as a complementizer, it has just the form *kong*$^{2-1}$.

Appendix: A spoken corpus of *kong*^{2}

This appendix lists all sentences in the episode 毋通做台語的青盲牛：台語文書寫 *M*$^{7-3}$ *thang*$^{1-7}$ *tso*$^{3-2}$ *Tai*$^{5-7}$*gi*^{2} *e*$^{5-7}$ *tshenn*$^{1-7}$*me*$^{5-7}$*gu*^{5}*: Tai*$^{5-7}$*gi*$^{2-1}$*bun*^{5} *tsu*$^{1-7}$*sia*^{2} 'Don't be illiterate in Taiwanese: Writing Taiwanese' of the radio show 做伙守台語 *Tso*$^{3-2}$*hue*^{2} *tsiu*$^{2-1}$ *Tai*$^{5-7}$*gi*^{2} 'Looking after Taiwanese together' in which *kong*^{2} ~ *kong*$^{2-1}$ was attested. The sentences are numbered in order of appearance, with a capital letter indicating the speaker, as follows:

Speaker A: Ngoo$^{5-7}$ Soo$^{3-2}$hong^{7} 吳素鳳 (host; f.)
Speaker B: Soo$^{1-7}$ Se$^{3-2}$hiong^{5} 蘇世雄 (co-host; m.)
Speaker C: Li$^{2-1}$ Khin$^{5-7}$huann^{7} 李勤岸 (guest; m.)
Speaker D: Lim$^{5-7}$ Ka$^{1-7}$i^{5} 林佳怡 (guest; f.)

1A Soo$^{2-1}$i$^{2-1}$ *=ne*, tak$^{8-4}$ke^{1} ki$^{7-3}$ tse^{1} pai$^{3-2}$lak^{8} hoo^{0}, tse^{1} *zōumò =de shíjiān*,
therefore =RLV everyone be.at[20] this:entity Saturday XPC this:entity weekend =SUB time

it$^{4-8}$ting$^{7-3}$ *=ne*, e$^{7-3}$ kam$^{2-1}$kak$^{4-8}$ **kong$^{2-1}$** e^{0} tse^{1} sim$^{1-7}$tsing^{5} khah$^{4-2}$ ho^{2}
definitely =RLV will feel **say** ATT this:entity emotional.state relatively good

khah$^{4-1}$ khin$^{1-7}$sang^{1}.
relatively relaxed

'So, this Saturday, right, this weekend, everyone will definitely feel that this feeling is rather nice, rather relaxed.' (04:57)

2A A^{0} ua$^{2-1}$ tsong$^{2-1}$si$^{7-3}$ **kong$^{2-1}$**, *māmājié bābājié*, *bābājié =ne*, *māmā yě yìqǐ*
RLT 1SG always **say** Mother's.Day Father's.Day Father's.Day =RLV mother also together

guò jié.
pass holiday

'And I always say that with Mother's and Father's Day, on Father's Day, the mother also celebrates it together with him.' (05:24)

3A A^{0} soo$^{2-1}$i$^{2-1}$ kho$^{2-1}$ling^{5} *=ne*, eh-- koh$^{4-1}$ ui$^{7-3}$ thiann$^{1-7}$tsiong$^{3-2}$ ping$^{5-7}$iu^{2} e$^{7-3}$ kam$^{2-1}$kak$^{4-8}$
RLT therefore maybe =RLV HST every place listening.audience friend will feel

20 I assume that the speaker intended to say *ti*^{7} 'be at'.

kong$^{2-1}$ m^{0} khak$^{4-8}$sit^{8} *=ne*, tsin$^{1-7}$tsiann^{2} tsainn$^{1-7}$iann$^{2-1}$-- beh$^{4-1}$ tsainn$^{1-7}$iann$^{2-1}$
say AGR indeed =RLV really know FUT know

i$^{1-7}$ e$^{5-7}$ i$^{3-2}$su^{3}.
3 SUB meaning

'So it may be that, uhm-- our listeners will feel that, hm yes, they really know-- want to know its meaning.' (05:40)

4A A^{0} soo$^{2-1}$i$^{2-1}$ **kong$^{2-1}$**, ti$^{7-3}$ tsia^{1} *=ne*, *tòuguò* *=ne*, Soo$^{1-7}$ Se$^{3-2}$hiong^{5} *lǎosī* hoo^{0},
RLT therefore **say** be.at here =RLV via =RLV PNSoo PNSè-hiông teacher XPC

ua$^{2-1}$ ma$^{7-3}$ hi$^{1-7}$bang$^{7-3}$ **kong$^{2-1}$** i$^{1-7}$ e$^{7-3}$tang$^{3-2}$ sio$^{2-1}$ kan^{9} kai$^{2-1}$sueh^{4} tsit$^{8-0}$e$^{7-0}$.
1SG also hope **say** 3 be.able.to a.little OBJ:1INCL explain a.while

'So that's to say that, here, through, you know, Mr. Soo Sè-hiông, I also hope that he can explain it a bit to us.' (06:07)

5A A^{0} ka$^{3-2}$ tsit$^{4-8}$ tsun^{7}, eh-- eh-- tau$^{3-2}$te^{2} *=ne*, lan$^{2-1}$ i$^{2-1}$tsing$^{1-7}$ u$^{7-3}$ **kong$^{2-1}$**
RLT arrive this moment HST HST in.the.end =RLV 1INCL already[21] EXST **say**

=kue$^{3-2}$ siann$^{2-1}$mih$^{8-3}$ hue^{7}?
=EXP what thing

'So and up to this moment eh- eh- what exactly have we already talked about?' (06:23)

6B A^{0} lan$^{2-1}$ kin$^{1-7}$a$^{2-1}$jit^{8} e$^{5-7}$-- pun$^{2-1}$ tsiu^{1} e$^{5-7}$ tsu$^{2-1}$te^{5} *=ne*, si$^{7-3}$ kio$^{3-2}$tso$^{3-2}$ **kong$^{2-1}$**
RLT 1INCL today SUB this week SUB topic =RLV be be.called **say**

Tai$^{5-7}$gi^{2} ma$^{7-3}$ tioh$^{8-3}$ **kong$^{2-1}$** Tai$^{5-7}$bun^{5}.
Taiwanese also get **say** written.Taiwanese

'So our topic of today-- of this week is called "If you speak Taiwanese, you have speak about written Taiwanese as well".' (06:40)

7A *Nà* koh$^{4-1}$ ui$^{7-3}$ thiann$^{1-7}$tsiong$^{3-2}$ ping$^{5-7}$iu^{2} li$^{2-1}$ thiann$^{1-7}$ kau$^{3-2}$ tsit$^{4-8}$ e$^{5-7}$
well every place listening.audience friend 2SG listen arrive this entity

kam$^{2-1}$kak$^{4-8}$ **kong$^{2-1}$** e^{0} lan$^{2-1}$ long$^{2-1}$ si$^{7-3}$ an$^{2-1}$ne^{1} tsit$^{8-4}$ poo^{7} tsit$^{8-4}$ poo^{7}
feel **say** ATT 1INCL in.every.case be like.this one step one step

lai$^{5-7}$ ʔun$^{5-7}$un$^{5-7}$a$^{2-1}$ **kong$^{2-1}$** hoo$^{7-3}$ tak$^{8-4}$ e^{5} lai$^{5-7}$ liau$^{2-1}$kai^{2} hoo^{0}.
come ʔslowly[22] **say** give every entity come understand XPC

21 Note that the speaker uses *i$^{2-1}$tsing^{1}* to mean 'already'. The regular Taiwanese pronunciation is *i$^{2-1}$king^{1}*. This is probably due to influence of the Mandarin form *yǐjīng* 'already'.

22 The speaker is very unclear here. My informant guesses speaker A says *un$^{5-7}$un$^{5-7}$a$^{2-1}$* 'slowly', but he is not certain.

‘Well, dear listeners, when you've listened up to this and feel like: “Hey, we always do it like this step by step, to talk slowly about it to you so you will understand it, you know.’ (06:57)

8A Lan$^{2-1}$ e$^{7-3}$hiau$^{2-1}$ **kong$^{2-1}$** Tai$^{5-7}$gi^{2} a^{0} tan$^{7-3}$si^{7} =*ne*, tsit$^{4-8}$ e$^{5-7}$ Tai$^{5-7}$gi^{2} tau$^{3-2}$te^{2}
1INCL know.how.to **say** Taiwanese RLT but =RLV this entity Taiwanese in.the.end

sann$^{2-1}$mih$^{8-3}$ hue^{7}, henn^{0}, **kong$^{2-1}$** sit$^{8-4}$tsai^{7} ua$^{2-1}$ ma$^{7-3}$ si$^{7-3}$-- tui$^{3-2}$ Ta$^{5-7}$gi^{2}
what thing CFM **say** truthful 1SG also be DIR Taiwanese

beh$^{4-1}$ su$^{1-7}$sia^{2} ma$^{7-3}$ si$^{7-3}$ tshenn$^{1-7}$me$^{5-7}$ gu^{5}.
FUT write also be blind buffalo

‘We can speak Taiwanese, but, what this Taiwanese actually is, yes, if I'm being honest, I am also-- when it comes to wanting to write Taiwanese, I am also illiterate.’ (07:05)

9A A^{0} tshenn$^{1-7}$me$^{5-7}$ gu^{5} m$^{7-3}$ bat^{4}, e$^{7-3}$hiau$^{2-1}$ **kong^{2}** be$^{7-3}$hiau$^{2-1}$ sia^{2}.
RLT blind buffalo NEG familiar.with know.how.to **say** NEG:know.how.to write

‘But an illiterate person doesn't understand; he can speak it, but he can't write it.’ (07:13)

10B A^{0} gua$^{2-1}$ koh$^{4-2}$ tsit$^{8-4}$ e$^{5-7}$ kuan$^{1-7}$he^{7}-- toh$^{8-3}$ si$^{7-3}$ **kong$^{2-1}$** gua$^{2-1}$ tsin$^{1-7}$-- tsin$^{1-7}$
RLT 1SG again one entity relation just be **say** 1SG real real

huann$^{1-7}$hi^{2} i$^{1-7}$ tik$^{8-4}$piat$^{8-4}$ koh$^{4-8}$ si$^{7-3}$ gua$^{2-1}$ hit$^{4-8}$ e$^{5-7}$ gian$^{2-1}$kiu$^{3-2}$soo$^{2-1}$ kai$^{1-7}$tuann^{7} e$^{5-7}$
happy 3 special again be 1SG that entity graduate.school phase SUB

tsi$^{2-1}$to$^{7-3}$ kau$^{3-2}$siu^{7}.
supervise professor

‘And I have an extra tie with him-- I mean, I am very-- very happy he is especially also my supervising professor during my time at graduate school.’ (07:57)

11B A^{0} soo$^{2-1}$i$^{2-1}$ kin$^{1-7}$a$^{2-1}$lit^{8} e$^{7-3}$tang$^{3-2}$ tshiann$^{2-1}$ Khin$^{5-7}$huann$^{7-3}$ lau$^{7-3}$su^{1} lai$^{5-7}$ kau$^{3-2}$
RLT therefore today be.able.to invite PNKhîn-huānn teacher come arrive

hian$^{7-3}$tiunn^{5}, a^{0} e$^{7-3}$sai$^{2-1}$ **kong$^{2-1}$** tsin$^{1-7}$ huann$^{1-7}$hi^{2} tsin$^{1-7}$ huann$^{1-7}$hi^{2} la^{0}
present.stage RLT be.allowed.to **say** real happy real happy WARN

honn^{0} <si^{7}>.
XPC be

‘So to be able to invite Mr. Khîn-huānn to come here today, you may say that I am very, very happy, you know <yes>.’ (08:04)

12A Bo$^{5-7}$ siunn$^{7-3}$-tioh$^{8-3}$ **kong$^{2-1}$** Tai$^{5-7}$gi^{2} tsia^{1} u$^{7-3}$ hak$^{8-4}$bun^{5} ho^{0}.
NEG:EXST think-get **say** Taiwanese that.much EXST erudition XPC

'I never would have thought that there was so much to speaking Taiwanese, you know.' (09:06)

13A Siunn$^{7-3}$beh$^{4-1}$ oh^{8} *=ne*, ha^{0}, lang^{5} long$^{2-1}$-- tsong$^{2-1}$si$^{7-3}$ hoo^{0}, eh-- be$^{7-3}$tang$^{3-2}$ **kong$^{2-1}$**
want.to learn =RLV RLT person in.every.case always XPC HST NEG:be.able.to **say**

eh-- s-- *nóu-chū zhèi yàng =de shíjiān chū lái*.
HST HST move-exit this kind =SUB time exit come

'If you want to learn it, oh, people can always-- never, you know, eh-- say eh-- s-- muster up that sort of time.' (09:09)

14A E$^{7-3}$tang$^{3-2}$ *tòuguò* lan$^{2-1}$ e$^{5-7}$ *jiémù* honn^{0}, hoo$^{7-3}$ eh-- tsit$^{8-4}$ kua$^{2-1}$ lan$^{2-1}$ lik$^{8-4}$sik$^{4-8}$
be.able.to via 1INCL SUB program XPC give HST one a.couple.of 1INCL green

ho$^{5-7}$ping^{5} e$^{5-7}$ thiann$^{1-7}$tsiong$^{3-2}$ ping$^{5-7}$iu^{2} *=a* ah$^{8-3}$si$^{7-3}$ ki$^{5-7}$thann^{1} *=de* ping$^{5-7}$iu^{2} honn^{0},
peace SUB listening.audience friend =EC or other =SUB friend XPC

liau$^{2-1}$kai$^{2-1}$ **kong$^{2-1}$** lan$^{2-1}$ long$^{2-1}$ e$^{7-3}$hiau$^{2-1}$ **kong^{2}** tan$^{7-3}$si$^{7-3}$ beh$^{4-1}$ an$^{2-1}$nua^{1}
understand **say** 1INCL in.every.case know.how.to **say** but FUT how

khi$^{2-1}$ poo^{7} honn^{0}, beh$^{4-1}$ an$^{2-1}$nua^{1} <tioh^{8}> si$^{7-3}$ *zǎo ménlù*.
raise step XPC FUT how right be look right.social.connection

'Through our show, we are able to let a couple of our Lik-sik Hô-pîng[23] listeners or other friends, you know, understand like we all know how to speak it, but where to start, you know, where to <right> look for the right people.' (09:21)

15A Soo$^{2-1}$i$^{2-1}$ ai$^{3-2}$ *tòuguò lǎosī*-- li$^{2-1}$ kin$^{1-7}$a$^{2-1}$lit^{8} ai$^{3-2}$ ta$^{7-3}$ uan$^{2-1}$-- ho$^{2-1}$-ho^{2} lai$^{5-7}$
therefore have.to via teacher 2SG today have.to PRTC[24] 1EXCL good-good come

kong$^{2-1}$ hoo$^{7-3}$ lan$^{2-1}$ thiann$^{1-7}$tsiong$^{3-2}$ ping$^{5-7}$iu^{2} tsai$^{1-7}$iann^{2} hoo^{0} <si^{7}>.
say give 1INCL listening.audience friend know XPC be

'So through Mr.-- today you have to let us-- nicely let our listeners know it <yes>.' (09:33)

16A Soo$^{2-1}$i$^{2-1}$ pun$^{2-1}$lai^{5} toh$^{8-3}$ an$^{2-1}$ne^{1} **kong^{2}** hoo^{0}, a^{0} **kong$^{2-1}$** tshenn$^{1-7}$me$^{5-7}$ gu^{5},
therefore originally just like.this **say** XPC RLT **say** blind buffalo

23 Lik-sik Hô-pîng (in full: Lik-sik Hô-pîng Tiān-tâi 'Green Peace Radio Station'; FM 97.3) is the name of the radio station this show is aired on. Note that this radio station is not to be confused with the environmental organisation that goes by the name of Greenpeace.

24 The speaker seems to have meant *ka$^{7-3}$* 'PRTC'.

tshenn$^{1-7}$me$^{5-7}$ gu^{5} ma$^{7-3}$ si$^{7-3}$ e$^{7-3}$tang$^{3-2}$ **kong$^{2-1}$** ue^{7} ma$^{7-3}$ si$^{7-3}$ e$^{7-3}$hiau$^{2-1}$ kap$^{4-8}$
blind bovine also be be.able.to **say** speech also be be.able.to with

lang^{5} kau$^{1-7}$thong^{1} tan$^{7-3}$si$^{7-3}$ =*ne*, beh$^{4-1}$ khuann^{3} khuann$^{3-2}$ bo^{5}, beh$^{4-1}$ sia^{2}
person communicate[25] but =RLV FUT look look NEG:EXST FUT write

sia$^{2-1}$ be^{7} tshut$^{5-0}$ lai$^{5-0}$.
write NEG:will exit come

‘So it's always been said like this, you know, and saying *tshenn-mê gû* ‘blind buffalo’, *tshenn-mê gû* are both able to speak and able to communicate with others, however, when they want to read it, they can't understand it; when they want to write it, they can't write it down.’ (09:39)

17A Soo$^{2-1}$i$^{2-1}$ ti$^{7-3}$ tsia^{1} *yào* tshiann$^{2-1}$ mng^{7} oo^{0}, *lăosī* hoo^{0}, lan$^{2-1}$ Tai$^{5-7}$gi^{2} e$^{7-3}$hiau$^{2-1}$
therefore be.at here want.to invite ask XPC teacher XPC 1INCL Taiwanese know.how.to

kong^{2}, a^{0} tan$^{7-3}$si$^{7-3}$ =*ne*, e-- Tai$^{5-7}$gi^{2} honn^{0}, kam^{1} u$^{7-3}$ ji^{7} e$^{7-3}$hiau$^{2-1}$ sia^{2}--
say RLT but =RLV HST Taiwanese XPC QW EXST character know.how.to write

e$^{7-3}$tang$^{3-2}$ hoo$^{7-3}$ lan$^{2-1}$ sia^{2}?
be.able.to give 1INCL write

‘So at this point, I'd like to ask, you know, you, you know, we can speak Taiwanese, but, b-- this Taiwanese, are there characters to write it, that enable us to write it?’ (09:58)

18A Long$^{2-1}$ m$^{7-3}$ si$^{7-3}$ tsiann$^{3-2}$thong^{2}-- m-- m-- gua$^{2-1}$-- tui$^{3-2}$ gua$^{2-1}$ lai$^{5-7}$ **kong^{2}** la^{0} hoo^{0},
in.every.case NEG be orthodox HST HST 1SG DIR 1SG come **say** WARN XPC

long$^{2-1}$ sia^{2} tshut^{0} lai^{0}, long$^{2-1}$ m$^{7-3}$ si$^{7-3}$-- kam$^{2-1}$kak$^{4-8}$ m$^{7-3}$ si$^{7-3}$ Tai$^{5-7}$gi^{2} =le^{0},
in.every.case write exit come in.every.case NEG be feel NEG be Taiwanese =RLV

soo$^{2-1}$i$^{2-1}$ <tioh^{8} a^{0}> *háisì* *lăosī* ai$^{3-2}$ **kong$^{2-1}$** khah$^{4-2}$ tshing$^{1-7}$tsho^{2} <henn^{0}>
therefore right PERF would.be.better teacher have.to **say** relatively clear CFM

e$^{5-0}$ hoo^{0}.
SUB XPC

‘It's all not the right-- uhm-- uhm-- I-- to me, you know, when you write in them they all don't give-- it feels like it's all not Taiwanese, so <right> it is still necessary that you make it clearer <yes>, you know.’ (10:24)

25 The speaker seems to make a mistake here and says *kau$^{1-7}$thong^{1}* 'traffic'; she most likely meant to say the formally similar word *koo$^{1-7}$thong^{1}* 'communicate'. Presumably this is influence from Mandarin: commonly used Mandarin words such as *gŏu* ‘dog’, *gòu* ‘enough’, *gōu* ‘hook’ all have etymological Taiwanese counterparts with /*au*/: *kau^{2}* ‘dog’, *kau^{3}* ‘enough’, *kau^{1}* ‘reel in’. The first syllable of Mandarin *gōutōng* ‘communicate’, however, is in its TSM counterpart not *kau^{1}*, but *koo^{1}*.

19C Lan$^{2-1}$ sing$^{1-7}$-- lan$^{2-1}$ sing$^{1-7}$ lai$^{5-7}$ **kong$^{2-1}$** hit$^{4-8}$ e$^{5-7}$-- <si^{7}> **kong$^{2-1}$** thau$^{5-7}$a^{2} **kong^{2}**
1INCL first 1INCL first come **say** that entity be **say** at.the.start **say**

e$^{5-7}$ tshenn$^{1-7}$me$^{5-7}$ gu^{5} la^{0} <hann^{0}> honn^{0} <si^{7}>.
SUB blind buffalo WARN CFM XPC be

'Let's first-- let's first talk about that-- <yes> about *tshenn-mê gû* 'blind buffalo' we talked about at the beginning, <hm> right <yes>.' (10:33)

20C A^{0} tshenn$^{1-7}$me$^{5-7}$ gu^{5}, lan$^{2-1}$ Tai$^{5-7}$uan$^{5-7}$ lang^{5} le^{1} **kong$^{2-1}$** tshenn$^{1-7}$me$^{5-7}$ gu^{5}
RLT blind buffalo 1INCL Taiwan person PROG **say** blind buffalo

toh$^{8-3}$ si$^{7-3}$ **Kong$^{2-1}$,** ah-- a^{0} gua$^{2-1}$ toh$^{8-3}$ e$^{7-3}$hiau$^{2-1}$ **kong$^{2-1}$** Tai$^{5-7}$gi^{2} honn^{0} <si^{7}>,
just be **say** HST RLT 1SG just know.how.to **say** Taiwanese XPC be

a^{0} m^{7}koh$^{4-1}$ =he^{0}, ji^{7} honn^{0}, m$^{7-3}$ bat^{4} gua$^{2-0}$ la^{0} <hann0> honn^{0} <tioh^{8}>.
RLT but =RLV character XPC NEG familiar.with 1SG WARN CFM XPC correct

'And *tshenn-mê gû* 'blind buffalo', if we Taiwanese are saying *tshenn-mê gû* then this means uhm-- "I can speak Taiwanese as you know <yes>, but, the characters, you know, they are not familiar with me" <hm>, right <right>.' (10:38)

21C A^{0} soo$^{2-1}$i$^{2-1}$ toh$^{8-3}$ **kong$^{2-1}$-tioh$^{8-3}$** hit$^{4-8}$ e$^{5-7}$ tshenn$^{1-7}$me$^{5-7}$ gu^{5}, sionn$^{7-3}$ **kong$^{2-1}$** lan$^{2-1}$
RLT therefore just **say-get** that entity blind buffalo want.to **say** 1INCL

le^{1} tshi$^{7-3}$ gu^{5}, an$^{2-1}$ne^{1} hit$^{4-8}$ e$^{5-7}$ gu^{5} i$^{1-7}$ toh$^{8-3}$ ah-- m$^{7-3}$ bat$^{4-8}$-- m$^{7-3}$
PROG feed buffalo like.this that entity buffalo 3 just HST NEG familiar.with NEG

bat$^{4-8}$ ji^{7} =ma^{0} <si^{7} a^{0}> honn^{0}.
familiar.with character =DASS be PERF XPC

'So we just say that it's a *tshenn-mê gû* 'blind buffalo', by which we mean to say that, the buffalos we are raising, such a buffalo, he just uhm-- doesn't know-- doesn't know characters, right <oh, yes>.' (10:50)

22C A^{0} i$^{1-7}$ loh$^{8-3}$ long$^{2-1}$ ka$^{7-3}$ gua$^{2-1}$ **kong$^{2-1}$** a^{0} gua$^{2-1}$ na$^{2-1}$ tshenn$^{1-7}$me$^{5-7}$ gu^{5} =*a*
RLT 3 just in.every.case PRTC 1SG **say** RLT 1SG be.like blind buffalo =EC

beh$^{4-1}$ an$^{2-1}$nua^{2} <hann^{0}>?
FUT how CFM

'And they always said to me: "So what if I'm like a *tshenn-mê gú* 'blind buffalo' <hm>?"' (10:59)

23C A^{0} tse^{1} si$^{7-3}$ te^{1} **kong$^{2-1}$** Hua$^{5-7}$gi^{2} a^{0} <onn> honn^{0}.
RLT this:entity be PROG **say** Mandarin ATT SRPS XPC

'And this is Mandarin we're talking about, <oh> mind you. (11:09)

24C Tong$^{1-7}$si^{5} toh$^{8-3}$ si$^{7-3}$ **kong$^{2-1}$** Hua$^{5-7}$gi^{2}-- lan$^{2-1}$ long$^{2-1}$ u$^{7-3}$ siu$^{7-3}$ =kue$^{3-2}$ kau$^{3-2}$iok^{8}
at.that.time just be **say** Mandarin 1INCL in.every.case EXST receive =EXP education

=ma^{0}, khi$^{3-2}$ hak$^{8-4}$hau^{7} thak$^{8-4}$ tsheh^{4} =ma^{0}, <si^{7}> a^{0} soo$^{2-1}$i$^{2-1}$ lan$^{2-1}$ toh$^{8-3}$ bat$^{4-8}$
=DASS go school read book =DASS be RLT therefore 1INCL just familiar.with

ji^{7} <hmm>.
character BKCH

'At that time it was Mandarin--, we have all received education, right, and went to school to study, right <yes>, and so we are literate <hmm>.' (11:12)

25C I$^{1-7}$ toh$^{8-3}$ **kong$^{2-1}$** i$^{1-7}$ m$^{7-3}$ bat$^{4-8}$ ji^{7} toh$^{8-3}$ tshenn$^{1-7}$me$^{5-7}$ gu^{5} <hmm>
3 just **say** 3 NEG familiar.with character just blind buffalo BKCH

'They just said they were illiterate and so were *tshenn-mê gû* 'blind buffalos' <hmm>.' (11:23)

26C A^{0}-- a^{0} soo$^{2-1}$i$^{2-1}$ tsit$^{4-8}$ e$^{5-7}$ tshenn$^{1-7}$me$^{5-7}$ gu^{5}, ki$^{5-7}$-sit^{8} ma$^{7-3}$ bo$^{5-7}$ it$^{4-8}$ting$^{7-3}$
RLT RLT therefore this entity blind buffalo actually also NEG.EXST definitely

kong$^{2-1}$ u$^{7-3}$ sim$^{2-1}$mih$^{8-3}$ phainn$^{2-1}$ i$^{3-2}$su^{3} a^{0} <si^{7}>.
say EXST what bad meaning PERF be

'And-- and so this *tshenn-mê gû* , actually, you know, it isn't necessarily the case that it has a bad meaning <yes>.' (11:25)

27C Toh$^{8-3}$ si$^{7-3}$ **kong$^{2-1}$,** lan$^{2-1}$ i-- Tai$^{5-7}$uan$^{5-7}$ lang^{5} tsa$^{2-1}$tsing^{5} toh$^{8-3}$ long$^{2-1}$ tso$^{3-2}$
just be **say** 1INCL HST Taiwan person before just in.every.case do

sit^{4} =ma^{0} <henn^{0}> soo$^{2-1}$i$^{2-1}$ long$^{2-1}$ iong$^{7-3}$ gu^{5} lai$^{5-7}$ tso^{3} pi$^{3-1}$ju^{7} e$^{5-0}$ <si^{7}>
work[26] =DASS CFM therefore in.every.case use buffalo come do analogy SUB be

tsok$^{4-8}$ tse^{7} la^{0} <henn^{0}>.
very many WARN CFM

'I mean, you know that before, we i-- Taiwanese all used to work in the fields <right>, and so we very often <yes> use buffalo as a metaphor <right>.' (11:29)

28C A^{0} soo$^{2-1}$i$^{2-1}$ toh$^{8-3}$ **kong$^{2-1}$** m$^{7-3}$ bat$^{4-8}$ ji^{7} kio$^{3-2}$ tso$^{3-2}$ tshenn$^{1-7}$me$^{5-7}$ gu^{5}
RLT therefore just **say** NEG familiar.with character call do blind buffalo

26 The combination *tso$^{3-2}$ sit^{4}* originally had the meaning 'to work in the field', nowadays, especially in urban area's, it has acquired the more general meaning of 'to work'. Here, however, the original meaning of 'to work in the field' is meant.

la^{0}, bo$^{5-7}$ sim$^{2-1}$mih$^{8-3}$ phainn$^{2-1}$i^{3}.
WARN NEG.EXST what bad.meaning

'And so we say that illiterate people are called *tshenn-mê gû* 'blind buffalos', right, that has no negative meaning whatsoever.' (11:36)

29C honn^{0}, a^{0} soo$^{2-1}$i$^{2-1}$ i$^{2-1}$tsing^{5} na$^{7-3}$-- na$^{7-3}$-- lang^{5} na$^{7-3}$ **kong$^{2-1}$,** a^{0}-- a^{0} li$^{2-1}$ hit$^{4-8}$ lo$^{7-3}$--
XPC RLT therefore before if if person if **say** RLT RLT 2SG that sort

Tai$^{5-7}$gi^{2} **kong$^{2-1}$** hiah$^{4-1}$ ho^{2}, a^{0} koh$^{4-1}$ m$^{7-3}$ bat$^{4-8}$ ji^{7} <henn^{0}> honn^{0},
Taiwanese **say** so good RLT again NEG familiar.with character CFM XPC

an$^{2-1}$ne^{1} bat$^{4-8}$ ji^{7} tsit$^{4-8}$ sann$^{1-7}$ ji^{7} ma$^{7-3}$ hong^{2} kam$^{2-1}$ma$^{2-1}$ **kong$^{2-1}$**
this.way familiar.with character this three character also give:person feel **say**

phainn$^{2-1}$-- phainn$^{2-1}$ i$^{3-2}$su^{3} <tioh^{8}>.
bad bad meaning right

'Right? So before, if-- if-- if the people said: you-- you speak uhm-- Taiwanese so well, but are illiterate <yes>, right, in this way the three words "you are illiterate" will give you the feeling that it has a bad-- bad meaning <right>.' (11:39)

30C Lan$^{2-1}$ **kong$^{2-1}$** m$^{7-3}$ bat$^{4-8}$ ji^{7}, a^{0} lan$^{2-1}$ na$^{7-3}$ khi$^{3-2}$ hak$^{8-4}$ hau^{7} t-- thak$^{8-4}$
1INCL **say** NEG familiar.with character RLT 1INCL if go school HST study

tsheh^{4} toh$^{8-3}$ bat$^{4-8}$ ji^{7} =ma^{0} honn^{0} <henn^{0} si^{7}>.
book just familiar with character =DASS XPC CFM be

'We say 'illiterate', but when we go to school and l-- learn, we are literate, right <yes, right>.' (12:02)

31A Bo$^{5-7}$ lang^{5}-- <bo^{5} la^{0}> gu^{5} e$^{7-3}$ **kong$^{2-1}$** hoo^{0}, *wǒ* toh$^{8-3}$ *hěn wúgū* =le^{0},
NEG.EXST person NEG.EXST WARN buffalo will **say** XPC 1SG just very innocent =RLV

<haha toh$^{8-3}$ tioh^{8}> i-n$^{1-7}$ na$^{2-1}$ tsai$^{1-7}$ gua$^{2-1}$ m$^{7-3}$ bat$^{4-8}$ ji^{7} <tioh^{8} a^{0} tioh^{8} a^{0}>?
haha just right 3-PL how know 1SG NEG familiar.with character right PERF right PERF

'There is no one-- <no, there isn't> the buffalo will think, you know, I'm completely innocent, <haha, that's right> how do they know I'm illiterate <yes, that's right>?' (12:22)

32C A^{0} soo$^{2-1}$i$^{2-1}$ **kong$^{2-1}$** ji^{7}, toh$^{8-3}$ si$^{7-3}$ ai$^{3-2}$ sing$^{1-7}$ **kong$^{2-1}$** sim$^{2-1}$mih^{8} kio$^{3-2}$ tso$^{3-2}$
RLT therefore **say** character just be have.to first **say** what call do

ji^{7} la^{0}.
character WARN

‘So if we talk about *jī* ‘characters’, we have to first say what we call *jī*, you know.’ (12:35)

33C A^{0} ji^{7} toh$^{8-3}$ si$^{7-3}$ **kong$^{2-1}$** lan$^{2-1}$ ka$^{7-3}$ i$^{1-7}$ sia^{2} tshut$^{4-0}$ lai$^{5-0}$, a^{0} lan$^{2-1}$ khi$^{3-2}$ ka$^{7-3}$ i$^{1-7}$
RLT character just be **say** 1INCL PRTC 3 write exit come RLT 1INCL go PRTC 3

thak^{8}, an$^{2-1}$ne^{1} khuann$^{3-2}$ u$^{7-3}$ i$^{3-2}$su^{3} e$^{5-0}$ kio$^{3-2}$ tso$^{3-2}$ ji^{7} =ma^{0} honn^{0} <hmm>.
study this.way look EXST meaning SUB call do character =DASS XPC hmm

‘With *jī* ‘characters’ it is that we write them down, and we go and read them, things like this in which we can see meaning are called *jī* ‘characters’, you know <hmm>.’ (12:39)

34C A^{0} soo$^{2-1}$i$^{2-1}$ ji^{7} lan$^{2-1}$ ti$^{7-3}$ Tai$^{5-7}$uan^{5} n-- thong$^{1-7}$siong^{5} long$^{2-1}$ e$^{7-3}$ hong^{9}
RLT therefore character 1INCL be.at Taiwan HST usually in.every.case will give:person

goo$^{7-3}$kai$^{2-1}$ **kong$^{2-1}$,** a^{0} ji^{7}, toh$^{8-3}$ si$^{7-3}$ Khong$^{2-1}$tsu$^{2-1}$ ji^{7}, toh$^{8-3}$ si$^{7-3}$
misunderstand **say** RLT character just be Confucius character just be

han$^{3-2}$ji^{7}, kio$^{3-2}$ tso$^{3-2}$ ji^{7}, <hmm> honn^{0}.
Chinese.character call do character hmm XPC

‘So *jī* ‘characters’, we in Taiwan n-- usually always misunderstand them like, well *jī* are *Khóng-tsú jī* ‘Confucius’ characters’, *hàn-jī* ‘Chinese characters’ in other words, that is what are called *jī* <hmm>, right?’ (12:45)

35C A^{0} khi$^{5-7}$sit^{8} li$^{2-1}$ na$^{7-3}$ thak$^{8-4}$ hit$^{4-8}$ lo^{7}-- lan$^{2-1}$ jin$^{5-7}$lui^{7} hit$^{4-8}$ e$^{5-7}$ bun$^{5-7}$ji^{7} e$^{5-7}$
RLT actually 2SG if study that sort 1INCL humanity that entity writing.system SUB

ian$^{2-1}$piat^{4} li$^{2-1}$ toh$^{8-3}$ tsai$^{1-7}$iann$^{2-1}$ **kong$^{2-1}$** khi$^{5-7}$sit^{8} tsit$^{8-4}$ tsiong$^{2-1}$ ji^{7} si$^{7-3}$
evolve 2SG just know **say** actually one kind.of character be

khah$^{4-1}$-- khah$^{4-1}$ tsa^{2} tsin$^{3-2}$tsing^{5} u^{7} e$^{5-0}$.
relatively relatively early before EXST SUB

‘And actually if you study uhm-- the evolution of mankind's writing systems, you know that actually one kind of *jī* ‘character’ existed relatively-- relatively early on.’ (12:55)

36C Soo$^{2-1}$i$^{2-1}$ **kong^{2}** khiai^{0} toh$^{8-3}$ si$^{7-3}$ khah$^{4-1}$ koo$^{2-1}$noo^{2} <henn^{0}>, khah$^{4-1}$ lau^{7} e$^{5-7}$
therefore **say** rise:come just be relatively ancient CFM relatively old SUB

bun$^{5-7}$ji^{7}.
writing.system

‘So that is just saying that it's a more ancient <yes>, an older writing system.’ (13:05)

37C A^{0} ki$^{5-7}$sit^{8} koh$^{4-1}$ kue$^{3-2}$ nng$^{7-3}$ tshing$^{1-7}$ tang^{1} i$^{2-1}$au^{7} toh$^{8-3}$ huat$^{4-8}$hian$^{7-3}$ **kong$^{2-1}$**
RLT actually again pass both thousand winter after just discover **say**

lan$^{2-1}$ e$^{5-7}$ ue^{7} lai$^{7-3}$te^{2} u$^{7-3}$ im^{1} =ne^{0}.
1INCL SUB draw within EXST sound =RLV

'And actually after another two thousand years, we discover that there are sounds in our drawings, you know.' (13:24)

38C Na$^{7-3}$-- na$^{7-3}$ e$^{7-3}$hiau$^{2-1}$ sia$^{2-1}$ hia^{1} e$^{5-7}$ im^{1}, gua$^{2-1}$ e$^{5-7}$ bun$^{5-7}$ji^{7} gua$^{2-1}$ e$^{7-3}$hiau$^{2-1}$
if if know.how.to write there SUB sounds 1SG SUB writing.system 1SG know.how.to

kong^{2} gua$^{2-1}$ tong$^{1-7}$jian^{5} ma$^{7-3}$ e$^{7-3}$hiau$^{2-1}$ sia^{2} a^{0}.
say 1SG of.course also know.how.to write PERF

'If-- if I know how to write those sounds, with my writing system, I know how to speak it, then of course I also know how to write it.' (13:34)

39C A^{0} Lo$^{5-7}$ma$^{2-1}$ ji^{7} u$^{7-3}$ e$^{5-7}$ lang^{5} long$^{2-1}$ e$^{7-3}$ ka$^{7-3}$ i$^{1-7}$-- ka$^{7-3}$ i$^{1-7}$ pai$^{5-7}$thik$^{8-4}$
RLT Rome character EXST SUB person in.every.case will PRTC 3 PRTC 3 exclude

kong$^{2-1}$ he^{1} Se$^{1-7}$hong^{1} e^{5}.
say that:entity west SUB

'But some people will always ... them saying "Those are from the West."' (13:58)

40C A^{0}-- A^{0} jin$^{5-7}$lui^{7} e$^{5-7}$ kiong$^{7-3}$iu$^{2-1}$ tsu$^{1-7}$san^{2}, a^{0} lan$^{2-1}$ na$^{7-3}$ long$^{2-1}$ h-- ka$^{7-3}$-- ka$^{7-3}$
RLT RLT humanity SUB shared property RLT 1INCL if in.every.case HST PRTC PRTC

pang$^{3-2}$-tiau$^{7-3}$ **kong$^{2-1}$** he^{1} long$^{2-1}$ Se$^{1-7}$hong^{1} e$^{5-0}$ a^{0} koh$^{4-1}$ lan$^{2-1}$ bo$^{5-7}$ ai^{3}...
release-away **say** that:entity in.every.case West SUB RLT again 1INCL NEG.EXST want

'But-- but mankind's shared property, but if we all th-- do away with it like "those are all from the West, and we don't want it"...' (14:11)

41D A^{0} **kong$^{2-1}$-tioh$^{8-3}$** tsit$^{4-8}$ e$^{5-7}$ han$^{3-2}$ji^{7} honn^{0}, ua$^{2-1}$ lai^{5}-- toh$^{8-3}$ si$^{7-3}$ **kong$^{2-1}$** honn^{0},
RLT **say-get** this entity Chinese.character XPC 1SG come just be **say** XPC

i-- in$^{1-7}$ui$^{7-3}$ ki$^{5-7}$sit^{8} lau$^{7-3}$su^{1} in$^{1-7}$ui$^{7-3}$ tui$^{3-2}$ tsit$^{4-8}$ e$^{5-7}$ bun$^{7-3}$te^{5} i$^{2-1}$king$^{1-7}$ tsiok$^{4-8}$
HST because actually teacher because DIR this entity problem already very

liau$^{2-1}$kai^{2}.
understand

'Now that you mention these *hàn-jī* 'Chinese characters', I will-- I mean, you know, e-- especially because Mr. Lí Khîn-huānn because he already has a very good understanding of this problem.' (14:35)

42D Tan$^{7-3}$si$^{7-3}$ it$^{4-8}$puann$^{1-7}$ lang^{5} =*a* ki$^{5-7}$sit$^{8-4}$ **kong$^{2-1}$-tioh$^{8-3}$** han$^{3-2}$ji^{7} i$^{1-7}$ ma$^{7-3}$
but regular person =EC actually **say-get** Chinese.characters 3 also

siunn$^{7-3}$ **kong$^{2-1}$,** a^{0} han$^{3-2}$ji^{7} kam^{1} m$^{7-3}$ si$^{7-3}$ Hua$^{5-7}$gi^{2} e^{5}?
think **say** RLT Chinese.characters QW NEG be Mandarin SUB

'But regular people, actually, when you mention *hàn-jī* 'Chinese characters', they will think like: "Don't Chinese characters belong to Mandarin?".' (14:41)

43D Toh$^{8-3}$ si$^{7-3}$ **kong$^{2-1}$,** lan$^{2-1}$ tsainn$^{1-7}$iann$^{2-1}$ **kong$^{2-1}$** ha-- Han$^{3-2}$gi^{2} e$^{7-3}$bin^{7} u$^{7-3}$ tsiann$^{5-7}$
just be **say** 1INCL know **say** HST Sinitic below EXST truely

tse$^{7-3}$ gi$^{2-1}$gian^{5}.
many language

'I mean, we know that there are really a lot of languages within the Si-- Sinitic family.' (14:52)

44D A^{0} lan$^{2-1}$ Tai$^{5-7}$gi^{2} ham$^{7-3}$ tsit$^{4-8}$ma^{2} te$^{7-3}$ tsit$^{4-8}$ e$^{5-7}$ Hua$^{5-7}$gi^{2} ah$^{8-3}$si$^{7-3}$
RLT 1INCL Taiwanese.language with now be.at:DUR this entity Mandarin or

kong$^{2-1}$ Kheh$^{4-2}$gi^{2} long$^{2-1}$ si$^{7-3}$ tsia^{1} e$^{5-7}$ gi$^{2-1}$gian$^{5-7}$ lai$^{7-3}$te^{2} e$^{5-7}$ ki$^{5-7}$tiong^{1} e$^{5-7}$
say Hakka.language in.every.case be here SUB language inside SUB among SUB

tsia^{1} e$^{5-7}$ gi$^{2-1}$gian$^{5-7}$ lai$^{7-3}$te^{2} e$^{5-7}$ ki$^{5-7}$tiong^{1} e$^{5-7}$ kui$^{2-1}$ tsiong^{2}.
here SUB language inside SUB among SUB how.many kind.of

'And our Taiwanese together with Mandarin or Hakka, for example, are all a couple of different varieties among those within this language.' (14:56)

45D A^{0} soo$^{2-1}$i$^{2-1}$ siunn$^{7-3}$ **kong$^{2-1}$** =*a*, u$^{7-3}$ e$^{5-7}$ khah$^{4-1}$ kan$^{2-1}$tan^{1} e$^{5-7}$ tsit$^{4-8}$ e$^{5-7}$
RLT therefore want.to **say** =EC EXST SUB relatively simple SUB this entity

han$^{3-2}$ji$^{7-3}$ =*a*, siunn$^{7-3}$ **kong$^{2-1}$,** Hua$^{5-7}$gi^{2} **kong$^{2-1}$** *tiāndì* =*de* *tiān*
Chinese.character =EC want.to **say** Mandarin **say** heaven.and.earth =SUB heaven

tioh^{8} <hmm> <hann^{0}> bo$^{5-0}$ honn^{0}, a^{0} lan$^{2-1}$ Tai$^{5-7}$gi^{2} ma$^{7-3}$ **kong$^{2-1}$** <thinn^{1}>
right BKCH CFM NEG.EXST XPC RLT 1INCL Taiwanese.language also **say** heaven

thinn$^{1-7}$ kah$^{4-1}$ thian^{1} =ma^{0} <henn^{0}> <tioh^{8}> honn^{0}?
heaven and heaven =DASS CFM right XPC

'So I want to say, some of the simpler characters, I want to say, in Mandarin we say the *tiān* 'heaven' of *tiāndì* 'heaven and earth', <hmm> right <yes>, in Taiwanese we also say <*thinn*> *thinn* 'heaven' and *thian* 'heaven', you know <yes> <right> right?' (15:11)

46D Lan$^{2-1}$ ting$^{2-1}$ ting$^{2-1}$ le$^{2-1}$pai^{3} **kong^{2}** e$^{5-0}$ <si^{7}> tsit$^{4-8}$ e$^{5-7}$ bun$^{5-7}$peh$^{8-3}$im^{1}
1INCL above above week **say** SUB be this entity literary.and.colloquial.reading

thian^{1} <tioh^{8}> kah$^{4-8}$ <hmhmhm> thinn^{1} tioh^{8} bo$^{5-0}$ honn^{0}?
heaven right and BKCH heaven right NEG.EXST XPC

'The literary and colloquial readings *thian* 'heaven' <right> and *thinn* 'heaven' <hmhmhm> we talked about <yes> two weeks ago, right?' (15:20)

47D A^{0} tshiunn$^{7-3}$ **kong$^{2-1}$** Hua$^{5-7}$gi^{2} **kong$^{2-1}$** *dì* *shàng* *=de* *dì* <hm> honn^{o}.
RLT resemble **say** Mandarin **say** acreage above =SUB acreage BKCH XPC

'Just like in Mandarin we say *dì* 'acreage' as in *dì shàng* 'on the acreage', <hm> right.' (15:24)

48D Tshiunn$^{7-3}$ **kong$^{2-1}$** tsit$^{4-8}$ e$^{5-7}$ thinn^{1} *=a* te^{7} la^{0} suann^{1} la tsui^{2}
resemble **say** this entity heaven =EC earth WARN mountain WARN water

la^{0} honn^{0} tsia^{1} e$^{5-7}$ han$^{3-2}$ji^{7} long$^{2-1}$ si$^{7-3}$ Tai$^{5-7}$gi^{2} ham$^{7-3}$ Hua$^{5-7}$gi^{2}
WARN XCP here SUB Chinese.character in.every.case be Taiwanese with Mandarin

i-n$^{1-7}$ iong^{7} e$^{5-7}$ ji^{7}, long$^{2-1}$ si$^{7-3}$ kang$^{7-3}$ khuan^{2} e^{0}.
3-PL use SUB character in.every.case be same type SUB

'What I mean is that these *thinn* 'heaven', *tē* 'earth', *suann* 'mountain', *tsuí* 'water', you know, their Chinese characters are all characters used by Taiwanese and Mandarin, they're all the same.' (15:28)

49D Si$^{7-3}$ **kong$^{2-1}$** tsia^{1} e$^{5-7}$ ji^{7} toh$^{8-3}$ si$^{7-3}$ ti$^{7-3}$ =le^{1} hia^{1}, toh$^{8-3}$ tshiunn$^{7-3}$
be **say** here SUB character just be be.at =DUR there just resemble

tu$^{2-1}$tsiah$^{4-2}$ lau$^{7-3}$su^{1} **kong^{2}** e$^{5-0}$ i$^{1-7}$ e$^{5-7}$ lik$^{8-4}$su^{2} tsiok$^{4-8}$ tng^{5} e$^{5-0}$ <tioh^{8}> <hm> honn^{0}.
just.now teacher **say** SUB 3 SUB history very long SUB right BKCH XPC

'That's to say, these characters are there, just like Mr. Lí said just know, their history is very long <yes> <hm> right.' (15:49)

50A Si^{7} si^{7}, soo$^{2-1}$i$^{2-1}$ *yìbān* lai$^{5-7}$ **kong^{2}** long$^{2-1}$ ti$^{7-3}$ lan$^{2-1}$ *Dōngnányǎ* te$^{7-3}$khu^{1}
be be therefore regular come **say** in.every.case be.at 1INCL South.East.Asia region

=a, <tioh^{8}> khah$^{4-1}$ tse^{7} =ma^{0} <tioh^{8} tioh^{8}> honn^{0}, honn^{0}, si^{7}.
=EC right relatively many =DASS right right XPC XPC be

'Right, right, so usually speaking they're all in our South East Asian region, <yes> they're more numerous, you <yes, yes> know, yes.' (16:31)

51A Soo$^{2-1}$i$^{2-1}$ tu$^{2-1}$a^{2}, henn^{0}, an$^{2-1}$ne^{1} **kong^{2}** khiai^{0} toh$^{8-3}$ si$^{7-3}$ i$^{1-7}$ tak$^{8-4}$ke^{5}
therefore just.now yes like.this **say** rise:come just be 3 everyone

long$^{2-1}$-- khah$^{4-2}$ tsa^{2} long$^{2-1}$ goo$^{7-3}$kai^{2} la^{0} honn^{0},
in.every.case relatively early in.every.case misunderstand WARN XPC

'So just now, yes, we were actually saying that everyone was just-- before, everyone misunderstood, right.' (16:37)

52A **Kong$^{2-1}$** Han$^{3-2}$gi^{2} han$^{3-2}$ji^{7}, a toh$^{8-3}$ si$^{7-3}$ le^{1} **kong$^{2-1}$** hua$^{5-7}$bun^{5}
say Chinese Chinese.character RLT just be PROG **say** written.Mandarin

hua$^{5-7}$ji^{7}-- Hua$^{5-7}$gi^{2} =ma^{0} <tioh^{8}> honn^{0}, iong$^{3-2}$ tse^{1} an$^{2-1}$ne^{1} *xiázǎi*
Mandarin.characters Mandarin =DASS right XPC use this:entity like.this narrow

lai$^{5-7}$ **kong^{2},** ki$^{5-7}$sit^{8} *=ne*, lan$^{2-1}$ tse^{1} si$^{7-3}$ *guǎngyì* e$^{5-0}$,
come **say** actually =RLV 1INCL this:entity be broad.meaning SUB

fànwéi *sī* *hén* *guǎng* *=de* <tioh^{8} la^{0}>.
scope be very broad =SUB right WARN

'When they said *Hàn-gí* 'Chinese' or *hàn-jī* 'Chinese characters' they just meant *huâ-bûn* 'written Mandarin' or *huâ-jī* 'Mandarin characters'-- just *Huâ-gí* 'Mandarin' <right>, you know, using these in a narrow way like this, but actually you know, this has a broad meaning with us, its scope is very broad <right>.' (16:42)

53B <*Qíng* *Sū*> U$^{7-3}$ <*lǎosī* a--> tiann$^{7-3}$tiann$^{7-3}$ te^{1} khuann$^{3-2}$ tian$^{7-3}$iann^{2} e$^{5-7}$
invite PNSū EXST teacher HST often PROG look movie SUB

lang^{5} honn^{0}, <henn^{0}> kho$^{2-1}$ling$^{5-7}$ e$^{7-3}$ h-- i^{1} tsu$^{2-1}$i$^{3-2}$-tioh$^{8-3}$ **kong$^{2-1}$**, hit$^{4-8}$ e$^{5-7}$
person XPC CFM possible will HST 3 notice-get **say** that entity

Hiong$^{1-7}$kang$^{2-1}$phinn^{3} honn^{0} <henn^{0}>, i-n$^{1-7}$ e$^{5-7}$ ue^{7} kah$^{4-8}$ i-n e$^{5-7}$ ji^{7}
Hong.Kongese.movie XPC CFM 3-PL SUB speech and 3-PL SUB character

he^{1} long$^{2-1}$ ham$^{7-3}$ gua-n$^{2-1}$ *sǒwèi* *=de* *Guóyǔ* bo$^{5-7}$ kang$^{7-3}$
that:entity in.every.case with 1-PL so.called =SUB National.language NEG.EXST same

khuan^{2} e$^{5-0}$ <tioh^{8} la^{0}> <tioh^{8} tioh^{8}>.
type SUB right WARN right right

'<I want to ask Mr.> There are <Sū a--> the people that often are watching films, right, <yes> maybe they will h-- they notice that, those Hong Kong movies, you know <yes>, their language and Chinese characters those are all different from our so-called *Guóyǔ* 'National language' <indeed> <right, right>.' (16:57)

54B Ia$^{2-1}$ koh$^{4-2}$ u^{7}, lan$^{2-1}$ Kheh$^{4-2}$gi^{2} ma$^{7-3}$ u$^{7-3}$ Kheh$^{4-2}$gi^{2} e$^{5-7}$ ji^{7} =ne,
still again EXST 1INCL Hakka.language also EXST Hakka.language SUB character =RLV

<hmhmhm> he^{1} u$^{7-3}$ ki$^{1-7}$hue^{7} tsiah$^{4-8}$ koh$^{4-2}$ lai$^{5-7}$ kang^{2} ngo^{0}, toh$^{8-3}$ si$^{7-3}$
BKCH that:entity EXST chance only.then again come discuss XPC just be

kong$^{2-1}$ <si^{7} si^{7}> tsia^{1} long$^{2-1}$ si$^{7-3}$ han$^{3-2}$ji^{7} e$^{5-7}$ tsit$^{8-4}$ tsiong^{2} <si^{7} si^{7}>.
say be be here in.every.case be Chinese.character SUB one kind.of be be

‘Furthermore, our Hakka language also has its characters, you know, <hmhmhm> those we can discuss some other time, I'm just saying <yes, yes> that these are all instances of Chinese characters <yes, yes>.’ (17:09)

55B Ia^{7}-- ia^{7}-- ia$^{7-3}$ e$^{7-3}$sai$^{2-1}$ **kong$^{2-1}$** han$^{3-2}$ji^{7} e$^{7-3}$sai$^{2-1}$ theh$^{8-3}$ lai$^{5-7}$
as.well as.well as.well be.allowed.to **say** Chinese.character be.allowed.to take come

tso$^{3-2}$ tsok$^{4-8}$ tse$^{7-3}$ tsiong$^{2-1}$ gi$^{2-1}$gian^{5} e$^{5-7}$ su$^{1-7}$sia$^{2-1}$ bun$^{5-7}$hian^{3} <oo> <tioh^{8}>.
do very many kind.of language SUB write document SRPS right

‘You can also-- also-- also say that Chinese characters can be used to make written documents for a lot of different kinds of languages <ow> <right>.’ (17:16)

56D Soo$^{2-1}$i$^{2-1}$ tsin$^{1-7}$ tse$^{7-3}$ <hm> Tai$^{5-7}$gi^{2} tsin$^{1-7}$ ho^{2} e$^{5-7}$ lang^{5} =*a*, <hm>
therefore real many BKCH Taiwanese real good SUB person =EC BKCH

i$^{1-7}$ toh$^{8-3}$ khuann$^{3-2}$-tioh$^{8-3}$ e$^{7-3}$ **kong$^{2-1}$** honn^{0}-- tse^{1} iong$^{7-3}$ Tai$^{5-7}$gi^{2} sia^{2} e$^{5-7}$
3 just look-get will **say** XPC this:entity use Taiwanese write SUB

bun$^{5-7}$tsiunn^{1} honn^{0}, i$^{1-7}$ toh$^{8-3}$ **kong$^{2-1}$** a^{0} li$^{2-1}$ sia$^{2-1}$ tse^{1} sannh? Gua$^{2-1}$
article XPC 3 just **say** RLT 2SG write this:entity what 1SG

khuann$^{3-2}$ bo^{5} =*a*.
look NEG.EXST =EC

'So a lot of people whose Taiwanese is very good, you know, <hm> they just look at them and will say, you know-- writing articles in Taiwanese with these, you know, they will just say "What are you writing? I can't read it." (17:42)

References

Âng, Uî-jîn 洪惟仁, 2002. “閩南語轉調的語言層次（一）：句法、語義與語用的制約 Mǐnnányú zhuǎndiào de yǔyán céngcì (yī): jùfǎ, yǔyì yú yǔyòng de zhìyuē [The linguistic layers of Southern Mǐn tone sandhi (1): Syntactic, semantic and pragmatic constraints]”. Paper presented at the Fourth International Conference on the Languages of Taiwan and their Teaching, Kaohsiung National Sun Yat-sen University, April 24th-28th, 2002.

Brinton, Laurel J. & Elizabeth Cross Traugott, 2005. *Lexicalization and language change.* Cambridge: Cambridge University Press.

Coulmas, Florian, 1986. “Reported speech: Some general issues” in Florian Coulmas (ed.) *Direct and indirect speech*, Berlin: Mouton de Gruyter, pp. 1-28.

Chao, Yuen Ren, 1930. “A system of tone letters”, *Le maître phonétique* 45, pp. 24-27

—, 1959. “Ambiguity in Chinese” in Søren Egerod en Else Glahn (eds.) *Studia serica Bernhard Karlgren dedicata*, København: Ejnar Munksgaard, pp. 1-13.

—, 2011. *A grammar of spoken Chinese*. Peking: The Commercial Press. First edition: 1968. Berkeley: University of California Press.

Chinese Tools, 2016. Chinese Calligraphy Editor. http://www.chinese-tools.com/jdd/public/callitext/141101105250149608720 3.png. Accessed, May 29, 2017.

Chung, Raung-fu, 1996. *The segmental phonology of Southern Min in Taiwan*. Taipei: Crane.

Diercks, Michael, 2013. “Indirect agree in Lubukusu complementizer agreement”, *Nat lang linguist theory* 31, pp. 357-407.

Douglas, Carstairs, 1899. *Chinese-English dictionary of the vernacular or spoken language of Amoy*. London: Publishing Office of the Presbyterian Church of England.

Foolen, A. et al., 2006. “Het quotatieve *van*. Vorm, functie en sociolinguïstische variatie”, *Toegepaste taalwetenschap in artikelen, nr. 76, Thema’s en trends in de sociolinguistiek*, 5:2, pp. 137-149.

Hopper, Paul J. & Elizabeth Closs Traugott, 2003. *Grammaticalization*. Camebridge: Camebridge University Press.

Jespersen, Otto, 1924. *The philosophy of grammar*. London: George Allan & Unwin.

Khóo, Kik-tūn 許極燉, 1998. 台灣語概論 *Táiwānyǔ gàilùn*. Taipei: 前衛出版社 Qiánwèi Chūbǎnshè, Avanguard Publishing.

Klöter, Henning, 2005. *Written Taiwanese*. Wiesbaden: Harrassowitz.

Kuo, Chen-Hsiu, 2013. *Perception and acoustic correlates of the Taiwanese tone sandhi group*. Dissertation: University of California.

Lau, Seng-hian, 2013. “On non-verbal *kongs* in Taiwanese”, 台灣學誌 *Táiwān xuézhì*, 7, pp. 57-87.

Li, Charles N., 1986. “Direct and indirect speech: A functional study” in Florian Coulmas (ed.) *Direct and indirect speech*, Berlin: Mouton de Gruyter, pp. 29-46.

Li, Charles N. & Sandra A. Thompson, 2009. *Mandarin Chinese: A functional reference grammar*. Taipei: 文鶴 Wénhè, Crane.
First edition: 1981. Berkeley: University of California Press.

Lin, Philip T., 2015. *Taiwanese grammar*. Greenhorn Media.

Lín, Qìngxūn 林慶勳, 2008. 臺灣閩南語概論 *Táiwān Mǐnnányǔ gàilùn* [An introduction to Taiwanese Southern Mǐn]. Taipei: 心理出版社 Xīnlǐ Chūbǎnshè.

Lôo, Kńg-tsiânn 盧廣誠, 2003. 台灣閩南語概要 *Táiwān Mǐnnányǔ gàiyào* [An outline of Taiwanese Southern Mǐn]. Taipei: 南天書局 Nántiān Shūjú.

—, 2011. 實用台語詞典 *Sit$^{8-4}$iong$^{7-3}$ Tai$^{5-7}$gi$^{2-1}$ su$^{5-7}$tian^{2}* / *Shíyòng Táiyǔ cídiǎn* [Practical Taiwanese dictionary]. Taipei: 文水出版社 Wénshuǐ Chūbǎnshè.

Luó, Chángpéi 羅常培, 1956. 廈門音系 *Xiàmén yīnxì* [The sound system of Amoy]. Peking: 科學出版社 Kēxué Chūbǎnshè.
First edition: 1931, Taipei: Ku-tin Book Store.

Matthews, P. H., 2007. *The concise Oxford dictionary of linguistics*. Oxford: Oxford University Press.

MoE (Ministry of Education), 2011a. 臺灣閩南語羅馬字拼音方案使用手冊 *Táiwān Mǐnnányǔ Luómǎzì Pīnyīn fāng'àn shǐyòng shǒucè* [Manual for the usage of the Taiwanese Romanization System]. Taipei: Ministry of Education.

—, 2011b. 教育部臺灣閩南語常用詞辭典 *Jiàoyùbù Táiwān mǐnnányǔ chángyòng cí cídiǎn* [Dictionary of the ministry of Education of frequently used words in Taiwanese Southern Mǐn].
http://twblg.dict.edu.tw/holodict_new/default.jsp; keyword: “誤解” *gōo-kái.* Accessed May 11, 2017.

Norman, Jerry, 2012. *Chinese*. Cambridge: Cambridge University Press.

Ōuyáng, Juéyà 欧阳觉亚, Ráo Bīngcái 饶秉才, Zhōu Yàowén 周耀文 & Zhōu Wújì 周无忌, 2005. 广州话、客家话、潮汕话与普通话对照词典 *Guǎngzhōuhuà, Kèjiāhuà, Cháoshànhuà yú Pǔtōnghuà duìzhào cídiǎn* [Comparative dictionary of Cantonese, Hakka and Teo-Swa with

Standard Chinese]. Canton: 广东人民出版社 Guǎngdōng Rénmín Chūbǎnshè.

Peyraube, Alain & Li Ming, 2012. "The semantic historical development of modal verbs of volition in Chinese" in Janet Zhiqun Xing (ed.) *Newest trends in the study of grammaticalization and lexicalization in Chinese*, Berlin: Mouton de Gruyter, pp. 149-167.

Pīnyīn, 2010. "Rules for using pinyin". http://www.pinyin.info/rules/index.html. Accessed, Nov. 8, 2016.

Pronk, Joren, 2014. "*Hōo*, wat geeft dat?: Een beschrijving van *hōo* 'geven' in het Taiwanees" [A discription of *hōo* 'give' in Taiwanese]. BA-thesis: Leiden University.

Schoppa, R. Keith, 2000. *The Columbia guide to modern Chinese history*. New York: Columbia University Press.

TML (臺灣母語聯盟 Tai$^{5-7}$uan$^{5-7}$ Bu$^{2-1}$gi$^{2-1}$ Lian$^{5-7}$bing^{5} [Alliance for the Native Languages of Taiwan]). Radio show: 做伙守台語 *Tso$^{3-2}$hue^{2} tsiu$^{2-1}$ Tai$^{5-7}$gi^{2}* [Looking after Taiwanese together]. Title: 毋通做台語的青盲牛：台語文書寫 *M$^{7-3}$ thang$^{1-7}$ tso$^{3-2}$ Tai$^{5-7}$gi^{2} e$^{5-7}$ tshenn$^{1-7}$me$^{5-7}$gu^{5}: Tai$^{5-7}$gi$^{2-1}$bun^{5} tsu$^{1-7}$sia^{2}* [Don't be illiterate in Taiwanese: Writing Taiwanese]. https://www.youtube.com/watch?v=3puhXw9V_i0. Accessed Aug. 28, 2016.

Tseng, Ming-hua, 2008. *The multifunction of Taiwanese Southern Min 'Kong2'*. MA-thesis: National Tsinghua University.

Wiedenhof, Jeroen, 1995. *Meaning and syntax in spoken Mandarin.* Leiden: CNWS publications.

—, 2015. *A Grammar of Mandarin*. Amsterdam: John Benjamins.

Wikimedia 2015. "Map of the most commonly used home language in Taiwan"; Author: Yoxem. https://commons.wikimedia.org/wiki/File:Map_of_the_most_commonly_used_home_language_in_Taiwan.svg. Accessed May 30, 2016.

Yang, Barry Chung-Yu, 2006. "Syntactic Structure of Ka-Construction in Taiwanese Southern Min". Paper presented at the National Conference on Linguistics at Hsinchu National Chiao Tung University, 2006.

Yáng, Xiùfāng 楊秀芳, 1991. 臺灣閩南語語法稿 *Táiwān Mǐnnányǔ yǔfǎ gǎo* [A grammatical sketch of Taiwanese Southern Mǐn].Taipei: 大安出版社 Dàān Chūbǎnshè.

Yip, Moira, 2002. *Tone*. Cambridge: Cambridge University Press.

Zhèng, Liángwěi 鄭良偉, 1994. 教學參考用的臺語輕聲規律 *Jiāoxué cānkǎo yòng de Táiyǔ qīngshēng guīlǜ* [Rules for the Taiwanese Neutral tone to be used for reference in education] in Taiwan Language and Literature Society (eds.) *Papers from the 1994 Conference on Language Teaching and Linguistics in Taiwan Vol.I: Southern Min*, pp. 259-295.

Zhāng, Fāngjié 張芳杰, 2009. *Far East Chinese-English dictionary*. Taipei: 遠東圖書公司 Yuǎndōng Túshūgōngsī.

Zhū, Déxī 朱德熙, 1985. 现代汉语语法研究 *Xiàndài Hànyǔ yúfǎ yánjiù* [Grammatical studies in modern Chinese]. Peking: 商务印书馆 Shāngwù Yìnshūguǎn.