國立台灣大學文學院語言學研究所

碩士論文

Graduated Institute of Linguistics

College of Liberal Arts

National Taiwan University

Master Thesis

霧台魯凱語言談中的指稱表達式研究

Referential Expressions in Discourse of Budai Rukai

吳妤蓁

Yu-Chen Wu

指導教授: 宋麗梅博士

Advisor: Li-May Sung, Ph. D.

中華民國 112 年 7 月

July, 2023

doi:10.6342/NTU202302199

謝辭

完成這本論文並不容易，過程中曾經以為永遠走不到終點。好在身邊有群溫暖的人們，給予不斷地支持與鼓勵，讓我即便跌跌宕宕，仍然可以堅持著不放棄。

由衷感謝我的指導教授宋麗梅老師，帶領我進入了南島語的世界，在我苦尋論文題目之時不厭其煩地指引我方向、理解我的困難，給我足夠的信任與空間，一次又一次與我討論語料細節，終於促成這本論文的誕生。感謝我的兩位口試委員吳靜蘭與洪媽益教授，在口試過程中提出的寶貴建議，讓整本論文可以臻於完善。謝謝我的霧台魯凱語老師：柯菊華老師、巴山光老師以及在霧台山上遇到的發音人們，你們傳承古老語言的熱情，賦予這本論文更多意義。

我會懷念與南島語研究室夥伴們(力行、凱弘與俊宏)，在好幾個早晨及午後，窩在語言中心一同鑽研南島語料的日子，我們相互砥礪成長，也共同各種分享生活軼事。也很幸運能夠遇見 R08 優秀的同儕：容榕、凱弘、俊宏、晴方、莊勻、峻維、永賦、飛揚與庭瑋，我們一同組成讀書會、上山田調以及舉辦聖誕晚會，過程中大家總是互相照顧，努力盡責，讓每件事最終得以盡善盡美。

誤打誤撞踏入台大語言所的大門，在這裡追逐理想的日子有眼淚有歡笑，最後，我感謝我的家人以及總是陪伴在側的每位摯友們，毫不吝嗇地支持我生活與心靈所需，讓我在每一次的跌倒後都能站起，終能收穫滿載地畢業。

doi:10.6342/NTU202302199

中文摘要

在言談中，同一個指涉對象 (referent)可能會以不同的語言形式 (syntatic form)出現。而影響言談中的說話者挑選不同語言形式的關鍵究竟為何，也成為近年來認知與功能語言學領域研究的一大重點。

此篇論文將透過兩個主要的研究方法:分別為 Givón (1983)的 Theory of Topic Continuity 以及 Gundel (1980)的 Givenness Hierarchy，來分析霧台魯凱語言談中不同的語言形式。

正如同其他許多不同的語言一樣，在魯凱語當中，語言外觀形式的複雜性以及說話者的認知狀態都會影響到指稱表達式(referential expression)的挑選。其中，我們發現四種主要的指稱表達式: 零指示詞(zero anaphora)、代名詞(pronoun)、定名詞(definite noun) 以及不定名詞(indefinite noun)，都符合 Givón 所提出的 Topic Continuity Scale，亦即零指示詞具有最高的主題連續性，代名詞次高，定名詞次低，而不定名詞為最低。四者當中，擁有較高主題連續性的零指示詞以及代名詞通常會用來指涉在言談中較為重要的主題，相反地，定名詞 以及不定名詞具有較低的主題連續性，指涉之主題也相對較不重要。

此外，在霧台魯凱語中我們也發現，當指涉對象存在於說話者不同的認知狀態當中時，其指稱表達式也會跟著不同，這也符合 Gundel 所提出的 Givenness Hierarchy 之假設。具體來說，當一個指稱對象處在 In Focus 的認知狀態當中，其通常會以零指示詞或代名詞的樣態出現。另一方面，處在 Referential 或 Type Indentifiable 等認知狀態的指稱對象則通常會以不定名詞的樣態呈現。值得一提的是，定名詞在幾乎所有認知狀態中的指稱對象上都可以看見；而尤其當指稱對象處在 Activated、Familiar 或 Uniquely Identifiable 這三種認知狀態時，基本上都是以定名詞的樣態展現。

關鍵字：指稱表達式、主題連續性、已知性結構、霧台魯凱

doi:10.6342/NTU202302199

Abstract

In the discourse, the same referent can be encoded by various kinds of syntactic forms. This current thesis attempts to provide some analysis on syntactic coding devices of NPs in the discourse of Budai Rukai. Two major approaches, Givón's theory of topic continuity (1983) and Gundel's hypothesis of the Givenness Hierarchy (1980), are mainly applied to our analysis.

The same as many other langueges, both the quantity of the syntactic coding devices and the cognitive status of a certain referent play a key role in the referential choice in Budai Rukai. The four major types of syntactic coding devices, zero anaphora, pronoun, definite noun, and indefinite noun, are found to follow the Givón's scale of topic continuity, with zero anaphora having the highest degree of topic continuity, pronouns the second highest, definite NP the second lowest, and indefinite NP having the lowest one. The two most continuous NP coding devices, zero anaphora and pronoun are often used to refer to the more important referents in the discourse, while the two least continuous NP coding devices, definite noun and indefinite noun, tend to refer to relatively unimportant topics.

Also, we found that referents in different cognitive statuses are coded differently in Budai Rukai, consisting with the Gundel's Givenness Hierarchy. A referent in the status of *in focus* can be coded as zero anaphora and pronoun. On the other hand, *referential* and *type indentifiable* refernts are in the grammatical device of indefinite noun. Interestingly, definite noun can be seen to code referents in almost every cognitive status except for *type identifiable*. Furthermore, the *activated, familiar* and *uniquely identifiable* referents are all primarily expressed in the form of definite noun.

Keywords: noun phrase, referential expression, topic continuity, the Givenness Hierarchy, Budai Rukai, Formosan languages

doi:10.6342/NTU202302199

Contents

doi:10.6342/NTU202302199

doi:10.6342/NTU202302199

List of Abbreviations

1S	First person singular
1P	First person plural
2S	Second person singular
2P	Second person plural
3S	Third person singular
3P	Third person plural
ACT	Active voice
BN	Bound Nominative
CAUS	Causative
CONJ	Conjunctor
FIL	Filler
FS	False start
FUT	Future tense
GEN	Genitive case
INV	Invisible
IRR	Irrealis
LNK	Linker
LOC	Locative
NEG	Negator
NMLZ	Nominalizer
NOM	Nominative case
OBL	Oblique case
P	Plural
PASS	Passive voice
PFV	Perfective
POSS	Possessive case
PN	Proper noun
PROG	Progressive
RED	Reduplication
REL	Relativiser
RLS	Realis
REFL	Reflexive
SG	Singular
STAT	Stative
TOP	Topic marker
VIS	Visible

doi:10.6342/NTU202302199

List of Figures and Tables

doi:10.6342/NTU202302199

Chapter 1

Introduction

1.1 Motivation and Research Questions

Noun phrases (NPs) referring to exactly the same entity may be encoded by a number of different syntactic forms. For instance, in English, a particular boy that exists in the real physical world can be referred to as *a boy, the boy, that boy, this boy, he,* or even *elided*. Such referential expressions have been investigated by a variety of approaches, inclusive of neo-Gricean pragmatic approaches (Huang 1991, 1994; Levinson 1983, 1987a, 1987b, 1991), cognitive approaches (Ariel 1988, 1990; Chafe 1979, 1987; Gundel 1980, 1988, 1995; Gundel et al. 1988, 1989, 1990, 1993; Tomlin and Pu 1991), linear approach (Givón 1983a, 1983b), and hierarchical approach (Fox 1987a, 1987b).

However, those studies mentioned above put their focus mainly on Indo-European or other renowned languages. Relatively fewer studies investigate this research topic in Austronesian languages, especially from a cognitive perspective, let alone Formosan languages. As a result, this thesis aims to fill this gap. By applying two cognitive approaches (Givón 1983; Gundel et al. 1993) to the examination of the referential expressions in Budai Rukai, one of the main offshoots of Proto Austronesian languages as well as one of the sixteen Formosan languages, this thesis tries to deal with the following research questions:

(1) **Research Question One:** What factors play essential roles in the selection of syntactic coding devices on NPs in discourse of Budai Rukai?

(2) **Research Question Two:** Does the syntactic coding system in Budai Rukai follow Givón's scale of topic continuity?

(3) **Research Question Three:** What is the relation between the cognitive status of a referent and the referential form in Budai Rukai, and does it align with Gundel's

doi:10.6342/NTU202302199

Givenness Hierarchy of NPs?

1.2 Introduction to Budai Rukai

Rukai is one member of the Austronesian language family in Taiwan, which is also known as Formosan languages. Most of the Rukai tribes are located on both sides of the south of the Central Mountain Range southern Taiwan, ranging from Kaohsiung City, Pintung County, and Taitung County. Figure 1.1 below shows the geographical distribution of Rukai language. Currently, the Rukai population is reported to be around 13,465 people in the 2020 survey conducted by the Council of Indigenous Peoples, Executive Yuan.

Figure 1.1 Geographic Distribution of Rukai (Wang 2003: 1)

There are six dialects spoken by the Rukai people distributed in different areas, including Budai, Tanaa, Labuan, Mantauran, Maga, and Tona (Li 1973, Wang 2003, and Zeitoun 2003, 2007). Maga, Mantauran, and Tona are spoken in Maolin District, in the south of Kaohsiung City. Budai and Labuan are situated in Wutai Township in the north of Pingtung County. Tanan is located in Peinan Township, in the west of Taitung County. The geographical distribution of six Rukai dialects is illustrated in Figure 1.2

doi:10.6342/NTU202302199

below.

Figure 1.2 Geographical Distribution of Rukai (Wang 2003; Zeitoun 2007)

Budai has a large number of linguistic similarities with Labuan and Tana, while it is a lot different from Maga, Mantauran, and Tona. However, still some lexical, phonological, as well as syntactic, characteristics are found shared among these six dialects of Rukai languages (Li 1977, Zeitoun 2000, 2018).

Rukai languages are found possessing several unique linguistic properties, causing its controversial status in the Austronesian family for long (Ferrell, 1969; Tsuchida, 1976; Li, 1977; Dahl, 1981; Ho, 1983; Starosta, 1995; Li, 1996; 1997a; Blust, 1999). As a result of the particular voice system and other special grammatical properties of Rukai, Starosta (1995) argued it to be the first offshoot in the Proto-Austronesian subgrouping. Figure 1.3 shows the subgrouping:

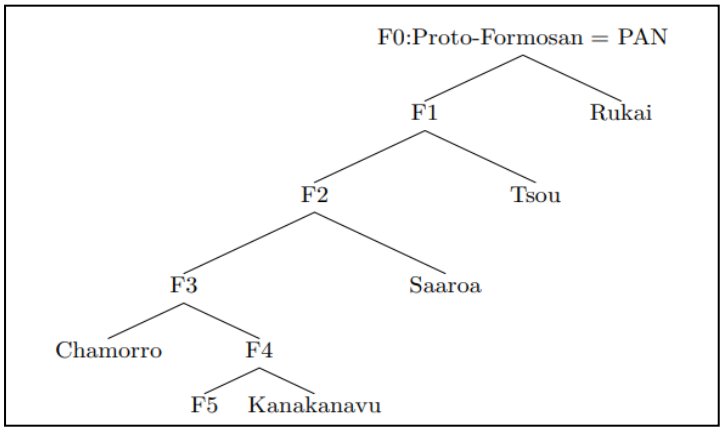


Figure 1.3 Proto-Austronesian subgrouping (Starosta 1995)

doi:10.6342/NTU202302199

Opposed to Starosta's subgrouping, Blust (1999) argued that the Austronesian subgrouping can be presented with nine Formosan branches as well as one Malayo-Polynesian branch, and Rukai stands as one of the nine main subgroups of Formosan language, as illustrated in Figure 1.4:

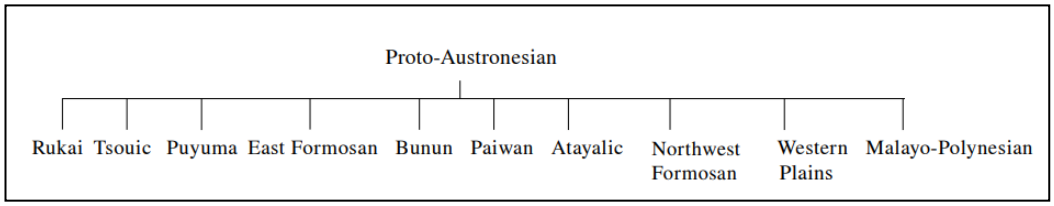


Figure 1.4 Blust's (1999) subgrouping (Chen 2017: 4)

Ross (2009), on the other hand, proposed that languages, such as Puyuma, Tsou, and Rukai should be split out from the rest of the Austronesian languages called Proto-Nuclear Austronesian languages (PNAn), which are all found undergoing the process of reanalyzing nominalizations into verbs. In Ross's subgrouping (2009), another major difference from Blust's (1999) is that Kanakanavu and Saaroa are removed from the Tsouic group and regrouped into PNAn because of their verbal affixes. Ross's subgrouping is indicated as below in Figure 1.5:

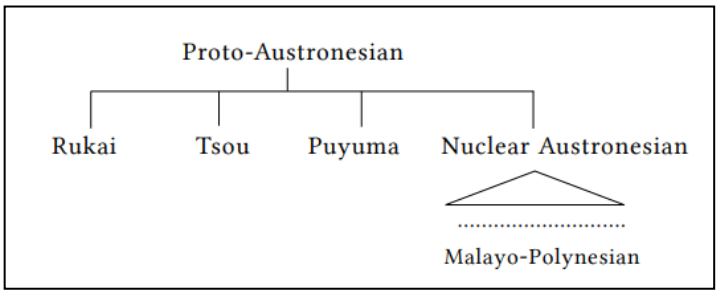
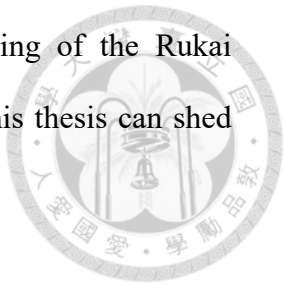


Figure 1.5 Ross's (2009) subgrouping (Chen 2017:202)

Many scholars have revised Ross's subgrouping and further given a finer subgrouping of Austronesian languages; however, there is no conclusion of the status

doi:10.6342/NTU202302199

of Rukai in the Austronesian languages. Although the subgrouping of the Rukai language is not the primary focus of this thesis, it is a hope that this thesis can shed some light on the understanding of this issue.



1.3 Grammar Sketch of Budai Rukai

For the ease of our readers' reference, we offer theem with a quick grammar sketch of Budai Rukai in the following section.

1.3.1 Basic Clause Structure

Just like in most Formosan languages, one sentence in Budai Rukai can be divided into two main parts, that is "subject" and "predicate". The subject is often nominal, denoting old information while the predicate can be either nominal or verbal, referring to new information. On top of that, Budai Rukai is characterized as a typical predicate-initial language, as shown in (1), in which the predicate is always seen followed by the subject.

(1) **a. Action Verbs as Predicate:**

[*w-a-udukri=nga*]Predicate	[*ka*	*Elrenge*]Subject
ACT-RLS-dance=PFV	NOM	Elrenge

'Elrenge danced'

b. Stative Verbs as Predicate:

[*ma-buruku*]Predicate	[*ka*	*daane*]Subject
STAT.RLS-crumble	NOM	house

'The house collapsed.'

c. Nouns as Predicate

[*ka*	*Lavausu*]Predicate	[*agi=li*]Subject
NOM	Lavausu	young.sibling=1S.BG

'Lavausu is my sister.'

doi:10.6342/NTU202302199

On the other hand, if there are more than one argument in a sentence, the word order is relatively flexible. In other words, both Verb-Subject-Object (VSO) and Verb-Object-Subject (VOS) are grammatical word order as shown in (2a) and (2b). It is noted that VOS word order is relatively preferred.

(2) **a. <u>VOS Word Order</u>**

w-a-punpungu	[*ki*	*Takanaw*]Object	[*ka*	*vali-ane*]Subject
ACT-RLS-hit	OBL	Takanaw	NOM	tooth-NMLZ

b. <u>VSO Word Order</u>

w-a-punpungu	[*ka*	*vali-ane*]Subject	[*ki*	*Takanaw*]Object
ACT-RLS-hit	NOM	tooth-NMLZ	OBL	Takanaw

'The boar hit Takanaw.'

As we can see in the (2a) and (2b), the position of the object *ki Takanaw* and the subject *ka valiane* are exchangeable. Besides the order of VOS and VSO, the order of SVO, in which the subject is in the sentence-initial position, is also allowed in Budai Rukai. Such sentence with a fronted subject is the so-called "topicalization" (Zeitoun 2000). The mechanism of topicalization in Budai is mainly used to indicate given information while the following predicate provides new information. One typical example of topicalization in Budai Rukai is given below in (3).

(3)

[*ku*	*ama*]TOP	*w-a-rubu*	*ku*	*angatu*
NOM	father	ACT-RLS-collect	OBL	tree

'My father, (he) collected trees.'

In (3), the preposed subject *ama* "father" in the sentence-initial position might have been known by the speaker and the hearer, and the speaker topicalizes it to bring

doi:10.6342/NTU202302199

it back to their current attention. Table 1.1 summarizes the possible types of word order in Rukai basic sentence structure.

Table 1.1: Word Order Variation of Rukai Simple Sentences (Chen 2008)

Initial	Internal	Final
V	O	S
V	S	O
S	V	O

1.3.2 Voice System

Different from most Formosan languages that generally have four grammatical voices, Budai Rukai is said to have a two-way active-passive voice distinction (Li 1977; Kuo 1979; Starosta 1995; Chen 1999; Zeitoun 2007; Chen 2008), as illustrated in (4) and (5).

(4) **Active Voice Structure:**

a. *kavay* *valis-ane* *wapunpungu* *ki* *Takanaw*
that.VIS tooth-NMLZ ACT-RLS-hit OBL Takanaw
'That boar hit Takanaw.'

b. *Ø-si<a>ludu* *ku* *vasaw* *ka* *ama*
ACT-<RLS>pick.up OBL leaf NOM father
'My father picked up the leaf.'

(5) **Passive Voice Structure:**

kavay *Takanaw* *ki-a-punpungu* *kavay* *ki* *valis-ane*
that.VIS Takanaw PASS-RLS-hit that.VIS OBL tooth-NMLZ
'That Takanaw was hit by that boar.'

doi:10.6342/NTU202302199

Based on examples (4) and (5), the voice of a sentence can be detected by the affixes attached to the verb. For instance, in active sentence structures like (4), the verbal predicates *punpungu* "hit" and *siludu* "pick up" are attached by either the prefix *w-* (4a) or the zero form *Ø-* (4b). On the other hand, if one sentence is in the passive voice like (5), its main verb *punpungu* "hit" takes the passive prefix *ki-*. Finally, Table 1.2 can summarize the voice marking system in Budai Rukai.

Table 1.2 The Voice System of Budai Rukai

	Active Voice	**Passive voice**
Marking	*w-, Ø-*	*ki-*

1.3.3 The Case Marking System and Demonstratives

A three-way case distinction among Nominative, marking a subject argument, Genitive, marking a possessor, and Oblique, marking a non-subject, is found in Budai Rukai (Li 1977; Chen 2008; Sung 2011; and many others). There are three case markers *ka, ku*, and *ki* in Budai language, and they are obligatorily present in front of noun phrases. The case markers *ka* and *ku* can be utilized to mark both nominative and oblique nominals, while *ki* is used to mark both oblique and genitive nominals. In addition, these case markers can be further categorized in terms of several semantic factors, including visibility, distance, animacy, humanness, and specificity. Table 1.3 summarizes the case marking system in Budai Rukai (Li 1977; Chen 1999 2008; Zeitoun 2000; Sung 2011, 2015).

doi:10.6342/NTU202302199

Table 1.3 Case Marking System in Budai Rukai (Sung 2011)

Nominative	Oblique	Genitive
ku (-visible, +distance, ±animate)	*ku* (-visible, +distance, -human, ±generic)	
ka (+visible, -distance, ±animate)	*ka* (+visible, -distance, -human)	
	ki (+specific, +human) (+generic, -human)	*ki* (±animate)

According to Table 1.8, the case markers *ka* and *ku* can be distinguished from each other by the features [±visible] and [±distance]. Looking at the following two examples:

(6) a. *ma-tuase* *ka* *Kui*
STAT.RLS-leave NOM Kui
'Kui left.'

b. *ma-tuase=nga* *ku* *Kui*
STAT.RLS-leave=PFV NOM Kui
'Kui has left.'

Although examples (6a) and (6c) look quite similar, there is a major difference between these two sentences. That is, *Kui* in (6a) can still be seen by the speaker while that in (6b) cannot.

Besides nominal case markers, demonstratives in Budai Rukai can also be distinguished in terms of visibility and distance to the speaker (Sung 2011). Table1.4 below summarizes demonstratives in Budai Rukai.

doi:10.6342/NTU202302199

Table1.4 Demonstratives in Budai Rukai

	+Visible	-Visible
+Distance	*kavay* “that”	-
-Distance	*kikay/kay* “this” *kuini/kui* “that”	*kudra* “that”

Demonstratives often cooccur with case markers, but sometimes in natural discourse, case markers are omitted so that the noun phrases are coded only by demonstratives (Shih, 2012).

1.3.4 The Pronominal System

There are case distinctions in the pronominal system in Budai Rukai, in which the nominative, genitive, oblique, and topic cases are clearly distinguished. On top of that, pronouns in Budai Rukai can be found in two types, comprising free forms and bound forms, as shown in Table 1.5 (Chen 1999, Zeitoun 2000,Sung 2011). Among the four cases in the pronominal system, the nominative and genitive cases are bound pronouns, in the forms of clitics, whereas the oblique and topic cases are free pronouns. It is noteworthy that both singular and plural third-person nominative pronouns are covert in Budai Rukai.

doi:10.6342/NTU202302199

Table1.5 Pronominal system in Budai Rukai (Sung 2011)

Person	Plurality	Visible/ Inclusive	Free		Bound	
			TOP	OBL	NOM	GEN
1	Singular		*ku aku*	*nakuane*	*=(c)aku, =naw*	*=li*
	Plura	±Inclusive	*ku ta*	*mitaane*	*=ta*	*=(i)ta*
		±Exclusive	*ku nay*	*nayane*	*=nay*	*=nay*
2	Singular		*ku su*	*musuane*	*=su*	*=su*
	Plura		*ku numi*	*numiane*	*=numi*	*=numi*
3	Singular		*ku ini*	*iniane*	∅	*=ini*
	Plura		*ku ini*	*liniane*	∅	*=lini*

1.4 Data Collection

The Budai Rukai data in this thesis consist of eleven narratives, primarily from the National Taiwan University Corpus of Formosan Languages (Sung et al. 2008; Su et al. 2008[1]). Nine of them are story-telling, and the other two are the traditional Rukai culture sharing. The details of the narrators and the texts are displayed in Table 1.6:

Table 1.6: List of native speakers and texts recorded

Name of Text	Genre	Speaker	Gender	Year of Birth
Millet	Narrative	Tagas	M	1938
Childhood	Narrative	Balenge	F	1961
Frog Story	Narrative	Kainguane	M	unknown
Frog Story	Narrative	Legeai	M	unknown
Frog Story	Narrative	Salrabu	M	1938
Pear story	Narrative	Ba Ching i	M	unknown
Pear story	Narrative	Balenge	F	1961

[1] NTU Corpus of Formosan Languages (台大台灣南島語語料庫): https://formcorp.netlify.app/#/

doi:10.6342/NTU202302199

Pear story	Narrative	Legeai	M	unknown
Pear story	Narrative	Salrabu	M	1938
Pear story	Narrative	Tagas	M	1938
Pear story	Narrative	Wauki	M	1933

1.5 Organization of the Thesis

The present thesis includes five chapters. Chapter Two reviews two major theories addressing the selection of referential forms in discourse as well as some previous studies on the syntactic coding devices in Budai Rukai.

Chapter Three checks on the topic continuity of syntactic coding system in Budai Rukai with Givón's hypothesis of topic continuity.

Chapter Four tries to figure out the relation between cognitive statuses and syntactic coding devices in Budai Rukai by the use of Gundel's theory of referential givenness.

Finally, in Chapter Five, the main findings of our three research questions are given, and several possible directions for future research are provided as well.

doi:10.6342/NTU202302199

Chapter 2
Literature Review

This chapter provides a broad introduction to the notions and concepts that are required for our investigation of the NPs in discourse of Budai Rukai. The whole chapter will be divided into two sections. Section 2.1 introduces two major theories dealing with how referential expressions interact with the cognitive statuses of the speakers and the hearers. Section 2.2 reviews some discussions on the syntactic coding devices in Budai Rukai, which were proposed by previous studies. And Section 2.3 explains how the reviewed studies imply the present study.

2.1 Referential Expressions and Cognitive/Discourse Statuses

It is proposed by previous studies (Givón 1983, Gundel et al. 1993, among others) that when more than one referential expression qualifies for the same cognitive status, referential choice is not decided randomly. Instead, the selection of the referential expression can be determined by several cognitive factors. The following subsections is the overview of some significant literature concerning such issue.

2.1.1 Givón (1983)

Givón assumed that a sentence, as a basic information processing unit in discourse, isn't produced randomly. Instead, sentences are connected with each other by topic chains so that the discourse can be coherent. In other words, a coherent discourse is likely to maintain similar topic over a span of sentences. In addition, both previous and subsequent contexts can have some influences on the selection of a referential expression, which is claimed to be a result of the interaction between memory and attention of the speaker.

To explain the phenomenon above, Givón (1983, 1992, 1994, 2017) further

doi:10.6342/NTU202302199

proposed the framework of topic continuity, which can be used to measure the continuity of noun phrases in connected discourse. Topic continuity is concerned with how the speaker instructs the hearer to mentally search for the topic in his/her episodic memory of the current text and to activate important topics but not unimportant ones (Givón 2017). Topic continuity can be measured by calculating the number of sentences between the current and the last mention of a referent in the previous context (anaphoric distance; hereinafter AD) and also by computing the number of times that such referent recurs after its present appearance (cataphoric persistence; hereinafter CP) (Givón 1994, 2017). If a referent have a lower value of AD and a higher value of CP, it is claimed to be more topically continuous, and thus more accessible and more important in the discourse. Furthermore, the degree of accessibility and importance can be coded by various grammatical devices which universally hierarchize along the continuum in (1). According to the Ironic quantity principles (Givón 1991), the less predictable information and more important information will be given more coding materials. In short, the quantity of syntactic codes plays a role in determining the order of the grammatical devices along the continuum in (1).

(1) **Syntactic Coding Devices in Different Degrees of Topic Continuity**

_____Highest Topic Continuity_____

a. Went to school. (Zero anaphora)
b. He went to school. (Unstressed anaphoric pronoun)
c. He' went to school. (Stressed independent pronoun)
d. The guy went to school. (Definite NP)
e. A guy with sunglasses went to school. (Indefinite NP)

_____Lowest Topic Continuity_____

doi:10.6342/NTU202302199

2.1.2 Gundel et al. (1993)

According to a number of previous studies (Chafe 1976, 1987, Hawkins 1978, 1991, Prince 1981, 1992, Yule 1981, Garrod and Sandford 1982, Ariel 1985, 1988, Gundel Hedberg and Zacharski 1993, Gundel, Hegarty and Borthen 2003), the speaker's selection of a referential expression in natural discourse is highly related to the cognitive status of its referent. Having observed several languages, such as English, Japanese, Mandarin Chinese, Russian and Spanish, Gundel and her colleagues (Gundel et al. 1993, Gundel, Hegarty and Borthen) claimed that there are six cognitive statuses a referent may have, and the different cognitive statuses are coded by different referential expressions. On top of that, a Givenness Hierarchy is proposed to show the implication relation among the cognitive statuses, seen in (2):

(2) **The Givenness Hierarchy (Gundel et al. 1993)**

In focus > Activated > Familiar > Uniquely Identifiable > Referential > Type Identifiable

Across languages, each status on the hierarchy is a necessary and sufficient condition for the appropriate use of a different form and forms (Gundel et al. 1993). For instance, if the hearer is able to access a representation of the type of object described by the expression, then the referent is in the status of "*type identifiable*", and such status is necessary for appropriate use of any nominal expression. When the speaker intends to refer to a particular object or objects, then the referent is in the status of "*referential*", and this status is necessary for appropriate use of all definite nominal expressions. The notion "*uniquely identifiable*" refers to the status in which the hearer is able to identify the speaker's intended referent on the basis of the nominal alone. Also, the status of uniquely identifiable is necessary for all definite reference. The notion of "*familiar*" refers to the status in which the hearer can uniquely identify the intended referent since

doi:10.6342/NTU202302199

s/he already has a representation of it in memory. Such status is necessary for all personal pronouns and definite demonstratives. "*Activated*" is the status in which the referent is represented in current short-term memory of the hearer, and it is necessary for appropriate use of all pronominal forms. Lastly, the referent is "*in focus*", when it is not only in the hearer's short-term memory, but is also at the current center of his/her attention. The status of in focus is the most restrictive among all forms and is necessary for the use of zero and unstressed pronominals.

Based on Gundel et al., the six statuses signaled by different referential expressions are not mutually exclusive. That is, a higher status can entail all lower statuses. A referring form can be replaced by forms that require lower cognitive statuses, and a form can also be distributed across over one status. Therefore, when necessary conditions for the use of more than one form are fulfilled, how a particular referential expression is chosen becomes a question. To answer the question, Gundel et al. (1993) claim that two conversational implicatures of Grice's Maxim of Quantity (Grice 1975) shown in (3) can be used to predict the interaction between the referential forms and cognitive statuses:

(3) **Maxim of Quantity (Grice 1975):**

Q1: Make your contribution as informative as required (for the current purposes of the exchange).

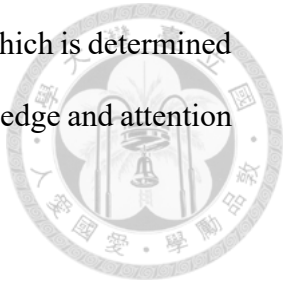
Q2: Do not make your contribution more informative than is required.

In Q1 implicature, the use of a weaker form implicates that a stronger form does not exist. In Q2, by contrast, the use of a weaker form implicates a stronger form. Nevertheless, Gundel et al. (1993) don't further explain how to define "informative".

In summary, Gundel et al.'s hypothesis reveals that the selection of certain

doi:10.6342/NTU202302199

referential expression is related to the cognitive status of a referent, which is determined by assumptions that the speaker make considering the hearer's knowledge and attention state in the particular context.



2.2 Previous Research on Syntactic Coding Devices of NPs in Budai Rukai

In the discussion of nominal structure in Budai Rukai, the nominal phrases (henceforth, NPs) generally occur with components of the case markers, the demonstratives and pronouns as well. The occurrences of these components in one nominal phrase are not only for syntactic reasons, such as case marking, but also for semantic reasons, such as the specification of deictic information (Chen 2008). For instance, both the case marker *ka* and *ku* can be adopted to mark subjects, but the former indicates a definite meaning and the latter indefinite, as shown in (4a-b).

(4)

a. Wagelregelrethe ka lavavalake.

W-a-gelregelrethe	ka	lavavalake
ACT-RLS-cry	NOM/DEF	child

'The child is crying.'

b. Wagelregelrethe ku lavavalake.

W-a-gelregelrethe	ku	lavavalake
ACT-RLS-cry	NOM/INDEF	child

'A/some child is crying.'

(Chen 2008)

On the other hand, all of the three case markers *ka, ku* and *ki* can be used to mark direct objects, but resulting in different semantic meanings. Examples in (5) show us that *ki* describes kind (indefinite), *ku* specific (indefinite) and *ka* definite.

doi:10.6342/NTU202302199

(5)

a. Wadruraku ki lrenege.

w-a-druru=aku	ki	lrenege
ACT-RLS-push=1S.BN	OBL	stone

'I pushed a stone.' (Not the other kind of object)

b. Wadruraku ku lrenege.

w-a-druru=aku	ku	lrenege
ACT-RLS-push=1S.BN	OBL	stone

'I pushed a stone.'

c. Wadruaku ka lrenege.

w-a-druru=aku	ka	lrenege
ACT-RLS-push=1S.BN	OBL	stone

'I pushed the stone.'

(Chen 2008)

Besides, definiteness, or deicticity, can be signified by demonstratives as well. Different demonstratives are able to indicate different "visibilities" and "distances" of the referents, as shown in the Table 2.1 below (Zeitoun 2000, Chen 2008).

Table 2.1 Demonstratives in Budai Rukai

Demonstrative	Visibility	Distance
kay "this"	+	-
kikay "this"	+	-
kuini "that"	+	-
kavay "that"	+	-
kudra "that"	-	+

2.3 The Implications to the Present Study

So far, we have looked into two dimensions that are claimed to influence the

doi:10.6342/NTU202302199

selection of the referential expressions. Firstly, the referential form with the least code quantity codes maximum accessibility and minimum importance in discourse (Givón 1983). Secondly, various grammatical forms on a NP may result from the referent in the speaker's different cognitive statuses (Gundel et al. 1993).

In the previous studies of Budai Rukai, the focus is often confined to the relation between grammatical coding devices and deicticity (Zeitoun 2000, Chen 2008). There is few investigating the association between syntactic coding devices and the discourse /cognitive status of a NP.

Consequently, this study aims to find the reasons behind the choice of the referential forms on NPs by examining the topic continuity (Givón 1983) and the referential givenness (Gundel et al. 1993) of NPs in Budai Rukai.

doi:10.6342/NTU202302199

Chapter 3
Topic Continuity in Discourse of Budai Rukai

3.1 Introduction

In recent decades, a growing number of studies have looked at certain syntactic phenomena from the viewpoint of their functional motivation in human communication and discourse pragmatics (Bolinger 1986, 1979; Chafe 1972, 1980; Givón 1979a, 1979b, 1981, 1982; Halliday & Hasan 1976; Labov 1972). Likewise, the main purpose of this chapter is to examine how different syntactic coding devices of NPs help connect topics in the discourse of Budai Rukai. Topic continuity, a concept proposed by Givón (1983), concerns primarily wirh how a speaker creates connections and coherences in an ongoing discourse, and also what kinds of syntactic coding device the speaker uses in order to help his/her hearer identify the topic in the discourse (Pu, 1989). Givón (1980, 1981a and 1982) hypothesized that the syntactic coding devices on NPs can be ordered hierarchically as shown in the topic continuity model below in (1):

(1) **The Scale of Topic Continuity** (Givón 1980)

most continuous (least surprising)
↑ Zero anaphora
Unstressed/clitic pronouns
Stressed/independent pronouns
Unmodified definite NP
Restrictively modified definite NP
↓ Referential indefinite NP
least continuous (most surprising)

The basic assumption of this model is that the more continuous the NP is, the less coding material the hearer requires to identify it, and consequently the less elaboration the speaker has to make. In other words, the syntactic devices at the top of the hierarchy possess higher topic continuity, which means that the referential identification of such

doi:10.6342/NTU202302199

NPs is easier. In contrast, those closer to the bottom of the hierarchy possess lower topic continuity and higher surprise to the hearer. Therefore, the hearer might have more difficulty in assigning referentiality of such NPs. In addition to syntactic coding devices, some other factors, including word order, case roles and humanness, are found to have some influences on topic continuity as well.

According to Givón, topic continuity of certain referential expression can be measured by some quantifiable parameters. In this chapter, such parameters are adopted to investigate the ways in which referential items are continued or discontinued in the discourse of Budai Rukai, and also to see whether the results found in Budai Rukai conform with Givón's topic continuity model.

This chapter is arranged as follows: Section 3.2 introduces the primary method used to measure the topic continuity of our data; Section 3.3 goes over the common syntactic coding devices in discourse of Budai Rukai; Section 3.4 shows the findings of our analysis, and later on discusses the topic continuity in Budai Rukai from the aspects of different factors, and Section 3.5 summarizes the chapter.

3.2 Description of Methodology

3.2.1 Measurements

The quantitative method used in this chapter was suggested by Givón (1979, 1980). This method assumes that each NP in the discourse has certain *degree of topic continuity*, or *topicality*, which refers to the degree of referential continuity of a NP on the clausal level. This method provides three separate parameters to measure the degree of topic continuity. Two of them will be considered in this chapter:

(a) Referential Distance (look-back): The number of clauses between the present mention of a NP and the last mention of the same referent.

doi:10.6342/NTU202302199

(b) Persistence (decay): The number of successive clauses in which a particular NP persists.

The parameter of *referential distance* measures the topic continuity of a certain NP in terms of how many clauses to the left intervene between the last mention of the NP and the current mention of it. The minimal value assigned is 1, in which case the NP last occurs in the immediately preceding clause, and the highest value is arbitrarily set at 20 because of the belief that a hearer will not normally be able to retrieve referential information prior to roughly 20 clauses to the left. In short, the value of referential distance ranges from 1 to 20, with 1 representing maximal continuity and 20 maximal discontinuity. On the other hand, the parameter of *persistence* measures how many clauses to the right the NP under study will persist. The minimal value of persistence is 0, indicating such NP doesn't persist. For this measure, the higher the score, the greater the continuity. Theoretically, there is no upper limit to the value of persistence. **It is typically expected that a highly topical and continuous NP in the discourse has a low value for referential distance and a high one for persistence** (Cooreman, 1983).

3.2.2 Texts

The data analyzed in this thesis is made up for eleven narrated texts from the NTU Corpus of Formosan Languages (Sung et al. 2008; Su et al. 2008[1]). Some details of the narrators and the texts are reproduced from Section 1.4 in Table 3.1.

Table 3.1: List of native speakers and texts recorded

Name of Text	Genre	Speaker	Gender	Year of Birth
Millet	Narrative	Tagas	M	1938
Childhood	Narrative	Balenge	F	1961

[1] NTU Corpus of Formosan Languages (台大台灣南島語語料庫): https://formcorp.netlify.app/#/

doi:10.6342/NTU202302199

Frog Story	Narrative	Kainguane	M	none
Frog Story	Narrative	Legeai	M	none
Frog Story	Narrative	Salrabu	M	1938
Pear story	Narrative	Ba Ching i	M	none
Pear story	Narrative	Balenge	F	1961
Pear story	Narrative	Legeai	M	none
Pear story	Narrative	Salrabu	M	1938
Pear story	Narrative	Tagas	M	1938
Pear story	Narrative	Wauki	M	1933

All noun phrases in these texts are labelled and some of their syntactic and semantic features, such as syntactic coding device, case role as well as humanness, are marked. Also, their average values of referential distance and persistence are calculated. By doing so, we expect to see the interactions between topic continuity and the syntactic or semantic features of NP in discourse of Budai Rukai. The results are shown in the following sections.

3.3 Syntactic Coding Devices Investigated

Before presenting the functional analysis of the syntactic coding devices on the NPs in Budai Rukai discourse, it seems useful to give a brief introduction to these syntactic devices. It is found that the NPs in our data are encoded by four different types of syntactic devices, including **zero-anaphora**, **pronoun**, **definite noun** and **indefinite noun**. All of them will be illustrated with certain examples as follows.

3.3.1 Zero Anaphora

This device is the use of a gap that "refers back" to something mentioned previously. In Budai Rukai, two types of zero anaphora are found. One of them is the completely deleted noun in the discourse, as shown in (2a), and the other type is the third person nominative, always in the zero form, as illustrated in (2b). In the following

doi:10.6342/NTU202302199

two examples, both of the zero anaphora are marked by **ϕ.**

(2a) Zero Anaphora: Completely Deleted Noun
(RukaiNr-frog_Legaile IU86-89)
[86]Yakay ku [87]tualay adringi ku [88]lu si [89]kaynganay pangituluku ***ϕ****.*

(IU86)		**(IU87)**			**(IU88)**	
i-a-kay	ku	tualay	adringi	ku	lu	si
LOC-RLS-this	OBL	from	inside	OBL	owl	and

(IU89)

kaynganay	pa-ngi-tuluku	**ϕ**
come.out	CAU-self-surprise	**(the.dog)**

"Then there is an owl coming from inside and surprising **(the dog)**."

(2b) Zero Anaphora: Third Person Nominative
(RukaiNr-pear_Wauki IU24-26)
[24]Sa katuatuasenga ***kuini vavalake*** *yaie, [25]madredresenge=****ϕ*** *kudra ki mua lrikilrikili kudra ababay, [26]si mapapungupungu=****ϕ****.*

(IU24)

sa	ka-tua-tuase=nga	**kuini**	**vavalake**	yaie,
when	STAT.IRR-RED-leave=PFV	that.VIS.PROX	child	TOP

(IU25)

ma-dre-dresenge(=**ϕ**)	kudra	ki	mua	lriki-lrikili
STAV.RLS-RED-meet(**=3S.BN**)	that.INV	OBL	go	RED-bicycle

(IU25)		**(IU26)**	
kudra	ababay	si	ma-pa-<pungu>pungu(=**ϕ**).
that.INV	girl	and	STAT.RLS-CAU-RED-bump.into(**=3S.BN**)

"When **the child** left, **(he)** met a girl who rode a bicycle and **(he)** bumped into (her)."

In IU 89 of (2a), the oblique "the dog" is in the form of zero anaphora, which is

doi:10.6342/NTU202302199

deliberately elided by the speaker. As for the subject *kuini vavalake* “that child” in IU 24 of (2b), it is later on covert in IU 25 and 26 because of its case role of third person nominative. Although both the subject and the oblique are allowed to be zero anaphora in Budai Rukai, the percentage of the subject with the device of zero anaphora (41%) is much higher than that of the oblique with the same device (4%), which is because most of the subjects in our data refer to third person nominative and behave in the form of zero anaphora.

3.3.2 Pronoun

This device is the use of a word to refer to a noun that has already been mentioned previously. According to our data, there are three types of pronouns found, including **personal clitic pronoun, personal full pronoun and demonstrative pronoun**. In many other languages, the most common pronouns are the personal pronouns, which can refer to the speaker (first person), the hearer (second person), or other people or things (third person). Personal pronouns also play a significant role in the pronoun system of Budai Rukai. Just similar to nouns, personal pronouns can function as different cases and might be presented by certain case forms. About 80% of the pronouns in our data are personal pronouns, with personal clitic pronouns accounting for 67%. It is especially noted that most of the **personal clitic pronouns** in our discourse data are third person genitive. In IU 65 of (3a), the third person singular genitive *ini* “his”, which is attached on the noun *drapale* “feet”, refers to the covert subject “the boy” in IU 63.

(3a) Personal Clitic Pronoun

(RukaiNr-pear_Balenge IU63-65)

*[63]Sa ka muadreke=ϕ la [64]kuini la navate la ngucilri patelre [65]si la angeange kuini drapale**ini**.*

doi:10.6342/NTU202302199

(IU63)				**(IU64)**		
sa	ka	mu-a-dreke(=ϕ)	la	kuini	la	navate
when	KA	go-RLS-fall**(=3S.BN)**	then	that.VIS.PROX	then	guava

(IU65)

la	ngu-cilri	patelre,	si	la	angeange	kuini
then	REFL-thown	all	and	then	hurt	that.VIS.PROX

drapale**=ini.**
feet=3S.BG

"When **(the boy)** fell, all the guavas were falling down, and his leg hurt."

As for the **personal full pronoun**, the third person singular oblique *iniane* is the only personal full pronoun found in our data, accounting for 13% of the pronouns in the data. In IU68 of (3b), *iniane* is used to refer to the covert subject "the boy" in IU67.

(3b) Personal Full Pronoun
(RukaiNr-pear_Balenge IU67-69)
[67]Sa nguadreke=ϕ la kela kudra tatulru kudra lamalala, [68]kudra mia ***iniane*** *ka [69]lavavalake.*

(IU67)

sa	ngu-adreke(=ϕ)	la	kela	kudra	tatulru	kudra
when	REFL-fall(**=3S.BN)**	then	come	that.INV	three	that.INV

(IU68)					**(IU69)**
la-mala-la	kudra	mia	**iniane**	ka	la-vavalake.
P-pal-male	that.INV	like	3S.FO	NOM	P-child

"When **(the boy)** fell, those three pals came and these kids were like **him**."

The demonstrative pronouns, which take up around 20% of pronouns in our data, are used to represent or identify a person, place, animal or thing. Unlike the

doi:10.6342/NTU202302199

demonstratives in English, there is no distinction between singular and plural among the demonstrative pronouns in Budai Rukai. But, they are distinct by the distance they refer to. For example, *kuini* "that" is farther than *kay* "this" while closer than *kudra* "that.INV". Among them, *kudra* and *kuini* are two demonstrative pronouns mostly used in the discourse. Taking a look at (3c), the demonstrative pronoun *kudra* in both IU 75 and IU 76 refer to "the dog" which has been mentioned in the previous discourse.

(3c) Demonstrative Pronoun
(RukaiNr-frog_Kainguane IU75-77)
*[75]Saalealeale **kudra** yaie, [76]la ikay **kudra** iya belenge ki [77]angatu ku sigu taiya.*

(IU75)				**(IU76)**		
sa-ale-ale-ale		**kudra**	yaie,	la	i-kay	**kudra**
when-RED-RED-bark		that.INV	TOP	then	LOC-this	that.INV

			(IU77)			
iya	belenge	ki	angatu	ku	sigu	taiya
say	above	OBL	tree	OBL	hornet	DM

"When **that (the dog)** was barking in this way, **that (the dog)** said that there was one hornet on the tree."

3.3.3 Definite Noun

This device is generally made up of a noun preceded by a definite article such as "the" in English. Although there is no definite article in Budai Rukai, demonstratives are found to have an alternative function of definite article, which is to indicate that the identity of the noun is recognized by the hearer. Therefore, the definite nouns in our data are manifested by a noun with a demonstrative preceding it. In IU 17 of (4a), *kuini kamadha* "the mangos" is a typical example of definite noun in Budai Rukai.

doi:10.6342/NTU202302199

(4a) (RukaiNr-pear_Tagas IU17-18)
[17]Ngisarade ***kuini kamadha*** *[18]ka kadaranane.*

(IU17)			**(IU18)**	
ngisarade	**kuini**	**kamadha**	ka	kadalranane
scatter	that.VIS.PROX	mango	OBL	road

"The mangos scattered all over the road."

3.3.4 Indefinite Noun

The so called indefinite noun is one kind of noun that is not specific in its reference. For example, in (5a), *ku vavalake* "child" in IU5 is modified by a number *tangea* "one", and this noun phrase *ku tangea ku vavalake* "one child" doesn't refer to any child that has been mentioned.

(5a) Indefinite Noun modified by number
(RukaiNr-pear_Tagas IU4-8)
[4]kaynganananganay ***ku [5]tangea ku vavalake*** *[6]ngucidinsya [7]alra kupa kuini [8]akuvaeva kuini karadrare si katuase.*

(IU4)		**(IU5)**			**(IU6)**
kay<ngana>nganay	**ku**	**tangea**	**ku**	**vavalake**	ngu-cidinsya
<RED>come	NOM	one.HUM	REL	child	NGU-bike

(IU7)			**(IU8)**		
alra	kupa	kuini	aku-vaeva	kuini	karadrare
take	steal	that.VIS.PROX	AKU-one	that.VIS.PROX	bamboo.basket

si	ka-tuase
and	STAT.IRR-leave

"There came **one child** riding on a bike and (he) stole one of those baskets and left."

Mostly, indefinite nouns are simply modified by a case marker, just like *kadalranane* "road" in IU 20 of (5b). Marked by oblique case marker *ki*, the indefinite

doi:10.6342/NTU202302199

noun phrase *ki kadalranane* “the road” represents general roads on earth instead of any specific roads.

(5b) Indefinite Noun modified by case marker
(RukaiNr-pear_Ba-ching-i IU17-20)
[17]Kudra ililukuini ku [18]tukunui [19]la mu [20]gugu patelre mua ***ki kadalranane****.*

(IU17)		**(IU18)**	**(IU19)**		**(IU20)**
kudra ililuku=ini	ku	tukunui	la	mu	gugu
that.INV carry=3S.BG	OBL	jelly	then	go	fell

patelre	mu-a	**ki**	**kadalranane**.
all	go-RLS	OBL	road

“All the jelly they had carried fell to **ground**.”

Sometimes, indefinite noun occurs without any modification or marker. Take (5c) for example, *talialalay* “leader” in IU 45 is in its bare form and what it actually talks about is any leader in Budai Rukai but not specific leader mentioned in the previous discourse.

(5c) Indefinite Noun in bare form
(RukaiNr-Becenge_Tagas IU45-47)
*[45]****Talialalay*** *la iluku ki [46]e marudrange [47]si la tucapicapili.*

(IU45)				**(IU46)**		**(IU47)**	
talialalay	la	iluku	ki	e	marudrange	si	la
leader	then	bring	OBL	FIL	senior	and	then

tu-capi-capili
do-RED-tandoor

“**The leader** will lead the seniors, and then start to do the millet-baking ceremony.”

Interestingly, some indefinite nouns in Budai Rukai can be attached by certain

doi:10.6342/NTU202302199

prefix and become verbalized such as *mubiabila* “go to the river bank” in (5d). This verbalized phrase is formed by an indefinite noun *biabila* “bank” with a prefix *mu-* “go”. It’s noted that when the speaker uses the phrase, s/he doesn’t mean to go to any specific banks but simply mean the general concept of “going to the river bank”.

(5d) Verbalized Indefinite Noun

(RukaiNr-frog_Salrabu IU132)

[132]*La katuase tupapapalra kuini ki angatu si ala mu**biabila**.*

(IU132)

la	ka-tuase	tu-pa-pa-palra	kuini	ki	angatu
then	STAT.IRR-leave	TU-RED-RED-follow	that.VIS.PROX	OBL	wood

si	ala	mu-**biabila**.
and	then	go-bank

“And then (they) left and were following that wood and going to **river bank**.”

3.4 Syntactic Coding Devices and Topic Continuity

After going over four common syntactic coding devices on the NPs in discourse of Budai Rukai, it’s time to look at the relation between the syntactic coding devices and topic continuity.

3.4.1 Numerical Results and Discussion

Table 3.2 below gives the average values for the measurements of referential distance and persistence applied to the various NP coding devices. As we have mentioned in 3.2.1, Givón’s definition of topic continuity reveals that a more continuous NP might possess lower value of referential distance and higher value of persistence. Accordingly, Table 3.2 indicates that different grammatical coding devices analyzed in this paper can be approximately ranked based on the topic continuity of the

doi:10.6342/NTU202302199

NP referents they encode in the discourse. It seems that the ranking can well correspond with the scale of topic continuity proposed by Givón, which is given again in (6).

Table 3.2 Measures of Topic Continuity by Syntactic Coding Devices

		Referential Distance	Persistence
Zero-anaphora		1.4	2.7
Pronoun	**Personal clitic**	1.3	2.3
	Personal full		
	Demonstrative		
Definite noun		1.8	1.4
Indefinite noun	**Case marker**	2.1	0.75
	bare		
	verbalized		

(6) **The Scale of Topic Continuity** (Givón, 1983)

most continuous

↕

Zero anaphora

Clitic pronouns

Independent Pronoun

Unmodified definite NP

Restrictively modified definite NP

Indefinite NP

most discontinuous

However, in order to apply Givón's scale to our findings in Budai Rukai, some slight adjustments are needed. That is, "clitic pronoun" and "independent pronoun" are regarded as the same "pronoun" category in the scale of Budai Rukai. Also, there is no distinction between "unmodified definite NP" and "restrictively modified NP" in our data. Therefore, the scale of topic continuity in Budai Rukai looks like (7), with zero anaphora the most continuous, pronoun the second, definite NP the second least and indefinite NP the least continuous.

doi:10.6342/NTU202302199

(7) **The Scale of Topic Continuity in Budai Rukai**

most continuous

↕ Zero anaphora

Pronoun

Definite NP

Indefinite NP

most discontinuous

Depending on Table 3.2, as two most continuous coding devices, **zero anaphora** and **pronoun** possess very close values of referential distance (1.4; 1.3) and persistence (2.7; 2.3), implying that they share quite similar properties of topic continuity. This finding does make sense. As mentioned in 3.3.1, since most cases of zero anaphora in our discourse are actually third person pronouns in the zero form, it isn't surprising that zero anaphoras perform similarly as pronoun does in topic continuity. On average, both of these two devices are used to refer back to a closer NP, which is often around one sentence away. What's more, the referents coded by these two devices tend to appear continuously for around two sentences. For instance, the subject "they" in IU 132 of (8) is a zero anaphora, referring to *kuini taupungu si vavalake* "the dog and the child" in the previous one sentence, and occurs in the next two sentences (IU133 and IU139) as subject, also in the form of zero anaphora.

(8) **RukaiNr-frog_Salrabu IU130-139**

*[130]Katuase si la ubelenge **kuini taupungu si vavalake** kuini ki angatu. [131]O [132]la katuase tupapapalra=**ϕ** kuini ki angatu si ala mubiabila. [133]Sa e mubiabilanga=**ϕ** yaie [134]e [135]ala ikay kudra ki [136]babiabila ki [137]aciaacilay yaie yakay kudra [138]e latakurauru. [139] La drele=**ϕ** kuini ki latakurauru si ala ikay kudra takurauruini.*

(IU130)

ka-tuase	si	la	u-belenge	**kuini**	**taupungu**
STAT.IRR-leave	and	then	go-up	that.VIS.PROX	dog

doi:10.6342/NTU202302199

si	**vavalake**	kuini	ki	angatu
and	child	that.VIS.PROX	OBL	wood

"**That child and dog** left and went up to that wood."

(IU132)

la	ka-tuase	tu-pa-pa-palra(**=ϕ)**	kuini
then	STAT.IRR-leave	TU-RED-RED-follow(**=3P.BN)**	that.VIS.PROX

ki	angatu	si	ala	mu-biabila
OBL	wood	and	then	go-bank

" And then **(they)** left and were following that wood and went to the bank."

(IU133) **(IU134)** **(IU135)**

sa	e	mu-biabila=nga(**=ϕ)**	yaie,	e	ala	ikay
when	FIL	go-bank=PFV(**=3P.BN)**	TOP	FIL	then	LOC-this

(IU136) **(IU137)**

kudra	ki	babiabila	ki	acia-acilay	yaie	i-a-kay
that.INV	OBL	bank	OBL	RED-water	TOP	LOC-RLS-this

(IU138)

kudra	e	la-takurauru
that.INV	FIL	P-frog

"When **(they)** had gone to the bank, there were frogs at the bank of river."

(IU139)

la	drele(**=ϕ)**	kuini	ki	la-takurauru	si	ala
then	see(**=3P.BN)**	that.VIS.PROX	OBL	P-frog	and	then

i-kay	kudra	takurauru=ini
LOC-this	that.INV	frog=3S.BG

"And **(they)** saw some frogs, and there was their frog."

doi:10.6342/NTU202302199

It is noted that the values in Table 3.2 are just average values. Therefore, sometimes zero anaphora or pronoun can persist longer than three sentences, just as the zero anaphora subject "he" in IU 39 and IU 40 of (9), which persists for nine sentences in total, eventually serving as a third person oblique *iniane* in IU70 and IU72.

(9) **RukaiNr-pear_Balenge IU39-43**

[39]Pua=ϕ ki lrikilini [40]si la katuase=ϕ ilukua kudra e [41]suaete [42]kudra e [43]navate kuini ki karadradre. (Nine sentences later) [70]Sa katuase sa dreleiniane yaie, [71]la e, [72]pararubuiniane [73]malra kudra navate [74]si [75]ala siludu [76]patelre kudra navate pua kuini ki karadrare.

(IU39)					**(IU40)**	
pu-a(=**ϕ**)	ki	lrikili=ini	si	la	ka-tuase(=**ϕ**)	
put-RLS(=**3S.BN**)	OBL	bike=3S.BG	and	then	STAT.IRR-leave(=**3S.BN**)	

			(IU41)	**(IU42)**		**(IU43)**
ilukua	kudra	e	suaete	kudra	e	navate
bring	that.INV	FIL	full.of	that.INV	FIL	guava

kuini	ki	karadradre
that.VIS.PROX	OBL	basket

"**(He)** put (it) into his bike, and then **(he)** left and brought that basket that was full of guavas."

……

……

(Nine sentences later)

(IU70)						**(IU71)**	
sa	ka-tuase	sa	drele	**iniane**	yaie	la	e,
when	STAT.IRR-leave	when	see	3S.FO	TOP	then	FIL

(IU72)		**(IU73)**			**(IU74)**	**(IU75)**	
pa-rarubu	**iniane**	malra	kudra	navate	si	ala	siludu
CAUS-help	3S.FO	take	that.INV	guava	and	then	pick.up

doi:10.6342/NTU202302199

(IU76)

patelre	kudra	navate	pu-a	kuini	ki	karadrare
all	that.INV	guava	put-RLS	that.VIS.PROX	OBL	basket

"When (they) were walking and saw **him**, then (they) helped **him** take the guavas and put all the guavas into the basket."

As for **definite NP**, according to Table 3.2, it can refer back to the referent that is about two sentences away in general, a little farther than zero anaphora and pronoun do. But, such referent has more difficulties persisting longer than two sentence. For instance, in IU 124 of (10), the definite subjects *kuini taupungu* "that dog" and *kuini vavalake* "that child" are both mentioned approximately two sentences before. Specifically speaking, *kuini taupungu* "that dog" appears just two sentences away in IU 118 as a definite NP subject, and *kuini vavalake* "that child" is a zero anaphora subject three sentences away in IU 115 and IU116. Although they are brought back into the speaker's attention, they don't tend to stay as topics for long. They serve as a zero anaphora subject for one more sentence in IU125 and then the speaker changes the topic into talking the frog (*kuini takurauru* in IU128) instead.

(10) **RukaiNr-frog_Legeai IU113-130**

*[113]Ala sa [114]e [115]muanga=ϕ kuini ki [116]valru muanga=ϕ kuini ki acilay. [117]Tarudrusa kuini ki e, [118]**kuini ki taupungu** ala ikay ku angatu kudra e.* ***(two sentences later)*** *[123]La mua kameamealane **kuini e [124]taupungu si kuini vavalake**. [125]Sa ngimianga=ϕ kuini yaie, [126]e [127]ala kamani kuini [128]e takurauru kudra na [129]silasilapelini matuase mu valru mua kudra ki [130]papalralini.*

(IU113)	**(IU114)**	**(IU115)**			**(IU116)**
ala	sa e	mu-a=nga(=ϕ)	kuini	ki	valru
then	when FIL	go-RLS=PFV=**3S.BN**	that.VIS.PROX	OBL	river

mu-a-nga(=ϕ)	kuini	ki	acilay.
go-RLS=PFV(=**3S.BN)**	that.VIS.PROX	OBL	water

doi:10.6342/NTU202302199

"Then when **(the boy)** went to this river, **(the boy)** went into this water."

(IU117)

tarudrusa	kuini	ki	e
two.HUM	that.VIS.PROX	OBL	FIL

(IU118)

kuini	**ki**	**taupungu**	ala	i-kay	ku	angatu
that.VIS.PROX	OBL	dog	then	LOC-this	OBL	wood

kudra	e.
that.INV	FIL

"These two, **that dog**…… There was a wood there."

……

……

(Two sentences later)

(IU123)

la	mua	ka-meameal-ane
then	go	real-RED-dry-NMZ

(IU124)

kuini	**e**	**taupungu**	si
that.VIS.PROX	FIL	dog	and

kuini	**vavalake**
that.VIS.PROX	kid

"Then **the dog** and **that kid** go to the river bank."

(IU125) **(IU126)** **(IU127)**

sa	ngi-mia=nga(=ϕ)	kuini	yaie,	e	ala	kamani
when	REFL-be.like=PFV**(=3S.BN)**	that.VIS.PROX	TOP	FIL	then	STAT.IRR-be

(IU128) **(IU129)**

kuini	**e**	**takurauru**	kudra	na	sila-silape=lini
that.VIS.PROX	**FIL**	**frog**	that.INV	PFV	RED-search.for=3P.BG

(IU130)

ma-tuase	mu-valru	mua	kudra	ki	papalra=lini
STAT.RLS-leave	go-river	go-RLS	that.INV	OBL	partner=3P.BG

doi:10.6342/NTU202302199

"When **(they)** are like that, (they say) this is **the frog** they have been looking for, and it went into the river and went to its partners."

Just like in many other languages, **the indefinite NP** is the least continuous coding device in Budai Rukai. Since this device is often used to encode a first-mentioned referent or a reintroduced referent that hasn't been in the focus for a long period of time, there is no doubt that its value of referential distance (2.1) is higher than those of any other three coding devices (1.3; 1.4; 1.8). What's more, an indefinite referent doesn't tend to stay long in the discourse. That is why its value of persistence (0.75) is the lowest among four coding devices (2.7; 2.3; 1.4) . Take (11) for example, the indefinite noun *ku kiw* " a goat" in IU 11 is a brand new referent and it isn't a significant topic in the discourse, so it doesn't persist in the following discourse.

(11) **RukaiNr-pear_Salrabu IU11-13**

[11]Lribate kuini talragini maililuku ***ku kiw*** *si [12]la katuase lribate naw pararubu la kai iya [13]sa ka bilrilane.*

(IU11)

lribate	kuini	talrag=ini	ma-il-iluku	**ku**	**kiyu**
pass	that.VIS.PROX	friend=3S.BG	STAT.RLS-RED-bring	OBL	goat

(IU12)

si	la	ka-tuase	lribate	naw	pararubu	la	kai	iya
and	then	STAT.IRR-leave	pass	want	help	then	NEG	say

(IU13)

sa	ka	bilril-ane
when	OBL	behind-NMLZ

'His friend passed and brought **a goat**, (but) (he) didn't not want to help in the end.'

Another example of indefinite noun is given in (12). The indefinite subject *ku lasigu*

doi:10.6342/NTU202302199

"bees" in IU122 is brought back into the context from IU103, which is eight sentences away. Despite being back into the focus, this indefinite referent doesn't seem to serve as an essential topic, therefore not persisting later on.

(12) **RukaiNr-frog_Kainguane IU102-122**

[102]La [103]um kirimu ponpon **kay lasigu** si *[104]la kasamali kay na kulrabau kudra si [105]eh nia tuluku taiya.* ***(eight sentences later)*** *[121]La ngibwale si la ngituluku kay vavalake si la tuaverevere kudra kiasaladhaladha ki taupungu [122]**ku lasigu** kudra ki [123]eh tuaverevere taiya.*

(IU102)	**(IU103)**					
la	um	kirimu	ponpon	**kay**	**la-sigu**	si
then	FIL	suddenly	humming	this	P-hornet	and

(IU104)					
la	ka-samali	kay	na	kulrabau	kudra	si
then	STAT.IRR-surprised	this	even	vole	that.INV	and

(IU105)			
eh	nia	tuluku	taiya
FIL	REFL	frightened	DM

'Then suddenly these hornets kept humming; the voles were surprised and frightened.'

……

……

(Eight sentences later)

(IU121)							
la	ngi-bwale	si	la	ngi-tuluku	kay	vavalake	si
then	REFL-appear	and	then	REFL-frightened	this	child	and

doi:10.6342/NTU202302199

la	tu-a-verevere	kudra,	ki-a-saladha-ladha	ki	taupungu
then	toward-RLS-fall	that.INV	PASS-RLS-chase-RED	OBL	dog

(IU122)				**(IU123)**		
ku	**la-sigu**	**kudra**	ki	eh	tu-a-verevere	taiya.
NOM	PL-hornet	that.INV	OBL	FIL	toward-RLS-fall	DM

'The owl appeared and the child was so frightened that he fell down from the tree.. The dog was chasing **the hornets**.'

In short, in the discourse of Budai Rukai, if one referent is right in the attention of the speaker and will continue to be the topic in the talking for a while, then the speaker tends to adopt the more continuous syntactic coding devices, such as zero anaphora or pronoun, to encode such referent. On the other hand, when a referent is out of the attention and has very little tendency to persist in the discourse, then the less continuous devices, like definite NP or especially indefinite NP, are chosen by the speaker instead. As a result, we can conclude that the syntactic coding devices used on the noun phrases in discourse of Budai Rukai correspond perfectly with the scale of topic continuity proposed by Givón.

3.4.2 Topicalization and Topic Continuity

In the previous section, we have investigated the relation between syntactic coding devices and the topic continuity in Budai Rukai by a quantitative method proposed by Givón. During the investigation, we have found one interesting category of coding device which isn't included in Givón's scale of topic continuity--**topicalization**.

The basic word order of Budai Rukai is VS. Even though the majority of the clauses adhere to this basic word order, there are some that follow the SV word order pattern for the pragmatic reasons. This phenomenon, in which certain referent is fronted to the sentence-initial position by the speaker and is often followed by a short pause, is

doi:10.6342/NTU202302199

called “topicalization”. In Budai Rukai, the **demonstrative pronoun** and **definite NP** are two elements that are most likely to be topicalized. These topicalized elements are used to emphasize the reintroduced referents, as illustrated in the following two cases, (13) and (14).

(13) **RukaiNr-frog_Kainguane IU8-18**

[8]Papangua dingidingi si sia-vavava ***ϕ*** *taiya. (Two sentences later) [11]**Kuini iyaw** kudra yaie, [12]alaka waituku. [13]Kay marakace **kay iyaw** yai. [14]Kudra si adravane kudra mia kuini ka didilrungu taiya [15]kudra marakace ituku [16]la ituku kuini [17]iyaw si katuase ubere taiya [18]la ngukay kuini ginganga kai kanicaebane.*

(IU8)

pa-pangua	dingi-dingi	si	sia-vavava	**ϕ**	taiya
RED-use	RED-shake	and	enjoy-<RED>toy	(3S.FO)	DM

‘It (the dog) shook and watched **(the frog)**.

……
……
(Two sentences later)

(IU11) **(IU12)**

kuini	**iyaw**	**kudra**	yaie,	alaka	wa-ituku
that.VIS.PROX	frog	that.INV	TOP	turn.out	ACT-RLS-jump

“**That frog** jumped out.”

(IU13)

kay	marakace	**kay**	**iyaw**	yaie
this	violently	this	frog	TOP

‘**This frog** jumped violently.’

doi:10.6342/NTU202302199

(IU14)

kudra	si	adravane	kudra	mia kuini	ka	di-dilrungu	taiya,
that.INV	and	no.matter	that.INV	say that	NOM	RED-bottle	DM

(IU15) **(IU16)** **(IU17)**

kudra	marakace	ituku	la	ituku	**kuini**	**iyaw**	si	ka-tuase
that.INV	violently	jump	then	jump	that	frog	and	STAT.IRR-leave

(IU18)

ubere	taiya	la	ngukay	kuini ginganga	kai	ka-ni-caeba-ane
run	DM	then	from	that bottle.mouth	NEG	real-NMLZ-cover-NMLZ

'No matter how small the bottle was, **that frog** jumped up and down, and finally ran away from the uncovered bottle mouth.'

In IU11 of (13), the speaker uses the topicalized definite NP *kuini iyaw* "that frog" to reintroduce the referent of the frog, which is mentioned two sentences before in IU8 , back to the discourse. And this reintroduced referent of the frog continues to be the topic of discourse in IU 13 and IU 17 of (13).

(14) **RukaiNr-frog_Kainguane 144-172**

*[143]La si tautautau kudra ki [144]taupunguini.[145]Saecenge kuini ki laungu kudra yaie, [146]**kudra** kai wathingathingale laka laungu ki salaungane. (Seven sentences later) [168]Sakela kuini kalrevesane yaie, [169]la ngituluku kuini, [170]la katuase ngithapilri kikay vavalake si. [171]Tuverevere kikay taupungu, [172]la tuverevere mua kavay ki aclay kudra.*

(IU143) **(IU144)**

La	si	tau-tau-tau	kudra	ki	taupungu**=ini**.
then	and	RED-RED-shout	that.INV	OBL	dog=3S.BG

(IU145)

Sae-cenge=ϕ	kuini	ki	laungu	kudra	yaie,
when-touch	that	OBL	horn	that.INV	TOP

'His (The boy's) dog kept barking.'

doi:10.6342/NTU202302199

(IU146)

kudra	kai	wa-thinga-thingale	laka	laungu	ki	salaungane.
that.INV	NEG	ACT-RLS-RED-know exactly	horn	GEN	goat	

'When the child touched it, he didn't know it was the goat's horn.'

……

……

(Seven sentences later)

(168)

Sa-kela	kuini	kalrevesane	yaie,
when-come	that.VIS.PROX	cliff	TOP

(169)

la	ngi-tuluku	kuini,
then	REFL-frightened	that.VIS.PROX

(170)

la	ka-tuase	ngi-thapilri	kikay	vavalake	si
then	STAT-IRR	REFL-spring	this	child	child

'Then the goat came to the cliff; it was so frightened that it stopped. Then the child was sprung from the goat.'

(IU171)

Tu-verevere	kikay	taupungu,
Toward-fall	this	dog

(IU172)

la	tu-verevere	mu-a		
then	toward-fall	go-RLS		
kavay	ki	aclay	kudra	
that.VIS.DIST	OBL	water	that.INV	

'The dog fell down into water.'

Similarly, the topicalized demonstrative pronoun *kudra* "that" in IU146 of (14) refers to the boy in the preceding sentence (IU144; in the form of genitive), remaining in the context for seven more sentences. And finally, the topic of the context changes into

doi:10.6342/NTU202302199

kikay taupungu “this dog” (IU171).

Givón (1983) claims that in a language with pragmatically controlled word-order flexibility, the preverbal position is relatively discontinuous. Accordingly, the value of referential distance for the topicalized NP is predicted to be higher while its value of persistence might be lower than other untopicalized NPs. Nonetheless, based on our findings shown in Table 3.3, the prediction above is only half-right.

Table 3.3 Measures of Topic Continuity for Topicalization

		Referential Distance	Persistence
Demonstrative Pronoun	**Untopicalized**	1.0	1.4
	Topicalized	1.3	2
Definite Noun	**Untopicalized**	1.8	1.2
	Topicalized	2.1	2.2

In Table 3.3, the values of referential distance and persistence for the topicalized NPs are both higher than those for the untopicalized ones. The high value of referential distance implies that the device of topicalization is often used to reintroduce an old topic, which was last mentioned in several sentences ago) back into the current context. On the other hand, the high value of persistence indicates that such reintroduced topic can remain in the discourse longer than any untopicalized NPs. With such features, topicalization seems hard to be included in Givón’s scale of topic continuity. Even so, there is no denying that topicalization is still an essential syntactic coding device in discourse of Budai Rukai.

3.5 Case Roles and Topic Continuity

In this section, our focus turns from various syntactic coding devices to different case roles. According to the hierarchy of case roles proposed by Givón (1976), the case

doi:10.6342/NTU202302199

roles are ranked based on their importance in the sentence as topics, with the subject ranking higher than the direct object which in turn ranks higher than indirect object and oblique as illustrated in (15).

(15) **The Topic Continuity of Different Case Roles (Givón 1976)**

High Subject > Direct object > Others **Low**

There are three main types of case roles found in our texts, including subject, oblique and genitive. Table 3.4 below gives overall average measures of referential distance and persistence for NPs in each case role, showing that the genitive is the most continuous case role and the subject is the second most continuous one, while the oblique possesses the least degree of topic continuity.

Table 3.4 Measures of Topic Continuity by Case Roles

	Referential Distance	Persistence
Genitive	1.3	2.3
Subject	1.5	2.2
Oblique	1.8	0.7

Although Givón doesn't include "genitive" in his hierarchy of case role, it has been found in other study that genitive is easily the most continuous case role in written English (Brown 1983). The reason why the genitive can be so continuous in the discourse might be that it often serves as a particular kind of bridge to link one NP to another NP. For example, the genitive *ini* "the dog's" in IU 33 of (16) is a linking between the subject *kikay taupungu* "this dog" (IU30) and the oblique *kay aulru* "this head" (IU33). Furthermore, it also links to the covert subject "the dog" (IU34), which enables the topic "the dog" to move from one sentence to the next smoothly.

doi:10.6342/NTU202302199

(16) **RukaiNr-frog_Legeai IU30-37**

*[30]Kikay taupungu yaie, [31]law yakay adringi ki didilungu la iya, [32]si puapulratuku,[33]kay aulru**ini** ki didilungu. [34]La kai maka=ϕ [35]adauthanenga ikay ki didilungu [36]la mua ki lribange si [37]e tautautau.*

(IU30)			**(IU31)**				
kikai	taupungu	yaie,	law	i-a-kay	adringi	ki	di-dilungu
this	dog	TOP	seem.like	LOC-RLS-this	inside	OBL	RED-jar

	(IU32)			**(IU33)**			
la	iya,	si	pu-a-pulratuku	kay	aulru**=ini**	ki	di-dilungu
then	say	and	put-RLS-put.in	this	head=3S.BG	OBL	RED-jar

"This dog (thought) might be in the bottle, and put **its** head into the bottle."

(IU34)			**(IU35)**			**(IU36)**	
la	kai	maka(=ϕ)	adauthane=nga	i-kay	ki	di-dilungu	la
then	NEG	can(=3S.BN)	pull.out=PFV	LOC-this	OBL	RED-jar	then

				(IU37)	
mu-a	ki	lribange	si	e	tau-tau-tau.
go-RLS	OBL	window	and	FIL	RED-RED-yell

"Then, (the dog) couldn't pull out (its head) and then went to the window, yelling."

As a result, in Budai Rukai, the hierarchy of case role can be adapted as (17) based on Givón's original version.

(17) **The Topic Continuity of Different Case Roles in Budai Rukai**

High Genitive > Subject > Oblique **Low**

3.6 Humanness and Topic Continuity

Besides syntactic coding device and case role, humanness also has some impacts on the topic continuity. It is found that human NPs are much more continuous than non-

doi:10.6342/NTU202302199

human NPs since speakers as human tend to speak more about humans while non-human NPs are, in most cases, just temporary focus or background information and therefore dropped relatively quickly (Pu, 1989).

Table 3.5 below shows us the values of referential distance and persistence for the human and non-human NPs in Budai Rukai. The value of referential distance for human NPs (1.5) is lower than that for non-human ones (1.9), indicating human NPs is slightly more continuous. On the other hand, in terms of persistence, the value for human NPs is a lot higher than that for non-human ones, implying that the human NPs have the tendency to remain in the following context for a longer time.

Table 3.5 Measures of Topic Continuity by Humanness

	Referential Distance	Persistence
Human	1.5	2.6
Non-human	1.9	0.8

According to our results above, the topic continuity of humanness in Budai Rukai can be concluded as in (18), coinciding with our prediction.

(18) **The Topic Continuity of Humanness in Budai Rukai**

High Human > Non-human **Low**

However, it is necessary to stress the importance of the content of the texts that are investigated. Our data here is primarily about culture sharings and folk tales, in which the characters included are mostly human. If texts with more non-human NPs are investigated, then results concerning humanness and topic continuity might differ.

doi:10.6342/NTU202302199

3.7 Summary

In this chapter, we have examined the topic continuity of NPs in discourse of Budai Rukai by a quantitative method proposed by Givón, in which the measurements of referential distance and persistence are analyzed from several different aspects, inclusive of syntactic coding devices, word order, case roles, and humanness.

According to our statistical results, the topic continuity of syntactic coding devices in Budai Rukai generally fits Givón's hypothesis, with zero anaphora and pronoun serving as the most continuous coding devices while definite NP and indefinite NP relatively less continuous.

In addition, we found topicalization another interesting but less discussed issue in the discussion of topic continuity. As NPs in the preverbal position often refer to reintroduced referents, the value of referential distance for topicalized NPs tends to be higher. Topicalized NPs can also be continued longer in the later discourse, so its value of persistence is higher than untopicalized ones as well.

Finally, we have also investigated the topic continuity of case role and humanness. In terms of case role, the subject is found a more continuous case role than the oblique. Our results show that the genitive is the most continuous case role due to its bridging function. As for humanness, the human NP is more continuous than the non-human one in our data as we have predicted.

doi:10.6342/NTU202302199

Chapter 4
The Givenness Hierarchy in Budai Rukai

4.1 Introduction

In the previous chapter, we examined the topic continuity of NPs in Budai Rukai, finding that NPs in this language are presented by four various types of syntactic coding devices, including zero anaphora, pronoun, definite NP and indefinite NP. Besides, it is concluded that different grammatical coding devices in this language reveal different degrees of topic continuity as Givón (1983) predicted in his hypothesis.

The research of reference has been a focus in the linguistic and psychological literature. Some linguists (Gundel, Hedberg and Zacharski 1993) especially put their emphasis on the distribution and understanding of different referential expressions in natural language discourse. They proposed the Givenness Hierarchy, in which they claim that different referential forms in the discourse are conventionally able to signal different cognitive statuses of the speaker.

In this chapter, our aim is to investigate how the speaker's cognitive status is implied by the use of syntactic coding devices in Budai Rukai. To fulfill this goal, the Givenness Hierarchy (Gundel, Hedberg & Zacharski 1993; Gundel et al. 2006; Hedberg 2014) is adopted to measure the referential givenness of all noun phrases in discourse of Budai Rukai.

This chapter is organized as follows: Section 3.2 introduces the Gundel et al.'s Givenness Hierarchy (1993); Section 3.3 takes a quick look at the narrated data and introduces the way we analyze such data; Section 3.4 discusses the referential properties of NPs in Budai Rukai based on the Givenness Hierarchy; Section 3.5 summarizes and concludes the chapter.

doi:10.6342/NTU202302199

4.2 The Givenness Hierarchy

The Givenness Hierarchy as given in (1) was first proposed by Gundel, Hedberg, and Zacharski (1993), and used to show multiple degrees of relation between the referring form and cognitive status of the speaker. In the Givenness Hierarchy, there are six cognitive statuses, each of which refers to a certain degree of referential givenness that a given referent possesses.

(1) **The Givenness Hierarchy (Gundel, Hedberg & Zacharski 1993):**
In focus > Activated > Familiar > Uniquely Identifiable > Referential > Type Identifiable

As shown above, the six cognitive statuses are presented in a descending order, with the leftmost referring to the highest degree of referential givenness, while the rightmost the lowest degree. That is, if a referent is *"in focus"*, it is right in the center of the speaker's and the hearer's attention; conversely, if a referent is *"type identifiable"*, then it is probably outside of the memory of the speaker and the hearer. It is worth noting that each status entails all lower statuses, but not vice versa. For example, an noun phrase that is *in focus* is, by default, *activated*, *familiar*, *uniquely identifiable*, *referential*, and *type identifiable* as well, but not all *familiar* noun phrases are *activated* or *in focus*. These six statuses are characterized with English examples in discourses as below (Gundel, Hedberg & Zacharski 1993; Gundel et al. 2006; Hedberg 2014). The underlined referents in the English examples belong to the statuses we are discussing.

1. IN FOCUS: The referent is at the current center of attention of the speaker's or the hearer's short-term memory. A referent is IN FOCUS if it meets at least one of the following criteria:
   (1) It is the interpretation of the main clause subject or the syntactic topic (inclusive of topicalized or dislocated phrases) in the immediately preceding

doi:10.6342/NTU202302199

sentence.

a. Midge pushed thick, wiry black hair back from her square forehead with a sturdy brown arm. Nothing unsubstantial or fairylike about her. (From Murder after Hours, Agatha Christie)

(2) It is part of the interpretation of previous part of the same sentence.

b. You can wear my scarf if you can find it.

c. If you stand on this chair, the chair will break.

(3) It is the interpretation of the syntactic focus of the immediately preceding clause (i.e., postcopular position of a cleft or existential sentence).

d. There was a mouse on the table. It was very large.

e. It was the dog that Bill was afraid of. He was very large.

(4) It is a higher level topic that is part of the interpretation of the preceding clause (whether it is overtly mentioned there or not).

f. The kitchen has a new countertops and a beautiful tile floor. There's also a big walk-through closet. Would you like to take a look at it? Both the kitchen (criterion 4) and the closet (criterion 3) are in focus.

(5) It is part of the interpretation of each of the two immediately preceding clauses.

g. It was the dog that Bill was afraid of. Small animals didn't usually frighten Bill. He was very large.

(6) It is the event denoted by the immediately preceding sentence.

h. John fell off his bike. This/it happened yesterday.

2. ACTIVATED: The referent is represented in current short-term memory, and it may have been retrieved from long-term memory, or it may arise from the immediate linguistic or extralinguistic context. A referent is ACTIVATED if it meets one of the following criteria:

(1) It is part of the interpretation of one of the immediately preceding two sentences.

doi:10.6342/NTU202302199

a. Central to the case was a Lewinsky-Tripp conversation that Mrs. Tripp taped on Dec. 22, 1997. This was the last talk between the two women that Mrs. Tripp recorded.

(2) It is something in the immediate spatial-temporal context that is activated by means of a simultaneous gesture or eye gaze.

b. (looking at the wrench) Please hand me that.

(3) It is a proposition, fact, or speech act associated with the eventuality (event or state) denoted by the immediately preceding sentence(s).

c. A. John fell off his bike.

B. That's not true.

3. FAMILIAR: The intended referent is able to be identified by the hearer, for it is represented in either short-term or long-term memory. A referent is FAMILIAR if it meets one of the following criteria:

(1) It was mentioned at any time previously in the discourse.

a. A Philippine Airlines jet with 290 people aboard was hijacked today by a man who took everyone's money and then parachuted to the ground outside Manila's airport and the passengers were let off safely. The jetliner left Davao City, in the southern Philippines, for the 90-minute flight to Manila with 278 passengers and 12 crew aboard, PAL said. The hijacker, wearing a blue ski mask and carrying a handgun…

(2) It can be assumed to be known by the hearer through cultural/encyclopedic knowledge or shared personal experience with the speaker.

b. If one takes a step back and looks at the rest of this week's music-group news, the situation looks bad for ugly, unpredictable rock 'n' roll: one of the most popular American rock bands of the 90's.

4. UNIQUELY IDENTIFIABLE: The intended can be identified by the hearer based

doi:10.6342/NTU202302199

on the nominal alone. A referent is UNIQUELY IDENTIFIABLE if it meets one of the following criteria:

(1) The referring form contains adequate descriptive/conceptual content to create a unique referent.

a. s: hello can I help you?

u: yeah I want t- I want to determine the maximum number of boxcars of oranges that I can get to Bath by 7 a.m. tomorrow morning so hm so I guess all the boxcars will have to go through oran- through Corning because that's where the orange juice factory is [Trains Corpus. Heeman & Allen 1995]

(3) A unique referent can be created via a 'bridging inference' by association with an already activated referent.

b. (Looking at a box) I think the bottom fell out.

5. REFERENTIAL: The particular referent is intentionally referred to by the speaker. A referent is REFERENTIAL, if it meets one of the following criteria:

(1) It is mentioned subsequently in the discourse.

a. When my youngest child was 3 or so, we were at a friend's house visiting and my friend was babysitting her infant nephew.

(2) It is evident from the context that the speaker intends to refer to some specific entity.

b. I want to tell you about this strange guy I saw today.

6. TYPE IDENTIFIABLE: The referent is able to be accessed as a type by the hearer without being referential. An interpretation is TYPE IDENTIFIABLE if the sense of the phrase (the descriptive/conceptual content it encodes) is understandable:

a. I don't have a VCR and neither does my neighbor.

To sum up, each status on the hierarchy is often manifested by the proper use of

doi:10.6342/NTU202302199

different forms. The relevant English forms for each status are listed in Table 4.1 below:

Table 4.1 The relevant forms for each status in the Givenness Hierarchy in English

In focus >	**Activated** >	**Familiar** >	**Uniquely Identifiable** >	**Referential** >	**Type Identifiable**
it	*that* *this* *that N*	*that N*	*the N*	*indefinite this N*	*a N*

4.3 Methodology and Data

In the previous section, we have gone through the six cognitive statuses of the Givenness Hierarchy. In this present study, our goal is to investigate all the noun phrases in the narrative data of Budai Rukai, and to see how the speakers' cognitive statuses are implied by the referring forms, or syntactic coding devices. At the same time, we also want to make sure whether the distribution of the referring forms follows the Givenness Hierarchy or not.

The data we utilize for analysis is the same eleven texts that we used in chapter three. This time, every noun phrase in these texts is classified into one of six corresponding cognitive statuses based on the guidelines pointed out by the Givenness Hierarchy Coding Protocol (Gundel et al. 2006; Hedberg 2014). Then, we examine the distribution of various syntactic coding devices in the six cognitive statuses. The results will be revealed and further discussed in the next section.

4.4 Cognitive Statuses and Syntactic Coding Devices in Budai Rukai

In this section, we will first present the statistical results concerning the cognitive statuses and the corresponding syntactic coding devices in Budai Rukai, and then discuss such results under every individual cognitive status.

doi:10.6342/NTU202302199

4.4.1 Results

Table 4.2 below shows the percentage of noun phrases distributing in six various cognitive statuses. We can see from this table that around half of the noun phrases (56%) in the data belong to *in focus* status. Furthermore, the noun phrases distribute almost evenly (13%) in the other cognitive statuses, except for *uniquely identifiable* and *type identifiable* statuses. The two lowest percentages of noun phrases are found in the cognitive statuses of *uniquely identifiable* (3%) and *type identifiable* (2%).

Table 4.2: The Percentage of NPs in Six Cognitive Statuses (Budai Rukai)

	Percentage
In focus	544 (56%)
Activated	125 (13%)
Familiar	123 (13%)
Uniquely Identifiable	34 (3%)
Referential	122 (13%)
Type Identifiable	24 (2%)
Sum	972 (100%)

We have known from Chapter Three that four major types of syntactic coding devices, including zero anaphora, pronoun, definite noun and indefinite noun, are used to code the referents in discourse of Budai Rukai. Besides, these four coding devices possess different degrees of topic continuity, with zero anaphora and pronoun serving as the most continuous coding devices while definite NP and indefinite NP relatively

doi:10.6342/NTU202302199

less continuous. As we have mentioned, if a NP is with higher topic continuity, it is easier for the speaker or the hearer to identify such referent in the discourse, which implies that such NP is at the center of the speaker's attention as well. Based on this, we predict that the coding devices with higher topic continuity is very likely to code the NPs that hold higher degrees of cognitive statuses, and vice versa.

Table 4.3 below shows us how noun phrases in various cognitive statuses of the speakers are manifested by the use of four noun phrase coding devices in Budai Rukai. Although almost all the four syntactic coding devices can be used to code noun phrases in six cognitive statuses, there is always one particular coding device that prevails other three ones in each cognitive status. More specifically, the *in focus* NPs are often in the form of zero anaphora (38%). The *activated*, *familiar,* and *uniquely identifiable* NPs are mainly coded as definite noun (62%; 68%; 74%). As for the NPs in *referential* status, half of them are coded as definite noun (50%) and the others are coded as indefinite noun (49%). And, most of the *type identifiable* NPs are indefinite noun (96%). Much more details will be further discussed in following subsections.

Table 4.3: Cognitive Statuses and Syntactic Coding Devices in Budai Rukai

	Zero anaphora	Pronoun	Definite noun	Indefinite noun	Sum
In focus	**205 (38%)**	**138 (25%)**	**169 (31%)**	32 (6%)	544 (100%)
activated	19 (15%)	12 (10%)	**78 (62%)**	16 (13%)	125 (100%)
familiar	7 (6%)	7 (6%)	**84 (68%)**	25 (20%)	123 (100%)
Uniquely Identifiable	0 (0%)	2 (6%)	**25 (74%)**	7 (20%)	34 (100%)
Referential	0 (0%)	1 (1%)	**61 (50%)**	**60 (49%)**	122 (100%)
Type Identifiable	0 (0%)	0 (0%)	1 (4%)	**23 (96%)**	24 (100%)

doi:10.6342/NTU202302199

4.4.2 In Focus

The *in focus* referent is claimed to be at the center of attention of the speaker's and the hearer's short-term memory. According to our data (see Table 4.2), more than half of the NPs, with the total of 544 tokens, or 56%, fall into *in focus* status. The *in focus* NPs that have been found in our data mainly abide by 2 of 6 standards stated in The Givenness Hierarchy Coding Protocol (Gundel et al. 2006; Hedberg 2014): the majority of them are the subjects or the topics in the immediately previous sentence, and the others are introduced by the existential or cleft sentence in the immediately previous sentence. Very interestingly, the *in focus* NPs in Budai Rukai can be seen in almost all the syntactic coding devices. Even so, the zero anaphora NPs (38%) are seen slightly more frequently than definite nouns (31%) and pronouns (25%). With only 6%, indefinite nouns are seen the least. Several examples of *in focus* NPs are illustrated as follows:

(1) **RukaiNr-frog_Salabu IU133-139**

*[133]Sa e mubiabilanga=**ø** yaie, [134]e [135]ala ikay kudra ki [136]babiabila ki [137]aciaacilay yaie yakay **kudra [138]e latakurauru** si.[139]La drele=**ø kuini ki latakurauru** si ala ikay kudra takurauruini si.*

(IU133)				**(IU134)**		**(IU135)**	
sa	e	mu-biabila=nga**(=ø)**		yaie,	e	ala	ikay
when	FIL	go-bank=PFV**(=3P.BN)**		TOP	FIL	then	LOC-this

	(IU136)		**(IU137)**		
kudra	ki	babiabila	ki	acia-acilay	yaie	i-a-kay
that.INV	OBL	bank	OBL	RED-water	TOP	LOC-RLS-this

(IU138)			
kudra	**e**	**la-takurauru**	si
that.INV	FIL	P-frog	and

doi:10.6342/NTU202302199

“When **(they)** went to the bank, there were **(many) frogs** at the bank of water.”

(IU139)

la	drele(=**Ø**)	**kuini**	**ki**	**la-takurauru**	si	ala	i-kay	
then	see**(=3P.BN)**	that	OBL	P-frog	and	then	LOC-this	

kudra	takurauru=ini	si.
that.INV	frog=3S.BG	and

“**(They)** saw **some frogs**, and there was his (the boy’s) frog.”

In IU 139 of (1), the subject “they”, in the form of zero anaphora, is counted as an *in focus* NP, since it refers to another covert subject “they” of the previous sentence (in IU 133). The subject of a sentence is often the most prominent element and the speaker wants to discuss more about it. Therefore, it must be the primary focus of the speaker and very likely to be the subject of later speech. As stated in Chapter 3, the third person nominative pronouns in Budai Rukai are coded as zero form, which explains why a great percentage of *in focus* NPs is in the form of zero anaphora.

Furthermore, the referent that is introduced by the previous existential sentence is in the *in focus* status as well. In IU 139 of (1), the definite noun *kuini ki latakurauru* ‘those frogs’ is in the status of *in focus* because it is introduced by the previous existential sentence in IU 137-138: *Yakay kudra latakurauru.* “There are many frogs.” Existential structure is used to refer to the presence of a certain referent in a particular time or place. And the introduction of the referent by the existential sentence is probably the speaker’s preparation for more discussion on such referent. That is why this kind of referent is considered to be *in focus* cognitive status in the later discourse.

Aside from zero anaphora and definite noun, personal clitic pronoun is another common coding device for the *in focus* NPs, accounting for 25% of them. In Chapter 3, we have concluded that genitive, the most continuous case role in Budai Rukai, is often

doi:10.6342/NTU202302199

used as one kind of bridge to link one NP to another NP in the following sentence. With this "bridging" function, a genitive pronoun helps transfer the speaker's attention from one topic to another. For instance, in IU 27 and IU 29 of (2), two third person singular genitive pronouns *ini* "his" are in the status of *in focus,* since they both refer to the zero form subject "he" in IU 25. The use of such genitive pronoun changes the speaker's focus from "the boy" to *kudra talrupunu**ini*** "his hat" and *kuini lrikil**ini*** "his bicycle."

(2) **RukaiNr-pear_Legeai IU23-29**

*[23]La sa madradresenge=ϕ yaie, [24]e [25]naw=ϕ dreeleana kikay. [26]Malralribate la iya, [27]la mualrane kudra talrupunu**ini** [28]lau kilrikilri ucucusu ki lrenege [29]kuini e kuini lrikil**ini** si muadreke.*

(IU23)				**(IU24)**	**(IU25)**
la	sa	ma-dradresenge(=ϕ)	yaie,	e	naw(=**ϕ**)
then	when	RECP-meet(=3P.BN)	TOP	FIL	want(=3S.BN)

dreele-ana	kikay.
see-first	this

"Then, when (they) met each other, (he) wanted to see this person first."

(IU26)			**(IU27)**		
ma-lra-lribate(=ϕ)	la	iya	la	mu-alra-ane	kudra
RECP-RED-pass(=3P.BN)	then	say	then	go-take-NMLZ	that.INV

	(IU28)				
talrupunu**=ini**	lau	kilrikilri	ucucusu	ki	lrenege
hat**=3S.BG**	then	trap	bump	OBL	stone

(IU29)					
kuini	e	kuini	lrikil**=ini**	si	mua-dreke
that.VIS.PROX	FIL	that.VIS.PROX	bicycle**=3S.BG**	and	go-fall

"(When) (they) passed each other, **his** hat fell and then he trapped and bumped **his** bicycle against a stone."

doi:10.6342/NTU202302199

4.4.3 Activated

The referents in the *activated* status may be in the speaker's current short-term memory, or may be retrieved from long-term memory of the speaker. According to the Givenness Hierarchy Coding Protocol (Gundel et al. 2006; Hedberg 2014), an *activated* referent is the one that is mentioned (not necessarily as subject) within two previous sentences. A total of 125 tokens, or 13%, of *activated* NPs are discovered in our data (see Table 4.2). In addition, more than half of the *activated* NPs (62%) are coded as definite noun (see Table 4.3). Considering the example (3)**,** the reference of 'the horn of the goat' is brought into the discourse in IU141 (*kuini ki laungu*). Then, the speaker switches the topic from the horn of the goat to the child's dog (*kudra ki taupunguini* "his dog") in the next sentence (IU142-IU143). One sentence later, the speaker retrieved "the horn of the goat" from the hearer's short-term memory by the use of *activated* definite noun *kuini ki laungu* in IU145. Often used for the speaker to draw the hearer's attention back to a just-mentioned referent, an *activated* NP can be viewed as a mechanism for rapid topic switches (Yang 2019).

(3) RukaiNr-frog_Kainguane IU139-146

[139]Kuini vavalake yaie, [140]sai kuini ki tarutugutugu, [141]kiaulau luka angatu amia si laecenge ***kuini ki laungu ki salaungane*** *taiya. [142]La si tautautau kudra ki [143]taupunguini. [144]Eh [145]saecenge* ***kuini ki laungu kudra*** *yaie, [146]kudra kai wathingathingale laka laungu ki salaungane.*

(IU139)				**(IU140)**		
kuini	vavalake	yaie,	sai	kuini	ki	tarutugu-tugu,
that	child	TOP	when	that.VIS.PROX	OBL	rock-RED

(IU141)

kiaulau	luka	angatu	a-iya	si	laecenge	**kuini**	**ki**	**laungu**
think.of	that	tree	RLS-say	and	touch	**that.VIS.PROX**	**OBL**	**horn**

doi:10.6342/NTU202302199

ki **salaungane** taiya
GEN **goat** DM
"That child on the rock thought of **that horn of the goat** as the tree and touched it."

(IU142) **(IU143)**

la	si	tau-tau-tau	kudra	ki	taupungu=ini.
then	and	RED-RED-shout	that.INV	OBL	dog=3S.BG

"His dog kept barking."

(IU144) **(IU145)**

eh	sa-ecenge=ϕ	**kuini**	**ki**	**laungu**	**kudra**	yaie,
FIL	when-touch	**that.VIS.PROX**	**OBL**	**horn**	**that.INV**	TOP

(IU146)

kudra	kai	wa-thinga-thingale	laka	laungu	ki	salaungane.
that.INV	NEG	ACT-RLS-RED-know	exactly	horn	GEN	goat

"When the child touched **that horn**, he didn't know it was exactly the goat's horn."

4.4.4 Familiar

Familiar referents are represented in either short-term or long-term memory and should have appeared at least once in the previous discourse. A total number of 123 tokens, or 13%, of *familiar* NPs are found in our data (see Table 4.2), which is very close to that of *activated* NPs (125; 13%). Similar to the *activated* NPs, *familiar* NPs in our data often refer to the referent that is brought back to the cognitive focus of the hearer by topic switching. Also, *familiar* NPs in Budai Rukai are primarily in the form of definite noun (68%) just like *activated* NPs (see Table 4.3). An example including *familiar* NP is given below:

(4) **RukaiNr-frog_Salrabu IU94-114**
[94]Ala kuini [95]kuici yaie, [96]la [97]pasana daidaisi kuini ki vavalake si [98]e ikay kudra

doi:10.6342/NTU202302199

*ki e [99]e kudra sakela **kuini taupunguini** yaie, [100]la ikay kudra lrenege matuatuase si, [101]la mu lrilrikudru si. (three sentences later) [111]La ikay kuini si tautautau **kuini ki taupungini** si silasilape si [112]kudra kuici, [113]la ibelenge ki angatu [114]si si sia-vavava iniane.*

(IU94)		**(IU95)**		**(IU96)**	**(IU97)**	
ala	kuini	kuici	yaie,	la	pasana	dai-daisi
then	that.VIS.PROX	owl	TOP	then	want.to	RED-grasp

					(IU98)	
kuini	ki	vavalake	si	e	i-kay	kudra
that.VIS.PROX	OBL	child	and	FIL	LOC-this	that.INV

(IU99)						
ki	e	kudra	sa-kela	**kuini**	**taupungu=ini**	yaie
OBL	FIL	that.INV	when-come	**that.VIS.PROX**	**dog=3S.BG**	TOP

(IU100)					
la	i-kay	kudra	lrenege	ma-tua-tuase	si
then	LOC-this	that.INV	stone	STAT.RLS-RED-leave	and

(IU101)			
la	mu	lrilrikudru	si.
then	go	behind	and

"Then this owl wants to catch this child; when his dog comes, there is a stone rolling, and it goes to the (stone's) back."

……
……
(Three sentences later)

(IU111)					
la	i-kay	kuini	si	tau-tautau	**kuini**
then	LOC-this	that.VIS.PROX	and	RED-call	**that.VIS.PROX**

						(IU112)
ki	**taupungu=ini**	si	sila-silape	si	kudra	kuici

doi:10.6342/NTU202302199

OBL dog=3S.BG and RED-find and that.INV owl
(IU113) **(IU114)**
la i-belenge ki angatu si si sia-<va>-vava
then LOC-up GEN tree and and <RED>-watch
iniane.
3S.FO

"(The boy) is calling and searching his fog, and then the owl on the tree keeps watching him."

In IU 99 of (4), the narrator has mentioned the boy's dog. While, in following IUs, the narrator changes the focus to discuss about the interaction between the boy and the deer. It is not until IU 111 (approximately after 3 sentences) does the narrator draw the hearer's attention back to "the boy's dog" with the definite noun, *kuini ki taupunguini*.

4.4.5 Uniquely Identifiable

The referent in *uniquely identifiable* status is characterized as one that shouldn't have been mentioned explicitly in the previous discourse, but generally must be inferred by the hearer via a "bridging inference to an already activated referent" or "adequate descriptive/ conceptual content" (Gundel et al. 2006; Hedberg 2014). Much more cognitive efforts are needed for the speaker and the hearer to process the *uniquely identifiable* referent, making the referent in such status less in the discourse. In our data, only 34 tokens, or 3%, of *uniquely identifiable* NPs are discovered. The *uniquely identifiable* NPs in our discourse data are all created by the "bridging inference". An associated example is offered below:

(5) **RukaiNr-frog_Legeai IU91-95**
[91]Tuverevere ucakena, [92]kuini taupungu la pauruu ***kudra****, [93]****sigu*** *si, la kisaladhaladhanga si [94]la kilangelangedre* ***kuini ki la lasigu****.*

doi:10.6342/NTU202302199

(IU91) **(IU92)**

tu-verevere	u-cakena,	kuini		taupungu	la	pa-ururu
TU-throw	go-ground	that.VIS.PROX		dog	then	CAUS-fall

(IU93) **(IU94)**

kudra	**sigu**	si	la	ki-saladhaladha=nga	si	la
that.INV	**bee.hive**	and	then	PASS-pursue=PFV	and	then

ki-lange-langedre	**kuini**	**ki**	**la**	**la-sigu**
PASS-RED-sting	**that.VIS.PROX**	**OBL**	**then**	**P-bee**

"(The beehive) dropped; that dog made **that beehive** drop, and the dog was chased and stung by **those bees**."

There are two tokens of *sigu* mentioned in the example (5). The former one in IU 93 means "the beehive" while the latter one in IU 94 means "the bees." Although the word *kuini ki la-sigu* "those bees" in IU 94 have never been mentioned previously, a highly associated referent *kudra sigu* 'that beehive' in IU 93 can serve as a bridge to introduce such definite noun ("those bees") into the context. Accordingly, the referent "those bees" in IU 94 is a typical *uniquely identifiable* NP in our data.

Additionally, to trigger the "bridging effect", the referent which the *uniquely identifiable* NP have a bridging inferential relation with should belong to at least *familiar* status. For instance, *kudra sigu* 'the beehive' in the example above is in *familiar* status, which is pre-existing around eight sentences before. That is, a never-activated referent is unable to be the "bridge" to introduce a *uniquely identifiable* NP.

4.4.6 Referential

A *referential* referent is intentionally referred to by the speaker and has to be subsequently mentioned in the following discourse. In our data, 122 tokens, or 13%, of referential NPs are discovered. Half of them (49%) are coded as definite noun and the

doi:10.6342/NTU202302199

other half of them (50%) are coded as indefinite noun. Additionally, the existential construction is the most often-used mechanism to introduce such entity into the discourse so as to prepare it for subsequent mention and topicality. In the example (6), the existential verb "*ikay*" is utilized to first introduce the definite noun *kudra vavalake* 'the child' into the discourse in IU 1. The child is the main character, being the subject and the topic in the major parts of story as well.

(6) **RukaiNr-frog_Salabu IU1-5**

*[1]Kudra nadruma **yakay kudra vavalake** si [2]la katalame [3]laulapu kudra kayki lawaudridripi.*

(IU1)

kudra	nadruma	**i-a-kay**	**kudra**	**vavalake**	si
that.VIS.PROX	before	LOC-RLS-this	that.VIS.PROX	child	and

(IU2)

la	ka-talame	lau-lapu	kudra	kay	ki
then	STAT.IRR-like	RED-raise	that.VIS.PROX	this	OBL

la-waudridripi
P-animal

"Long time ago, there was **a child** and then (he) liked to raise animals."

Another example of *referential* NP is given in (7), in which an indefinite noun *ku tadulru* "three (people)" is introduced into the context by the preceding existential verb. After its first introduction, such *referential* NP continues to be the subjects of two successive sentences.

(7) **RukaiNr-pear_Legeai IU32-33**

*[32]La kaynganay la **ikay ku e** [33]**tadulru ku.***

doi:10.6342/NTU202302199

(IU32)						(IU33)	
la	kaynganay	la	**i-kay**	**ku**	**e**	**tadulru**	ku
then	come	then	LOC-this	OBL	FIL	three.HUM	FS

"Then, there wre **three (people)** coming."

4.4.7 Type Identifiable

A *type identifiable* referent is able to be accessed as a type by the hearer without being referential. It is typically grammatically indefinite, such as *a frog* or *a kid* in English. In our data, there are 24 tokens, 2%, of type identifiable NPs, which is the fewest among the six cognitive statuses (see Table 4.2). Unsurprisingly, most of them (98%) are in the form of indefinite noun. For instance, the indefinite noun *ku becenge* "millet" in unit 25 of (8) doesn't refer to any specific millet but a general type of millet.

(8) **RukaiNr-becenge_Tagas IU25-26**

*[25]Lu ngukalra **ku becenge** [26]pakubalriyu.*

(IU25)				(IU26)
lu	ngu-kalra	**ku**	**becenge**	pakubalriyu
when	take-be.many	**OBL**	**millet**	PN

"If (they) take much **millet**, they will do the Pakubalriyu."

4.5 Overall Discussion and Concluding Remarks

With our statistical results and explanations with examples in the sections above, we have already gained a better understanding of the relation between cognitive statuses and syntactic coding devices in Budai Rukai. We found that the Givenness Hierarchy of NPs also works in Budai Rukai, as indicated in Table 4.4 below.

doi:10.6342/NTU202302199

Table 4.4 The Most Relevant Forms for Each Cognitive Status in Budai Rukai

In focus >	**Activated** >	**Familiar** >	**Uniquely Identifiable** >	**Referential** >	**Type Identifiable**
Ø pronoun definite N	definite N	definite N	definite N	indefinite N definite N	indefinite N

In Table 4.4, we can see that in Budai Rukai the referent with the highest degree of givenness, or *in focus* referent, tends to be coded by the syntactic coding devices of zero anaphora, pronoun, and definite noun. On the other hand, if a referent possesses lower degree of givenness, such as *referential* or t*ype identifiable* referent, it usually takes indefinite noun as its syntactic coding. It is noted that definite noun serves as a common type of syntactic coding device in Budai Rukai, adopted to express the referents in almost all kinds of cognitive status except for *Type Identifiable.* Especially, it is found that the *activated*, *familiar*, and *uniquely identifiable* referents are all mainly coded as definite N. Even so, some distinctions among these cognitive statuses are still found in Budai Rukai, depending on the demonstratives adopted to form various types of definite nouns. In other words, the speakers prefer different demonstratives when using definite nouns to code the referents in different cognitive statuses. For instance, *kuini* is very often seen on the *in focus*, *activated*, *familiar* and *uniquely identifiable* NPs, while *referential* NPs are coded by the demonstrative *kudra*. However, the choice of the demonstrative is not complementary, which means that all the demonstratives in Budai Rukai, including *kuini, kudra, kay, and kavay*, are able to occupy the position of demonstrative on the definite noun, only with different frequencies. Besides, since the frequencies of *kuini* and *kudra* are much higher than those of *kay* and *kavay*, our following discussion will focus on *kuini* and *kudra*. To be more specific, it seems that

doi:10.6342/NTU202302199

all the demonstratives can be the possible candidate to code the referent in each cognitive status, but *kuini* plays a more salient role on coding the referents with the higher cognitive statuses, from *in foucus* to *uniquely identifiable*; that is, more than 50% of the definite nouns in these four statuses are formed by this demonstrative, *kuini*. Although in these four cognitive statuses, the frequency of *kudra* is not as salient as that of *kuini* is, an interesting tendency is still found: the percentage of *kudra* increases gradually (ranging from 15% to 29%) as the cognitive status becomes higher and higher. And, with the degree of cognitive status becoming lower, the coding role of *kuini* is less salient **(**dropping abruptly from 60% to 20%.) Instead, the demonstrative, *kudra*, which is less salient in the previous four statuses, takes over the primary coding role, accounting for 60% of all the demonstratives. Accordingly, in Budai Rukai, the four highest cognitive statuses, inclusive of *in foucs, activated, familiar*, and *uniquely identifiable,* can be distinguished from the relatively lower cognitive status, *referential*, by their different preferences to the coding demonstratives. Even so, these four highest cognitive statuses themselves cannot be easily distinguished from one another since all of them are primarily coded by the demonstrative, *kuini*.

This phenomenon might be associated with the semantic features of various demonstratives. As we have mentioned in 2.2, different demonstratives can indicate different semantic features such as “visibilities” and “distances.” Based on these semantic features, since the demonstrative *kuini* refers to the referent that is in the speaker’s sight and is therefore closer to the speaker, this can explain why such demonstrative is very often used to code the definite nouns in the higher degrees of cognitive status. On the other hand, the demonstrative, *kudra*, implies that the referent is not only far away from the speaker, but also out of the sight of the speaker, which explains its high frequency on the definite nouns in lower cognitive status.

What’s more, the choice of referential form for each cognitive status is found to

doi:10.6342/NTU202302199

have something to do with topic continuity. That is, the referent in the highest degree of cognitive status (*in focus* referent) is coded by the two most continuous coding devices in Budai Rukai (zero anaphora and pronoun), whereas the referent in the lowest degree of cognitive status (*type identifiable* referent) is coded by the least continuous coding device in Budai Rukai (indefinite N). Besides, is found to. Despite the slightly weaker topic continuity of definite noun, it is able to code referents in almost all cognitive statuses except for type identifiable status. The interaction between the scale of topic continuity and the givenness hierarchy in Budai Rukai is summarized as follows in Table 4.5 below.

All in all, we can conclude that the cognitive status of referent can play a decisive role in the selection of syntactic coding devices on NPs in Budai Rukai.

Table 4.5 Interaction between Topic Continuity and Cognitive Status

Highest Cognitive Status ⟷ **Lowest Cognitive Status**					
In focus >	**Activated** >	**Familiar** >	**Uniquely Identifiable** >	**Referential** >	**Type Identifiable**
Ø pronoun definite N	definite N	definite N	definite N	indefinite N definite N	indefinite N
Most Continuous ⟷ **Least Continuous**					

doi:10.6342/NTU202302199

Chapter 5

Conclusion

5.1 Major Findings

In this present study, we have examined all the syntactic coding devices in the discourse of Budai Rukai, with the aim to respond to the following three research questions, repeated in (1), (2) and (3) from Chapter 1. To recapitulate this thesis, we will discuss these research questions one by one, and provide the main findings of all these questions in this section.

(1) **Research Question One**: What factors play essential roles in the selection of syntactic coding devices on NPs in Budai Rukai?

Reading the discourse data of Budai Rukai, we found that the same referent can be expressed in different types of syntactic coding devices, which raises our curiosity in the mechanism of choosing referential forms. Givón's and Gundel's research on the referential expressions with functional methods have inspired us a lot.

According to their findings, the referential choice is associated with not only the quantity of the syntactic coding devices but also the cognitive status of certain referents. With their research methods to examine our data, it is found that the quantity of syntactic coding devices as well as the cognitive status of certain referent do have some impacts on the determination of the referential expressions in Budai Rukai. More details about these two factors will be given in the findings of the next two research questions.

(2) **Research Question Two:** Does the syntactic coding system in Budai Rukai follow Givón's scale of topic continuity?

Based on our data of Budai Rukai narratives, there are four major types of

doi:10.6342/NTU202302199

syntactic coding devices, including zero anaphora, pronoun, definite noun and indefinite noun, and they perfectly follow the Givón's scale of topic continuity, with zero anaphora having the highest degree of topic continuity, pronoun the second highest, definite noun the second lowest and indefinite noun having the lowest one. As the two most continuous NP coding devices, zero anaphora and pronoun are often used to refer to the more important referents in the discourse, which are often not very far away from each other (perhaps within two sentences), and such referents can easily be continued for sentences later. Moreover, zero anaphora is in zero form, and pronouns in Budai Rukai are either in clitic form or one-word form. In other words, both of them possess relatively less coding quantity just as Givón predicted (1983). As for the two least continuous NP coding devices: definite nouns and indefinite nouns, they are in more complicated coding forms, in which a noun phrase is often combined with a demonstrative or a case marker, referring to relatively unimportant topics, which is often several sentences away from the next mention, and such topics don't tend to persist in the context for long.

Topicalization, as one of the common syntactic coding devices, is not mentioned by Givón in the scale of topic continuity. A topicalized NP, in the preverbal position, is not very continuous in terms of referential distances because it is usually a way to reintroduce the old information back to the context, causing its high value of referential distance. However, topicalization is found to be pretty continuous when it comes to persistence. That is because such topicalized NP is very often continued in the later context for long, leading to its higher value of persistence than untopicalized NPs.

(3) **Research Question Three:** What is the relation between the cognitive status of a referent and the referential form in Budai Rukai, and does it align with Gundel's Givenness Hierarchy of NPs?

doi:10.6342/NTU202302199

According to our findings, the Gundel's Givenness Hierarchy of NPs can also be seen in Budai Rukai. That is, referents in different cognitive statuses are coded differently. To be more specific, a referent in the status of *in focus*, the highest degree of cognitive status, can be coded as zero anaphora or pronoun. On the other hand, the grammatical device of indefinite noun is used to code *referential* and *type identifiable* referents, which are in the two lowest degrees of cognitive status. Interestingly, definite noun is a widely used grammatical device, which can be seen to code referents in almost every cognitive status except for *type identifiable*. Furthermore, the *activated, familiar* and *uniquely identifiable* referents are all primarily expressed in the form of definite noun. However, the demonstratives used on the definite nouns differ depending on the degree of the cognitive status of the referent. For instance, *kuini* is mostly found on the definite noun with slightly higher cognitive status, including *in focus, activated, familiar and uniquely identifiable.* As for the *referential* definite noun*,* which is in the relatively lower cognitive status, it is most coded by *kudra.* Such choice of different demonstratives is assumed to be connected to the semantic features of the demonstratives. With the speakers' preferences to different demonstratives, a clear distinction between *uniquely identifiable* status and *referential* status can be seen. On the other hand, since all the four highest cognitive statuses (from *in focus* to *uniquely identifiable*) take the same coding demonstrative, it seems hard to distinct these four statuses from one another.

Last but not the least, we also discover an interesting relation between the topic continuity and the referential givenness or cognitive status. When the referent is in the higher degree of cognitive status, such as *in focus*, then it tends to be coded by syntactic coding devices that are relatively more continuous, zero anaphora as well as pronoun included. Conversely, if such referent is in the lower degree of cognitive status, such as *referential* or type *identifiable*, a less continuous syntactic coding

doi:10.6342/NTU202302199

device, for instance, indefinite noun, is chosen.

5.2 Limitations and Recommendations for Future Research

The present study has investigated the NPs in a number of Budai natural data so as to figure out the factors that influence the selection of referential expressions in this languages. There are a few limitations that can be further considered in the future research.

To begin with, the genre of the data might have some effects on the choice of the referential expressions. Therefore, looking over as many data as possible can make the results more precise. Nevertheless, coding and analyzing the natural data are such trifling tasks. Only dealing with eleven pieces of data has taken us a lot of efforts and time. Most of our investigated data are storytelling and only two of them refer to culture sharing narratives. In these two genres, for instance, the speakers tend to use distal demonstratives, such as *kudra* and *kuini,* to refer to something on the picture book, on the screen or in their memory. If possible, the future research can try to examine more data and make the genres of the data balanced. With more and various genres included, we expect to see more distributions of other demonstratives, and we also expect that it will be more likely to distinct *in focus, activated, familiar* and *uniquely identifiable* in Budai Rukai.

In addition, most of the syntactic coding devices in Givón's or Gundel's studies are based on Indo-European languages. Some unique grammatical expressions in Austronesian languages, such as nominalization, haven't been tested in this study, left unsolved for the future research.

doi:10.6342/NTU202302199

References

Ariel, M. (1988). Referring and accessibility. *Journal of linguistics*, *24*(1), 65-87.

_____. (2014). *Accessing noun-phrase antecedents (rle linguistics b: Grammar)*. Routledge.

Bentivoglio, P. (1983). Topic continuity and discontinuity in discourse: A study of spoken Latin-American Spanish. *Topic continuity in discourse: A quantitative cross-language study*, *3*, 255-311.

Blust, R. (1999). Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. In *Selected papers from the eighth international conference on Austronesian linguistics* (Vol. 1, pp. 31-94).

Brown, C. (1983). Topic continuity in written English narrative. *Topic continuity in discourse*, 317-341.

Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. *Subject and topic*.

Chen, C. F. (2008). *Aspect and tense in Rukai: Interpretation and interaction*. The University of Texas at Austin.

Chen, C. F., & Sung, L. M. (2005). Interrogatives as polarity items in Kucapungan Rukai. *Concentric: Studies in Linguistics*, *31*(1), 95-117.

Fox, A. (1983). Topic continuity in Biblical Hebrew narrative. *Topic continuity in discourse: A quantitative cross-language study*, *3*, 215-254.

Gasser, M. (1983). Topic continuity in written Amharic narrative. *Topic continuity in discourse: A quantitative cross-language study*, *3*, 95-139.

Givón, T. (1983). Topic continuity in discourse: An introduction. *Topic continuity in discourse: A quantitative cross-language study*, *3*, 3-41.

_____. (1983). Topic continuity and word order pragmatics in Ute. *Topic Continuity in Discourse: A Quantitative Cross-Language Study. Amsterdam/Philadelphia: John Benjamins*, 141-214.

_____. (1983). Topic continuity in spoken English. *Topic continuity in discourse: A quantitative cross-language study*, 343-363.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 274-307.

Gundel, J. K., Hedberg, N., Zacharski, R., Mulkern, A., Custis, T., Swierzbin, B., ... & Watters, S. (2006). Coding protocol for statuses on the givenness

doi:10.6342/NTU202302199

hierarchy. *Unpublished manuscript (1993/2006). http://www. sfu. ca/hedberg/Coding_ for_ Cognitive_ Status. pdf*.

Hedberg, N. (2013, December). Applying the givenness hierarchy framework: Methodological issues. In *International workshop on information structure of Austronesian languages*.

Hinds, J. (1983). Topic continuity in Japanese. *Topic continuity in discourse*, 43-93.

Miyao, M. (2017). *The processing of referential expressions in discourse by Chinese, English, and Japanese native speakers and by Chinese and Japanese learners of English*. PhD Thesis. University of Hawai'i at Manoa.

Nie, X. (2020). *Information Structure, Discourse and Distribution of Referring Expressions in Chinese.* PhD Thesis. The University of Arizona.

Pu, M. M. (1989, November). Topic continuity in written Mandarin discourse. In *Annual Meeting of the Berkeley Linguistics Society* (Vol. 15, pp. 256-267).

Ross, M. (2009). Proto Austronesian verbal morphology: A reappraisal. In *Austronesian historical linguistics and culture history: A festschrift for Robert Blust*. Asia-Pacific Linguistics, College of Asia and the Pacific, The Australian National University.

Shi, Y. (1998). *Referring expressions in Chinese and English discourse*. Ball State University.

Shih, W. C. L. (2012). Are there case markers in Budai Rukai. *Taipei: National Taiwan University (MA thesis)*.

Shih, C. H. (2022). Adverbial Constructions in Budai Rukai. *Taipei: National Taiwan University (MA thesis)*.

Su, L. I., Sung, L.-M., Huang, S., Hsieh, F., & Lin, Z. (2008). NTU corpus of Formosan languages: A state-of-the-art report. Corpus Linguistics and Linguistic Theory, 4(2), 291–294.

Sung, L.-M., Su, L. I., Hsieh, F., & Lin, Z. (2008). Developing an Online Corpus of Formosan Languages. Taiwan Journal of Linguistics, 6, 79–117.

Sung, L. M. (2011). Clausal nominalization in Budai Rukai. *Nominalization in Asian languages: Diachronic and typological perspectives*, 523-559.

Starosta, S. (1988). A grammatical typology of Formosan languages. *Bulletin of the Institute of History and Philology*, *59*(2), 541-576.

Tanangkingsing, M. (2013). A study on the behavior of Cebuano pronouns in

doi:10.6342/NTU202302199

discourse. *Concentric: Studies in Linguistics*, *39*(1), 59-89.

Wang, K. H. (2022). Spatial Conceptualizations in Budai Rukai. *Taipei: National Taiwan University (MA thesis)*.

Young, C. J. (2019). The am Construction in Yami. *Taipei: National Taiwan University (MA thesis)*.

Young, L. X. (2022). Generalized Universal Quantifiers in Budai Rukai. *Taipei: National Taiwan University (MA thesis)*.

Zeitoun, E. (1997). The pronominal system of Mantauran (Rukai). *Oceanic Linguistics*, 312-346.

_____. (2000). *A Reference of Grammar of Budai Rukai* [Series on Formosan Languages 8]. Taipei : Yuanliou Publishing Company. (In Mandarin Chinese).

doi:10.6342/NTU202302199