

The Focus Marker *Si*⁷ 是 and Lexicalization of *Si*⁷ *Mih*⁸ 是乜 into *What* Wh-words in Earlier Southern Min Texts^{*}

Chinfa Lien

National Tsing Hua University

The paper explores the focus marker *si*⁷ 是 and lexicalization of the syntagm *si*⁷ *mih*⁸ 是乜 into *what* wh-words in earlier Southern Min texts. *Si*⁷ 是 shows a dual role functioning both as copula and focus marker. It can be a focus marker of a wide range of syntactic categories and, among others, a focus marker of subjects and objects expressing new information. As attested in texts, this is due to information structure constraints; focus is marked by *si*⁷ as opposed to topic as a dislocated sentence-initial element carrying old information. It is found that context or sentence processing provides a clue as to whether *si*⁷ 是 is a focus marker or loses its independent status and becomes a desemanticized part of the lexicalized *what* wh-word *si*⁷ *mih*⁸ 是乜. Textual evidence lends support to a thesis featuring a three-stage development of *si*⁷ *mih*⁸ 是乜 in Southern Min that bears on the development of *shi*⁴-*wu*⁴ 是物 in earlier Modern Chinese. Southern Min shows a measure of conservatism in the lexicalization of 是物.

Key words: focus marker, lexicalization, information structure, earlier Southern Min texts

1. Introduction

A survey of earlier Southern Min plays (Wu 2001a, b, c, d), some dating to the 16th century and written in a mixture of Chaozhou and Quanzhou dialects, shows that *si*⁷ 是 'be' is a multifunctional element in that as a lexeme it functions as a main copular verb expressing a relationship of equation or inclusion between subject and nominal

^{*} This paper is based on research partially supported by NSC 95-2411-H-007-013-MY3 and Project on Linguistic Corpora and Explorations of Theoretical Issues. It was presented at *The Past Meets the Present: A Dialogue Between Historical Linguistics and Theoretical Linguistics* hosted by the Institute of Linguistics, Academia Sinica, July 14-16, 2008. I am grateful to the participants for their comments and suggestions, in particular Christine Lamarre, Chaofen Sun, Feng-fu Tsao and Pei-chuan Wei. Thanks are also due to the anonymous reviewers' perceptive comments, which served to set me on the right track on many intractable issues. I owe a debt of gratitude to Paula Rogers for the editorial help and extremely useful suggestions for the penultimate version of this paper.

predicates.¹ But besides its status as a lexeme it has taken on a very general role as a marker of focus for the sequence following it. As (1) shows, *si*⁷ 是 is used as a marker underscoring the following element *li*² 你 ‘you’ as the focus² (see Erteschik-Shir 1997, 2007).

- (1) kiann³ si⁷ li² phah⁴ phoa³ (9.433, SZ)³
 鏡 是 你 打 破
 mirror FM you strike broken⁴
 ‘It is you who broke the mirror.’

Furthermore, *si*⁷ often precedes an interrogative wh-word to focus it, as in (2):

- (2) si⁷ mih⁸-lang⁵ cho³ siann¹ (25.130, JJ)
 是 乜 人 作 聲
 FM what person make sound
 ‘Who is it that made the sound?’ ‘Who made the sound?’

Whereas *si*⁷ in (2) can be construed either as an element independent from the following interrogative wh-word it marks as focus, or as an element that has been fused into the wh-word, the only interpretation in (3) is that *si*⁷ has become an integral part of the *what* wh-word evidently as a result of lexicalization that merges the focus marker and the *what* wh-word (Brinton & Traugott 2005).⁵

¹ The Southern Min texts comprise four plays, viz. the Jiajing (1522-1566), Wanli (1573-1619), Shunzhi (1644-1661), and Guangxu (1875-1908) edition of the *Legend of Litchi Mirror* dating back to the sixteenth century (see Wu 2001a, b, c, d). The approximate publication date of each text is given in parentheses.

² The spelling of Southern Min in this paper is based on the Church Romanization codified in Douglas (1873). Some modifications have been made. In particular, the diacritic tone marks have been abandoned in favor of numerical superscripts. No distinction is made between *ch* and *ts* or *chh* and *tsh* as they do not involve phonemic contrast. Open *o* and closed *o* are rendered as *oo* and *o*, as in *thoo*⁵ 塗 ‘earth’ and *tho*⁵ 桃 ‘peach’. Nasalization is indicated by a double *n*. The rendition of Southern Min sounds is based on the modern pronunciation as a way to approximate the original sound values of earlier Southern Min. Pinyin spelling is used to render Mandarin examples.

³ The source of each example is given in parentheses. The numerals separated by a dot stand for act and line, and the capital letters are abbreviations of the four editions of scripts of play featuring the same legend: JJ (Jiajing), WL (Wanli), SZ (Shunzhi) and GX (Guangxu).

⁴ Abbreviations used in this paper: AM=aspect marker; CP=copula; FM=focus marker; PRF=prefix; and RL=relativizer. A dot means the following syllable is unstressed.

⁵ Lexicalization is taken as a process of unifying a construction into an inseparable unit (Brinton

- (3) li² pat⁴ si⁷ mih⁸ (4.366, SZ)
 你 識 是 乜
 you know (FM) what-matter
 ‘What do you know?’

The dual nature of *si*⁷ can be taken as an indication of grammaticalization in line with the principles of grammaticalization that Hopper proposes (Hopper 1991, Hopper & Traugott 2003). In a nutshell then, this paper aims at establishing a case of the grammaticalization and lexicalization of the copular verb *si*⁷ 是 into a focus marker and the head of wh-words through re-analysis (Heine & Reh 1984) in tandem with the concomitant adjustment of phonological aspects including contracted forms and prosodic constraints. I shall also substantiate the thesis that wh-words sentences are formed as a result of a shift from a focus structure fed by pragmatic information to a grammaticalized syntactic structure. This paper also explores multiple functions of *si*⁷ 是 focusing on the function of focus marker in earlier and later Southern Min texts dating back to sixteenth century at its earliest state. The issues addressed in the paper include syntactic categories and grammatical functions of focused elements marked by *si*⁷ 是, the syntactic positioning of topic and focus, the internal makeup of *what* wh-words, the criteria for identifying *si*⁷-*mih*⁸ 是乜 as a syntagm or a lexicalized *what* wh-word, focus as opposed to topic, and the development of the syntagm *si*⁷-*mih*⁸ 是乜 into a *what* wh-word.

2. The multiple functions of *si*⁷ 是

*Si*⁷ 是 carries many functions that fall into two main types: copula and focus marker.

2.1 *Si*⁷ 是 as a copula

When functioning as copula, *si*⁷ 是 may yield an equative or specificational reading, as in (4a) and (4b).⁶ If it is equative, the postcopular term is referential, and if it is specificational, it is non-referential.

& Traugott 2005:48-50, Lehmann 1995:6-8). Among the nine characteristics of lexicalization that Brinton & Traugott (2005:96-97) touch on and discuss are the loss of semantic components, fusion, and creation of new lexical/contentful forms. The last characteristic may not apply to this case since the *what* wh-word *si*⁷-*mih*⁸ 是物 already belongs to functional category.

⁶ See Cann (2007) for the distinction between equative and specificational reading in the use of copulas.

- (4) a. gun² si⁷ a¹-niu⁵ sin¹ pinn¹ kan²-ji⁵ (17.118, SZ)
 阮 是 啞娘 身 邊 簡兒
 I CP PRF-lady body side maidservant
 'I am my lady's maidservant.'
- b. lin² si⁷ sinn¹-hun⁷ lang⁵ (9.086, SZ)
 恁 是 生分 人
 you CP unacquainted person
 'You are a stranger.'

2.2 *Si*⁷ 是 as a focus marker

Unlike the copula *be* in English which can function as a modal bearing tense and features of agreement, *si*⁷ 是 in Southern Min as well as other Chinese languages does not carry such a function when followed by predicates like verbs and adjectives. It fulfills other functions, one of which to mark the focused element in a sentence.

3. Syntactic categories of focused elements marked by *si*⁷ 是

When *si*⁷ 是 does not occur in predicate position, it functions as a focus marker before an NP. In (5a), for example, it occurs as a focus marker of the noun in subject position. Unlike English where the copula has to bear tense and agreement features, no copula appears before adjectives or verbs in Southern Min unless *si*⁷ 是 functions as a focus marker. *Si*⁷ 是 is a focus marker of adjectives as in (5b) & (5c), and of verbs as in (5d). It can be used to focus modals as in (5e), prepositions as in (5f), reciprocal adverbs as in (5g), time adverbs as in (5h), place adverbs as in (5i), and even whole sentences as in (5j) & (5k).

- (5) a. si⁷ i¹ phah⁴ phoa³ (9.427, SZ)
 是 伊 打 破
 FM he strike break
 'It is he who broke it.'
- b. sit⁸ si⁷ chhing¹-khi³ (07.000, WL)
 實 是 清氣
 really FM clean
 'It is really clean.'

- c. ho² chu¹-niu⁵ si⁷ chhin¹-chhinn² (6.166, JJ)
 好 諸娘 是 親淺
 good woman FM pretty
 ‘The fair lady is pretty.’
- d. guan² si⁷ lai⁵ boa⁵ kiann³
 阮 是 來 磨 鏡
 I FM come polish mirror
 ‘It is for polishing the mirror that I came.’
- e. si⁷ gau⁵ tann³ oe⁷ (16.061, GX)
 是 肴 咀 話
 FM good-at speak word
 ‘(He) is indeed a good talker.’
- f. i¹ pah⁴-poann¹ khoo²-thiann³ to¹ si⁷ ui⁵-tioh⁸ gun² (15.045, SZ)
 伊 百般 苦疼 都是 爲著 阮
 he hundred-fold hardship pain PF for us (exclusive)
 ‘It is for my sake that he suffered traumatic pain.’
- g. i¹ sann¹ lang⁵ si⁷ sann¹ chhoa⁷ chau² (40.022, JJ)
 伊 三 人 是 相 拽 走
 he three person FM mutually take escape
 ‘Three of them eloped.’
- h. jit⁸ loh⁸ se¹-san¹ si⁷ mi⁵-hng¹ (16.030, JJ)
 日 落 西山 是 冥昏
 sun fall west mountain FM night-evening
 ‘It is at dusk that the sun sets.’
- i. goan⁵-lai⁵ chin⁵-nia² chiann³ si⁷ chi² te³ (49.024, JJ)
 原來 秦嶺 正 是 只 處
 original Qin-ridge exactly FM this place
 ‘Qinling mountain range turns out to be here.’
- j. chi² si⁷ goan² iah⁴-chhun¹ tan³ thit⁴-tho⁵ (14.231, SZ)⁷
 只 是 阮 益春 揆 得桃
 this FM we (exclusively) Yichun throw play
 ‘It is that my maid threw it for fun.’

⁷ Such a focused sentence is often preceded by an anaphoric deictic term *chi*² 只 (proximal) or *hu*² 許 (distal).

- k. hu² si⁷ sio²-ti⁷ sit⁴-le² (3.139, SZ)
 許 是 小弟 失禮
 that FM young-brother (humble) lose-etiquette
 'I am sorry for it.'

The focus marker is often preceded by emphatic adverbs like *chiann*³ 正 'very', *si*⁸ 實 'really', and *chiah*⁴ 即 'so', as shown above.

4. Grammatical functions of focused elements marked by *si*⁷ 是

If we take the grammatical function of arguments into consideration, *si*⁷ 是 can mark a subject in situ, as in (6a), or a subject in a dislocated (viz. post-predicate) position, as in (6b).

- (6) a. si⁷ i¹ ma⁷ li² m⁷ si⁷ goa² (31.21, SZ)
 是 伊 罵 你 不 是 我
 FM he scold you not FM I
 'It is he rather than I who scolded you.'
- b. khi² loo⁵ hoat⁴ he² to¹ si⁷ li² (26.233, JJ)
 起 爐 發 火 都 是 你
 light furnace start fire all FM you
 'It is you who started the fire in the furnace.'

It can also mark an object, although less frequently, as in (7a), (7b), and (7c). Although *si*⁷ 是 does not immediately precede the object *Yichun* 益春, we know that the object is in focus from the fact that there is a contrast between *Yichun* and *you* (sg.).

- (7) a. chhoo¹-thau⁵ tiann¹ si⁷ chiau²-siann¹ (13.082, SZ)
 初頭 聽 是 鳥聲
 at.first listen FM bird.sound
 'It is the bird sounds that I listened to at first.'
- b. hong⁵-te³ sinn¹ e si⁷ nng⁷ (11.713, SZ)
 皇帝 生 个 是 蛋
 emperor bear RL FM egg
 'What the emperor begot are eggs.'

- c. goa² si⁷ ma⁷ iah⁴-chhun¹, m⁷ kann² ma⁷ li² (11.606, SZ)
 我 是 罵 益春, 不 敢 罵 你
 I FM scold Yichuan not dare scold you
 ‘It is Yichun that I scolded, rather than you that I dare not scold.’

5. Syntactic positioning of topic and focus

There are various means of realizing information structure such as with stress, tone patterns, word order, and particular sentence patterns.⁸ Here we focus on the operation of word order and the use of function words. For sentences involving a main verb taking two nominal arguments, a pattern often encountered is the preposing to sentence-initial position of an object noun phrase as topic and the reliance on the focus marker *si*⁷ 是 to mark the subject noun phrase bearing new information, viz. the focused element, as in (8a-e).

- (8) a. kiann³ si⁷ i¹ boa⁵ si⁵ phah⁴ phoa³ ·liau² (17.134, WL)
 鏡 是 伊 磨 時 打 破 了
 mirror FM he polish strike break AM
 ‘It is he who broke the mirror while polishing it.’
- b. choe³ loo⁵ si⁷ i¹ kam¹-sim¹ cheng⁵-goan⁷ (24.096, WL)
 做 奴 是 伊 甘心 情願
 do slave FM he willingly
 ‘It is out of his own will that he is a slave.’
- c. chhat⁸-chhien⁷ pi⁷ si⁷ a¹-kong¹ a¹-ma² ma⁷ ·e (22.327, JJ)
 賊賤婢 是 啞公 啞媽 罵 个
 thief.mean maidservant FM grandpa grandma scold RL
 ‘It is only grandpa and grandma that can scold the despicable maid.’
- d. ho² khiap⁴ to¹ si⁷ li² choe³ ·e (14.282, JJ)
 好 怯 都 是 你 做 个
 good bad all FM you do RL
 ‘It is you who did it, be it good or bad.’

⁸ For information structure, see Teng (1979), Tang (1980), and Lee (2005) for treatment of both sentence examples. Cleft sentences and pseudo-cleft sentences are special syntactic devices reflecting patterns.

- e. chui² pian⁷ si⁷ iah⁴-chhun¹ chho³-chhiu² phoah⁴ tioh⁸ li² (22.159, JJ)
 水 便 是 益春 錯手 潑 著 你
 water just FM Yichun inadvertently splash AP you
 ‘It is Yichun that inadvertently splashed water on you.’

6. The internal makeup of *what* wh-words in Southern Min

What wh-words involve a construction consisting of ‘focus + class’ preceded by a focus marker *si*⁷ 是, derived from a copula. The focus as a variable is inevitably realized as *mih*⁸ 乜, whereas the class can be *mih*⁸ 物 for things, *tai*⁷ 事 for affair, and *lang*⁵ 人 for humans (< 儂 *nong⁵). A feature chosen from a class can be regarded as the pinning down of the subcategory of focus. In contrast to modern English where wh-words like *what* and *who* are portmanteau words, modern Southern Min, as an analytic language, features wh-words where focus and class are realized as separate words, as shown in the following table. Thus, *what* in English corresponds to *mih*⁸-*mih*⁸ 乜物, and *mih*⁸-*tai*⁷ 乜事, and *who* corresponds to *mih*⁸-*lang*⁵ 乜人.⁹

focus marker	focus	class	Southern Min
是 <i>si</i> ⁷ be	乜 <i>mih</i> ⁸ what	物 <i>mih</i> ⁸ thing	是乜物 <i>si</i> ⁷ - <i>mih</i> ⁸ - <i>mih</i> ⁸ what
是 <i>si</i> ⁷ be	乜 <i>mih</i> ⁸ what	事 <i>tai</i> ⁷ affair	是乜事 <i>si</i> ⁷ - <i>mih</i> ⁸ - <i>tai</i> ⁷ what
是 <i>si</i> ⁷ be	乜 <i>mih</i> ⁸ what	人 <i>lang</i> ⁵ humans	是乜人 <i>si</i> ⁷ - <i>mih</i> ⁸ - <i>lang</i> ⁵ who

7. Criteria for identifying *si*⁷ *mih*⁸ 是乜 as a syntagm or a lexicalized *what* wh-word

The sequence *si*⁷ *mih*⁸ 是乜 is open to two interpretations: (1) *si*⁷ 是 ‘be’ and *mih*⁸ 乜 ‘what’ are separate constituents forming a construction, or (2) *si*⁷ *mih*⁸ 是乜 as a whole denotes ‘what’ and the two elements have been reanalyzed as a grammatical function word, viz. *what* wh-word, where the phonetic shape *si*⁷ 是 has lost its meaning

⁹ The focus marker *si*⁷ 是 in *si*⁷ *mih*⁸ 是乜 can be incorporated and becomes part and parcel of the wh-word ultimately leading to the disyllabic wh-word *sim*²-*mih*⁸ 甚麼 corresponding to *shen-me* 甚麼 in Mandarin. Though 乜 *mih*⁸ and 物 *mih*⁸ share a common origin, 乜 as a focused element and *what* wh-word bears the upper entering tone, and 物 as a thing-denoting word features the lower entering tone (Douglas 1873). For a general discussion of the evolution of 是物, see Shimura (1984:169-226) and Oota (1988:124-141, 1991:88-104).

as well as its independent status. I take *si*⁷ 是 ‘be’ as a focus marker and *mih*⁸ 乜 as a focused element.

Since the *what* wh-word *mih*⁸ 乜 occupies an argument position, I shall focus on the behavior of *si*⁷ *mih*⁸ 是乜 in subject and object position and see if it occurs as a construction featuring focus marker + focus or simply as a *what* wh-word. The clue to the resolution of ambiguity lies in context which will help determine whether *mih*⁸ 乜 alone or *si*⁷ *mih*⁸ 是乜 as a whole is the focus. In the first case, 是乜 is a construction consisting of focus marker and focus, whereas in the second case, 是乜 is a lexicalized grammatical function word.

7.1 *Si*⁷ *mih*⁸ 是乜 in subject position

If a sentence featuring *si*⁷ *mih*⁸ 是乜 in subject position occurs out of context, it may be subject to ambiguous interpretation, as in (9a). That is, it is not clear whether *si*⁷ 是 + *mih*⁸ 乜 should be analyzed as a lexicalized unit or a sequence consisting of the focus marker *si*⁷ and the focused element *mih*⁸. But a question-answer pair will help resolve the ambiguity. (9a) and (9b) form such a pair. In response to the question in (9a), *si*⁷ reappears and functions as a focus marker in the answer (9b). It is immediately clear that *si*⁷ *mih*⁸ *lang*⁵ 是乜人 should be interpreted as *si*⁷ + *mih*⁸ *lang*⁵ where *si*⁷ is a focus marker rather than an incorporated element.

- (9) a. Q: *si*⁷ *mih*⁸ *lang*⁵ *kio*³ *mng*⁵ (29.013, WL)
 是 乜人 叫 門
 FM what-person call door
 ‘Who is calling at the door?’

- b. A: *hiau*²-*tit*⁴ *iau*¹ *si*⁷ *li*²-*po*⁵ *ti*⁷ *hu*² (29.015, WL)
 曉得 天 是 李婆 在 許
 know still FM li-po be there
 ‘I get it, it is Li the old lady who is there.’

I have found no case of *si*⁷ *mih*⁸ 是乜 in subject position that can be interpreted as a lexicalized *what* wh-word in context.

7.2 *Si*⁷ *mih*⁸ 是乜 in object position

When *si*⁷ *mih*⁸ 是乜 occurs in object position, *si*⁷ 是 may function as a focus marker or become lexicalized as part of the *what* wh-word. The resolution of this ambiguity hinges on the context in which it occurs. I first consider the case where *si*⁷ 是 is a focus

marker. In the following dialogue between the comic character (CM) and the secondary female character (SFC, viz. the maid Yichun), the repeated occurrences of *si*⁷ 是 show that it is a focus marker in that it is followed by *lang*⁵ 人, *mih*⁸ 乜 or *nng*⁷ 蛋.¹⁰

- (10) a. CM: *lang*⁵ *sinn*¹ *si*⁷ *lang*⁵, *hong*⁵-*te*³ *sinn*¹ *si*⁷ *mih*⁸ (11.711, SZ)
 人 生 是 人, 皇帝 生 是 乜¹¹
 human bear FM human emperor bear FM what
 ‘Humans give birth to humans. What does the emperor give birth to?’
- b. SFC: *sinn*¹ *e* *ia*⁷ *si*⁷ *lang*⁵, *cheng*¹ *cho*³ *thai*³-*chu*² (11.712, SZ)
 生 个 亦是 人, 稱 做 太子
 bear RL also human, call be prince-regent
 ‘What (he) begot is also a human called prince-regent.’
- c. CM: *m*⁷ *tioh*⁸, *hong*⁵-*te*³ *sinn*¹ *e* *si*⁷ *nng*⁷ (11.713, SZ)
 不 著, 皇帝 生 个 是 蛋
 not correct emperor bear RL FM egg
 ‘That’s wrong. What the emperor begot is an egg.’

By contrast, in the following dialogue between the buffoon and the runner, *si*⁷ *mih*⁸ 是乜 as a whole functions as *what* wh-word in focus. What is repeated in (11b) is *tiau*³ 吊 rather than *si*⁷ 是. Such a distribution shows that *si*⁷ 是 is not a focus marker in (11a). Rather, it has lost its independent status and has been reanalyzed as being a part of the *what* wh-word *si*⁷-*mih*⁸ 是乜 as a result of lexicalization.

- (11) a. Q: *tiau*³ *si*⁷-*mih*⁸ *lang*⁵ (40.144, WL)
 吊 是乜人
 hung what.person
 ‘Whose name tag was hung?’
- b. A: *tiau*³ *tan*⁵ *pek*⁴-*kheng*¹ *mia*⁵-*ji*⁷,
 吊 陳伯卿 名字,
 hung Chen Boqing name

¹⁰ The terms *comic character*, *secondary female character*, and *buffoon* as the renditions of *jing* 淨, *tie* 貼, and *chou* 丑 are due to Loon (1992:38).

¹¹ 乜 and 物 stand for *what*, a grammatical function word, and a thing-denoting lexeme respectively in the text, even though, as in their descendants in modern Southern Min, they mostly probably share the same phonological form.

khi²-ke² ngai⁵-chiu¹ ·khi³ (40.146, WL)
 起解 涯州 去
 dispatch Aizhou off
 ‘Chen’s name tag is hung. He is to be escorted to Aizhou.’

Apart from context, sentence processing may help resolve possible ambiguity. Take (12) for instance. The most plausible interpretation is that *u*⁷ 有 is a verb meaning ‘have’. Given such an interpretation, one would expect that what follows would be an NP, not a verb. If *si*⁷ 是 is construed as a focus marker, however, it would constitute a jarring situation hard to cope with. So the solution is to take *si*⁷-*mih*⁸ 是乜 as a lexicalized element.

- (12) *u*⁷ *si*⁷-*mih*⁸ *gun*⁵-*chinn*⁵ *lai*⁵ *pe*⁵ *goa*² (17.148, WL)
 有 是乜 錢銀 來 賠 我
 have what money-silver come repay me
 ‘What can you repay me with?’

I have shown above that context provides clues to resolve possible ambiguities. There seems to be two factors that explain why *si*⁷ *mih*⁸ 是乜 tends to be interpreted as a lexicalized *what* wh-word in object position. First, *si*⁷ *mih*⁸ 是乜 in the V-O construction is more amenable to an interpretation of being an NP rather than the syntagm featuring focus marker + focus where the focus marker *si*⁷ 是 cannot shed its verbal nature. Verb + *si*⁷ 是 as a succession of two verbs constitutes a sequence that is difficult to process. Second, Chinese, an inflection-less language, relies on word order as one of its strategies for signaling information status, and post-verbal position often indicates that a bare noun phrase carries new information. There is no need for a focus marker on the bare noun phrase that occupies object position.

Therefore object position, as opposed to subject position, tends to provide a better ground for *si*⁷ 是 to lose its independent status and be reanalyzed as a part of the lexicalized *what* wh-word *si*⁷ *mih*⁸ 是乜. The asymmetry manifested in the acceptability of *si*⁷ as a focus marker in subject position rather than object position shows that there is an important constraint on the position of focus marking in the hierarchical stacking of functional categories (Benincà & Poletto 2004).

8. Focus as opposed to topic

In terms of information structure, a sentence can be construed as consisting of two parts: focus and topic, standing for new and old information respectively (Lambrecht

1994, Erteschik-Shir 2007 among others). There are many devices used to implement information structure. Here we concentrate on the use of *si*⁷ 是 to mark the *what* wh-word as a focused element. Below, the focused element is in italics and the topic is unitalicized, as in (13a-b) and (14a-b). There is a mismatch between syntax and semantics in (14a-b), in that the focused element has taken on an adverbial function even though it originates as a noun phrase.

- (13) a. *chi*² *te*⁵ *si*⁷ *mih*⁸-*lang*⁵ *sai*² *li*² *phang*⁵ *lai*⁵ (19.175, JJ)
 只 茶 是 乜人 使 你 捧 來
 this tea FM what person order you to take come
 ‘Who had you bring me the tea?’
- b. *Iah*⁴-*chhun*¹, *hu*² *chhiunn*⁵ *goa*⁷ *si*⁷ *mih*⁸ *choh*⁴ *siann*¹ (20.063, WL)
 益春, 許 牆 外 是 乜 做 聲
 Yichun that wall outside FM what make sound
 ‘What was that sound outside the wall, Yichun?’
- (14) a. *si*⁷ *mih*⁸ *to*⁷-*li*² *m*⁷ *lai*⁵ *ak*⁴ *hoe*¹ (24.192, JJ)
 是 乜 道理 不 來 沃花
 FM what reason not come water the flower
 ‘Why won’t you water flowers?’
- b. *i*¹ *tann*¹ *choe*³ *phai*² *si*⁷ *mih*⁸ *sim*¹-*i*³ (21.034, JJ)
 伊 今 做 歹 是 乜 心意
 he now do bad FM what intention
 ‘Why is he behaving wickedly?’

Based on ample evidence adduced from African languages, Heine & Reh (1984: 109-110, 147-182) (see also Harris & Campbell 1995:151-168) proposes a three-stage development of focus structure: (1) the formation of cleft sentences where the part marked by a copula carries new information whereas the rest is presupposed, (2) the development from cleft structure to focus structure in which the copula becomes an optional focus marker, and (3) the further functional shift of the cleft structure in which the focus marker become obligatory. If we pursue the chronological profile of the Southern Min focus structure along these lines, we can see that the sentence featuring the focus marker + the focus *si*⁷ *mih*⁸ 是乜 developed from a cleft structure where the gap representing new information in the sentence is filled by the variable *mih*⁸ 乜 preceded by the copula *si*⁷ 是 and the rest is the presupposed part.¹² It is at this stage

¹² The presupposed element as a subordinate part is not necessarily marked by a relativizer or a

that the copula *si*⁷ 是 becomes a desemanticized focus marker. But the focus marker remains optional, as *mih*⁸ 乜 may also be unmarked or preceded by another focus marker *cho*³ 做 to form a *why/how* wh-word.¹³

9. Development of the syntagm 是物 into a *what* wh-word

In this section I shall address the issue of how 是物 evolved into a *what* wh-word in the context of the historical development of Chinese grammar. The periodization of historical development of Chinese can be undertaken in terms of grammatical, phonological and lexical criteria (Tai & Chan 1999). According to Oota (1988:3-105, 1991: 1-73) (cf. Lü 1985: preface) the development of Chinese grammar can be roughly divided into three major periods: Old Chinese (15th–1st cent. BC), Ancient Chinese (3rd–6th cent. AD) and Modern Chinese (7th cent. – present). Modern Chinese can be further distinguished into Early Modern Chinese (7th–14th cent.) and Late Modern Chinese (17th cent. – present).

I postulate the following three stages of development of the syntagm of 是物 into a lexicalized functional word and the eventual de-semanticization of 是 as a result of fusion:

- (1) 是物 > (2) 甚物 > (3) 是甚物

What follows is an elaboration of this three-stage development.

9.1 The first stage

Besides being a copula, *si*⁷ 是 is a focus marker that marks focused elements such as *mih*⁸ 物 ‘what’ and other non-interrogative focused elements as shown in §3.

乜, which shares with 物 a common etymological origin, is a demotic character. As a *what* wh-word, it is written as 麼 in Mandarin. The coexistence of 物 & 乜 in Southern Min and 麼 & 物 in Mandarin shows that the two graphic representations of the *what* wh-words, viz. 乜 and 麼, came about as a result of lexical split of the

particle. There is much debate as to whether *de* in the discontinuous construction *shi ... de* in Mandarin is a relativizer, nominalizer, or a particle. See Lee (2005) for a detailed survey of cleft sentences. There are scanty examples of *e*⁵ as a relativizer in the cleft sentences that occur in the texts we examine. So the subordination is not optionally signaled by the marker *e*⁵ which takes on a neutral tone indicated by · (a raised dot).

¹³ There are 177 tokens of *si*⁷ *mih*⁸ 是乜 and 185 tokens of *cho*³ *mih*⁸ 做/佐乜 as well as tokens of unmarked 乜 *mih*⁸ among the 1,159 tokens of constructions featuring 乜 *mih*⁸ in the four Southern Min texts.

lexeme 物. Various proposals have been advanced to motivate the development of the interrogative function of 是物. Shimura (1984:153-226) proposes that 是 in 是物 is an interrogative word, a cognate of the interrogative word 底. Lü (1985:128-130) claims that 是物 is a telescoped form of 是何物 and inherits its interrogative force from 何, an earlier interrogative word, which disappears later. Oota (1988:139-140, 1991:101-103) attributes the interrogative function of 是物 to the demonstrative 是 functioning as a universal quantifier. Jiang (1997:512) also shows that 是 has the interpretation of both *what* and *any*. Norman (1988), on cross-linguistic grounds, proposes that the *what* interpretation of 物 comes from its thing-denoting meaning. Jiang (1995) claims that the *what* meaning of 物 comes from its kind-denoting property. Whatever the theory, it is evident that 物 functions as a variable that can be realized as a *what* wh-word or as a universal quantifier.

Based on colloquial texts, Oota (1988:138, 1991:101) dates the occurrences of 是物 in the Earlier Modern Chinese to the first half of the eighth century during the Tang dynasty. It is safe to surmise that the 是乜 attested in the sixteenth century Southern Min text still preserved the syntactic status of 是物 as a syntagm; i.e. it had not yet been completely fused into a lexical item as a result of lexicalization.¹⁴ As shown towards the end of §8, *si*⁷是 at this stage functions as an optional rather than an obligatory element of the *what* wh-word.

9.2 The second stage

In this second stage *si*⁷ + *mih*⁸ as a sequence of focus marker + focus becomes reanalyzed as a *what* wh-word. In other words, a change from syntagm to lexeme takes place. This is accompanied by phonetic modification realized as fusion involving regressive segmental copying. Specifically, this involves the copying of the onset of the second syllable to the coda of the preceding syllable, i.e. *si-mih* > *sim-mih*, forming the geminate *-mm-*. At this point, *si*⁷是 in *sim-mih* has lost its independent status (Brinton & Traugott 2005); it has lost its relationship with other instances of *si*⁷. While *si*⁷ in other contexts is still written as 是, it is now written as 甚 in *sim-mih*.

/Sim/ has several subdialectal variants such as /siam/, /siann/, and /sann/ in modern Southern Min. The last two variants, featuring nasalization of vowels as a further development of the nasal finals, are sometimes written as 啥 rather than 甚. 甚 in its earlier pronunciation had a bilabial nasal coda (/m/), which evolved into /n/ in modern

¹⁴ The thing-denoting 物 'thing' as a lexeme and the *what*-word 乜 'what' as a grammatical function word parted company at this stage, as they were given two different graphic representations even though they still share the same phonetic shape, viz. /mih⁸/, in modern Southern Min (Douglas 1873:330).

Mandarin as a result of the merger of /-m/ and /-n/. It is therefore pronounced /shen/ in modern Mandarin. The use of the character of 啥 makes it evident that the psychological link has been lost between *sim*², the modern descendant of *si*⁷, and the newly developing focus marker. *Sim*² 甚 as a metamorphosized form of *si*⁷ 是 has become part and parcel of, or rather an obligatory element of, the lexicalized *what* wh-word *sim*²-*mih*⁸ 甚物 (=甚乜) ‘what’ (corresponding to *shen-me* 甚麼 ‘what’ in Mandarin). Another piece of evidence for the obligatoriness of *sim*² 甚 (< *si*⁷ 是) is that the syllable *mih*⁸ 乜 alone can no longer function as a *what* wh-word in this stage, whereas there are many instances of *mih*⁸ 乜 used by itself as a *what* wh-word in the first stage.

Oota (1988:138, 1991:101) puts the appearance of 甚 in the ninth century. This character did not appear in Southern Min texts until the second half of the 19th century, in a morality play published in 1886 (Wu 2005). Based on careful examination of the colloquial texts, Oota (1988:138, 1991:101) ascertains that it was not until the beginning of the Song dynasty, viz. the second half of the tenth century, that 甚麼 emerged. The Southern Min texts I examined indicate that 甚, together with 麼, did not appear until the second half of the 19th century (Wu 2005).

9.3 The third stage

At this stage it becomes possible to use the new focus marker *si*⁷ 是 to mark the *what* wh-word, if necessary, as in 是甚物.¹⁵ Note that in the first stage it was impossible to add a *si*⁷ 是 to *si*⁷ *mih*⁸ 是乜, since *si*⁷ 是 was a focus marker and still an independent element. It was only after *si*⁷ 是 and *mih*⁸ 物 merged into 甚麼 that “another” *si*⁷ 是 could be used to serve as a focus marker. Such a development, as shown in (15a-c), occurs in modern Taiwanese Southern Min.¹⁶

- (15) a. li² si⁷ siann²-mih⁸ lang⁵ (100.08, Da-an-xiang-III)¹⁷
 你 是 啥物 人
 you FM what person
 ‘Who are you?’

¹⁵ *Si*⁷ 是 and *mih*⁸ 物 underwent lexical split into 是/甚 and 物/麼, respectively. 甚 is sometimes written as 什.

¹⁶ The examples are due to a collection of folktales compiled by Hu (1998). Only this reference is given here due to space constraints.

¹⁷ *Sim*², which acquires -m through regressive assimilation and the nasal final -im, can be further nasalized. *Siann*²-*mih*⁸, written as 啥物, is an outcome of such further change.

- b. si^7 $siann^2$ - mih^8 mih^8 - a^2 $phah^4$ m^7 $kinn^3$ (108.04, Luo A-feng & Chen A-bing)
 是 啥物物仔 拍毋見
 FM what-SUF become-not-see
 ‘What is missing?’
- c. si^7 $siann^2$ - mih^8 tai^7 - chi^3 (126.14, Xin-she I)
 是 啥物 代誌
 FM what affair
 ‘What is it?’

The scenario discussed thus far is like Jespersen’s cycle used to account for the evolution of negation in some languages (Jespersen 1917, Roberts 2007). The introduction of a new focus marker in the third stage is preceded by the lexicalization of si^7 mih^8 into a *what* wh-word, which resulted in the loss of its link to si^7 ’s use in other contexts (Bauer 1992, Brinton & Traugott 2005).

The perceptive reader will see here that the pace at which a syntagm lexicalizes into a grammatical function word varies across languages. Southern Min in the 16th or the 17th centuries, as attested in these texts, seems to be a more conservative language than Late Modern Chinese during this same period.¹⁸

10. Conclusion

At the outset, si^7 是 was an amphibious word, in that it functioned both as a copula and as a focus marker. Using earlier Southern Min texts, I examined in detail (a) the syntactic categories and grammatical functions of the focused elements marked by si^7 是, and (b) the syntactic positioning of topic and focus in terms of information structure. From the perspective of language typology, I fleshed out the internal makeup of *what* wh-words. This revealed the analytic nature of Southern Min in the coding of syntactic and semantic features in comparison to English, a relatively synthetic language. Previous studies (Shimura 1984, Oota 1988, 1991) show that the *what* wh-word *shen-me* 甚麼 can be traced back to *shi-wu* 是物 in earlier Modern Chinese period. In light of these studies, I established that 是物 (written as 是乜 in the earlier Southern Min texts) was still preserved in Southern Min at least as late as the 16th century, although some instances of si^7 mih^8 是乜 had already begun to show signs of lexicalization. I put forward textual evidence showing that context and sentence processing can be relied on to identify

¹⁸ I refrain from using the term standard Chinese as the notion of standard language did not come into being until the early part of 20th century. I cannot pinpoint the geographical specificities of the language as attested in earlier modern colloquial texts either.

examples of *si*⁷-*mih*⁸ 是乜 which had become lexicalized *what* wh-words. From the examples of these lexicalized *what* wh-words where *si*⁷ 是 had not undergone phonetic modification into *sim*⁷ 甚, it is not wide of the mark to assume that lexicalization of *si*⁷ foreshadowed or prompted its phonetic modification. In short, it was the re-analysis of the syntagm *si*⁷ *mih*⁸ 是乜 that brought about the fusion of *si*⁷ 是 into the *what* wh-word, yielding 甚物 at a later stage, not the other way around. It was also shown that once *si*⁷ 是 lost its independent status and was desemanticized, the newly developing *what* wh-word 啥物 could be focused with a preceding focus marker 是, i.e. 是啥物. So a new cycle in the sense of Jespersen (1917) is underway.¹⁹

References

- Bauer, Laurie. 1992. Lexicalization and level ordering. *Linguistics* 30.3:561-568.
- Benincà, Paola, and Cecilia Poletto. 2004. Topic, focus and V2: defining the CP sublayers. *The Structure of CP and IP*, Vol. 2: *The Cartography of Syntactic Structures*, ed. by Luigi Rizzi, 52-114. Oxford & New York: Oxford University Press.
- Brinton, Laurel J., and Elizabeth Closs Traugott. 2005. *Lexicalization and Language Change*. Cambridge & New York: Cambridge University Press.
- Cann, Ronnie. 2007. Towards a dynamic account of BE in English. *Existence: Semantics and Syntax*, ed. by Ileana Comorovski & Klaus von Heusinger, 13-48. Dordrecht & London: Springer.
- Douglas, Rev. Carstairs. 1873. *Chinese-English Dictionary of the Vernacular or Spoken Language of Amoy with the Principal Variations of the Chang-chew and Chin-chew Dialects*. London: Trübner.
- Erteschik-Shir, Nomi. 1997. *The Dynamics of Focus Structure*. Cambridge & New York: Cambridge University Press.
- Erteschik-Shir, Nomi. 2007. *Information Structure: The Syntax-Discourse Interface*. Oxford & New York: Oxford University Press.
- Harris, Alice C., and Lyle Campbell. 1995. *Historical Syntax in Cross-linguistic Perspective*. Cambridge & New York: Cambridge University Press.
- Heine, Bernd, and Mechthild Reh. 1984. *Grammaticalization and Reanalysis in African Languages*. Hamburg: Helmut Buske.

¹⁹ It should be noted that when *si*⁷ is fused into the *what* wh-word, this does not mean that the instances of *si*⁷ in the capacity of copula and focus marker have disappeared. What has happened is that only the focus marker *si*⁷ in the syntagm *si*⁷ *mih*⁸ has been fused. The other functions of *si*⁷ have also remained robust.

- Hopper, Paul J. 1991. On some principles of grammaticalization. *Approaches to Grammaticalization*, Vol. 1: *Focus on Theoretical and Methodological Issues*, ed. by Elizabeth Closs Traugott & Bernd Heine, 17-35. Amsterdam & Philadelphia: John Benjamins.
- Hopper, Paul J., and Elizabeth Closs Traugott. 2003. *Grammaticalization* (2nd edition). Cambridge & New York: Cambridge University Press.
- Hu, Wanchuan. (ed.) 1998. *Da-an Xiang Minnanyu Gushi Ji* [*A Collection of Southern Min (Da-an) Folktales*]. Taichung: Taichung County Cultural Center.
- Jespersen, Otto. 1917. *Negation in English and Other Languages*. Copenhagen: A. F. Høst.
- Jiang, Lansheng. 1995. Shuo ma yu men tongyuan [On the etymological cognates *ma* and *men*]. *Zhongguo Yuwen* 1995.3:180-190.
- Jiang, Lihong. 1997. *Dunhuang Bianwen Wenzi Tongshi* [*An Etymological Concordance of Words in Dunhuang Bianwen Texts*]. Shanghai: Shanghai Chinese Classics Publishing House.
- Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representation of Discourse Referents*. Cambridge & New York: Cambridge University Press.
- Lee, Hui-chi. 2005. *On Chinese Focus and Cleft Constructions*. Hsinchu: National Tsing Hua University dissertation.
- Lehmann, Christian. 1995. *Thoughts on Grammaticalization* (revised and expanded version). München: LINCOM Europa.
- Loon, Piet Van der. 1992. *The Classical Theatre and Art Song of South Fukien: A Study of Three Ming Anthologies*. Taipei: SMC Publishing Inc.
- Lü, Shuxiang. 1985. *Jindai Hanyu Zhidaici* [*Demonstratives and Pronouns in Early Modern Chinese*]. Shanghai: Xuelin Press.
- Norman, Jerry. 1988. *Chinese*. Cambridge & New York: Cambridge University Press.
- Oota, Tatsuo. 1988. *Chuugoku Goshi Tsuukoo* [*A Historical Study of Chinese Language*]. Tokyo: Hakuteisha.
- Oota, Tatsuo. 1991. *Hanyushi Tongkao* [*A Historical Study of Chinese Language*]. Trans. by Lansheng Jiang & Weiguo Bai. Chongqing: Chongqing Press.
- Roberts, Ian. 2007. *Diachronic Syntax*. Oxford & New York: Oxford University Press.
- Shimura, Ryooji. 1984. *Chuugoku Chuusei Gohoo Shi Kenkyuu* [*A Study of the History of Middle Chinese Grammar*]. Tokyo: Santoosha.
- Tai, James H-Y., and Marjorie K.-M. Chan. 1999. Some reflections on the periodization of the Chinese language. *Studies on Chinese Historical Syntax and Morphology: Linguistic Essays in Honor of Mei Tsu-lin*, ed. by Alain Peyraube & Chaofen Sun, 223-239. Paris: École des Hautes Études en Sciences Sociales, Centre de Recherches Linguistiques sur l'Asie Orientale.

- Tang, Ting-chi. 1980. Guoyu fenlieju, fenlie bianju, zhun fenlieju de jiegou yu xianzhi zhi yanjiu [Cleft and pseudo-cleft constructions in Chinese: structure, function and constraint]. *Journal of National Taiwan Normal University* 25:249-296.
- Teng, Shou-hsin. 1979. Remarks on cleft sentences in Chinese. *Journal of Chinese Linguistics* 7.1:101-113.
- Traugott, Elizabeth Closs, and Bernd Heine. (eds.) 1991. *Approaches to Grammaticalization*, Vol. 1: *Focus on Theoretical and Methodological Issues*. Amsterdam & Philadelphia: John Benjamins.
- Wu, Shouli. 2001a. *Ming Jiajing Kan Li Jing Ji Xiwen Jiaoli* [Annotated Texts of the Romance of Li Jing Ji of Ming Jiajing Edition (JJ)]. Taipei: Ts'ung-I Workshop.
- Wu, Shouli. 2001b. *Ming Wanli Kan Li Zhi Ji Xiwen Jiaoli* [Annotated Texts of the Romance of Li Zhi Ji of Ming Wanli Edition (WL)]. Taipei: Ts'ung-I Workshop.
- Wu, Shouli. 2001c. *Qing Shunzhi Kan Li Zhi Ji Xiwen Jiaoli* [Annotated Texts of the Romance of Li Zhi Ji of Qing Shunzhi Edition (SZ)]. Taipei: Ts'ung-I Workshop.
- Wu, Shouli. 2001d. *Qing Guangxu Kan Li Zhi Ji Xiwen Jiaoli* [Annotated Texts of the Romance of Li Zhi Ji of Qing Guangxu Edition (GX)]. Taipei: Ts'ung-I Workshop.
- Wu, Shouli. 2005. *Qing Guangxu Xin Kan Xuanjiang Xiwen Jiaoli* [Annotated Texts of the Morality Play of Qing Guangxu New Edition]. Taipei: Ts'ung-I Workshop.

[Received 15 August 2008; revised 3 February 2009; accepted 6 May 2009]

Graduate Institute of Linguistics
National Tsing Hua University
101, Sec. 2, Kuang-fu Road
Hsinchu 300, Taiwan
cflien@mx.nthu.edu.tw

早期閩南語文獻的焦點標記“是”： 兼論“是乜”之詞彙化爲“甚麼”疑問詞

連金發

國立清華大學

本文探討早期閩南語文獻中焦點標記“是”，兼論構式“是乜”之詞彙化爲“甚麼”疑問詞。“是”兼具繫詞和焦點標記兩種角色。“是”除了充當一系列詞類的焦點標記外還可以標明主語或賓語爲句子的焦點，承載新信息。基於信息結構的要求“是”標明焦點，而主題是前移的成分，反映舊信息。“是”不是焦點標記就是已經失去獨立性，併入詞彙化的“甚麼”疑問詞中，失去原有的語意。本文根據早期閩南語文獻的證據提出“是乜”三階段的演變說並與近古漢語的“是物”的演變進行比較，閩南語的演變顯然較具存古性。

關鍵詞：焦點標記，詞彙化，信息結構，早期閩南語