國立臺灣大學文學院語言學研究所

博士論文

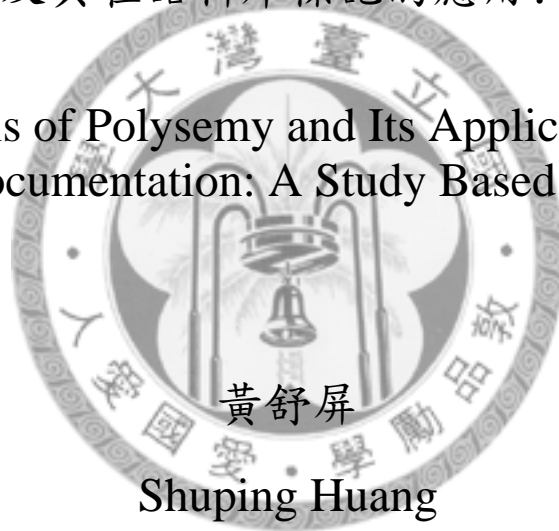Graduate Institute of Linguistics

College of Liberal Arts

National Taiwan University

Doctoral Dissertation

多義詞的分析及其在語料庫標記的應用：以賽夏語為例

Analysis of Polysemy and Its Applications to
Corpus Documentation: A Study Based on Saisiyat

黃舒屏

Shuping Huang

指導教授：蘇以文 博士

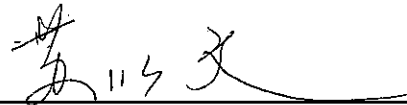Advisor: Lily I-wen Su, Ph.D.

中華民國 97 年 7 月

July 2008

國立臺灣大學博士學位論文
口試委員會審定書

多義詞的分析及其在語料庫標記的應用：以賽夏語為例
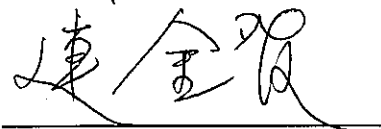## Analysis of Polysemy and Its Applications to Corpus Documentation: A Study Based on Saisiyat

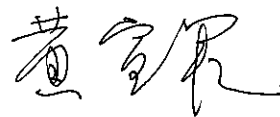　　本論文係黃舒屏君（F90142008）在國立臺灣大學語言學研究所、所完成之博士學位論文，於民國 97 年 7 月 4 日承下列考試委員審查通過及口試及格，特此證明

口試委員：

_____（簽名）
　　（指導教授）

_____　　_____

_____　　_____

Acknowledgements

　　寫作論文的過程是漫長而艱辛的，我也由此深深感覺到自己是個幸運的人，每當我感到困惑的時候，總有來自師長親友的鼓勵和指點，才能順利完成這本論文。

　　第一個要感謝的人是我的指導老師蘇以文教授。在語言所的七個年頭，老師從學術上到生活上，都給我很多寶貴的建議，這個研究最初的構想也是來自蘇老師的語料庫計畫。我無法忘記和老師一起去美國參加會議的時候，老師親自煮了咖啡，和我閒話家常，如同慈母般的關心，讓我永難忘懷。在我研究的過程中難免有挫折與不順遂，老師總是想盡辦法給我們各種協助；在我痛哭的時候，老師溫暖的關懷和耐心的開導，是支持我堅持到最後的主要動力。

　　我也要感謝我的口試委員黃宣範老師、連金發老師、宋麗梅老師、張永利老師，以及給予我很多幫助和鼓勵的畢永峨老師。黃宣範老師給我許多的建議，幫我指出研究的盲點。此外，在語言所的日子裡，從很多小細節上，我也親身體驗到黃老師對學生的關懷和照顧。連金發老師在口試的時候給了我很多寶貴的意見，讓我獲益良多。他不但博學而且親切。清晰的思路和耐心的指正，提供了我許多分析的方向。我要特別感謝宋麗梅老師，她不但帶我認識台灣南島語的研究，也在我人生中的許多關卡給我鼓勵和協助，更讓我擔任她的教學助理，與我分享許多教學的寶貴經驗。宋老師對學生的熱情和誠摯，更讓我深深感受到一種師長的風範。另外，對於張永利老師，除了感謝之外，也讓我深感欽佩。張老師在我的口試中提出許多的指正，讓我的想法更周全，他的好學不倦和對新知的追求，更帶給我很多的思考跟啟發。另外，畢永峨老師的幫助與鼓勵讓我銘感在心。畢老師一直都是我相當崇敬的學者，在提案的時候畢老師給予我相當寶貴的建議，而且她親切的態度，永遠令我感到如沐春風。

　　此外，我要感謝曾經教導過我的語言所師長。已經退休的張裕宏老師，是我心目中的長者典範，他的教導讓我對自己的母語——閩南語，有了更深一層的認識與了解。還有張顯達老師，他喜歡和學生一起用餐聊天，對學生的各種疑難雜症總是可以提出很多解決方案，我也感謝他對求職的問題給了我一些實質的建議。江文瑜老師對生活、人性、和社會的關懷，是我相當欽佩的，受到她的啟發，讓我對語言學有新一層的認識，也開始在研究中努力嘗試多面向的思考。安可思老師讓我學習到許多在語意學知識以及研究方法，對我的研究影響良多。此外，我有興能修到馮怡蓁老師的語音學課。馮老師的講

i

解清晰，由淺入深，有時更讓我對自己所學的語意學和語用學觸類旁通。記得通過提案時，馮老師是第一個給我祝福道賀的人，她的關懷讓我非常感動。

感謝所上的學長姐、學弟妹和同學們。和我同在蘇老師研究團隊的同組成員們，最能了解我的研究從無到有的過程，而其中有很多的成分來自於他們的寶貴建議。這中間當然也少不了互吐苦水的時候，謝謝他們分擔我部分的心理壓力，特別是貼心的季樺跟佑旻。還要感謝我的三位摯友：富惠學姊、紀昕、跟 Michael 學長。富惠學姊是我的心理輔導師兼老媽子，分享我生活中的大小事。紀昕陪我度過很多個在電腦室寒冷寂寞的日子，而 Michael 學長提供我很多的輕鬆笑料。我還要感謝熱情開朗的 Sally，我永遠的同學芳梅，還有陪我聊是非的鴻鈞。另外，謝謝可愛的立心每次都和我一起上山做田野調查。

這個研究用到很多寶貴的賽夏語語料。真的很感謝語料庫的工作人員們和組織語料庫工作的宋麗梅老師，還有國科會計劃和「多媒體整合實驗室」計劃的資助，才有這些資源可資使用。當然還有賽夏發音人朱阿良老師、高德盛老師、趙山河老師、跟風德輝老師。他們都是熱情親切的老人家，認真耐心的提供我們珍貴的語料，也讓我得以體會到賽夏語的奧妙。從事語料庫工作和田野調查的過程中，我非常的愉快而充實。

感謝我從大學以來的至交筱茹跟小建，不管我開心還是難過，她們永遠樂意當我分享情緒的對象。還有男友布萊恩，在論文寫作的最後階段，他可能永遠不曉得我為什麼老是唉聲歎氣，不過開朗樂觀的他，總是不忘給我一個陽光笑容。

最後，我要感謝從小到大一直支持我的家人。不管我好與壞，我的父母和哥哥們永遠都鼓勵我追求自己的夢想。我也希望在天堂的阿公，能跟我分享這喜悅的一刻。

每次想起這些背後的助力，我都會很感激的告訴我自己：「還好我不是自己一個人。」

Abstract

Being an important issue, polysemy has received a great much attention from different fields such as linguistics, philosophy, and psychology. In everyday conversation, language users typically generate appropriate interpretation of a polyseme with no difficulty (Taylor) — Polysemy is found to cause problems mainly in cross-linguistic contexts (Riemer 2001). We are thus motivated to investigate polysemy in a cross-linguistic context. Based on our experience of working on the creation and maintenance of the NTU Corpus of Formosan languages, we aim to explore documentation of polysemes by case studies on Saisiyat polysemy, using Mandarin as a meta-language.

In the course of language fieldwork, there is no known method of determining precisely the meaning of a lexical item. Two factors are directly responsible: contextual modulation and mismatches of semantic partitioning. The meaning of a lexical item is modulated in situated contexts, and new meaning emerge almost whenever a lexical item is used in a novel context. In addition, semantic partitioning is also a major source of misunderstanding. For concepts that are conceived as relevant in Saisiyat, their relations may not be highlighted in Mandarin and are thus expressed with formally irrelevant linguistic forms. When one Saisiyat lexical item yields multiple Mandarin translations, we need careful collation to determine which translation reflects the core meaning of the delimited lexical item.

Our analysis of three Saisiyat lexical items reveals that meta-language translation has a direct impact on our interpretation of the linguistic data. For example, Saisiyat *nanaw* denotes limitation of a quantity, like English 'only,' and it is also used to express an affirmative attitude to the factual status of a statement, like English 'exactly.' These two meanings are related, but such extension is found to be typologically-unimportant ones manifested in few languages, and its meta-language translations are seemingly irrelevant, which can yield a homonymy reading. Another lexical item *nahaen* is used to denote repetition, succession, and precedence of activities. Similar semantic network are more likely to be found in other languages, and many of its Mandarin translations exhibit functional overlaps, which may help corpus users to discover the relations between instances of *nahaen*. We may also come across cases of very high degree of cross-linguistic predictability, yielding similar ways of conceptual categorization in genetically-unrelated languages, as well as relatively consistent direct translation. Our study of Saisiyat *ma'* 'also' is one of the examples.

Theoretically, language-specificity of semantic partitioning urges us to take a social-cultural view on perspectivization. It is commonly held that the development of a

network exhibits tendency of "subjectification"— the speaker tends to include his own epistemic attitude and personal evaluation when using an expression (Langacker 1990). Traugott (2003) further postulates that speaker's point of view has to align with that of his addressee, a tendency known as "intersubjectification." The intersubjectification view is supported by our study, but we claim that the speaker's construal of a scene for the purpose of verbalization has to take into the shared linguistic and non-linguistic knowledge of the entire speech community. A "collective" view of intersubjectification is postulated to account for language-specificity.

Empirically, we propose that a consistent gloss of a polyseme can be sought on the basis of its prototype. We agree that inconsistent and imprecise translation of a lexical item in different contexts is inevitable in linguistic fieldwork. A polyseme is nevertheless advisable to be glossed consistent by one meta-language gloss (Lehmann 1982). Consistent glossing yields five advantages: 1) reflecting fundamental division between semantics and pragmatics, 2) facilitating user's identification of a linguistic item, 3) exhibiting economy and precision of data presentation, 4) facilitating search of a lexical item in a corpus, and 5) preserving conceptual categorization of the target language. Based on the categorization view of polysemy, we propose that the meta-language gloss of a polyseme should reflect the prototype of the polyseme. When a researcher wishes to gloss a polyseme by a consistent cover term, he can exploit the notion of prototype.

Overall, our investigation integrates intra-language investigation of polysemy with inter-language comparison of semantic partitioning. In addition to examination of theoretical issues, we tackle empirical problems of semantic analysis in language fieldwork with special focus on the possible applications to language documentation.

Key words: polysemy, comparative linguistics, categorization, prototype, language documentation, perspectivization

摘要

詞彙的多義現象是語言學、哲學及心理學共同感興趣的議題。雖然一詞多義理論上會造成溝通的誤解，但是這樣的誤解很少出現在日常語言使用中 (Taylor 2003)；多義詞造成困擾，多半是在跨語言溝通的情境 (Riemer 2001)。這促使我們將多義詞的研究延伸到跨語言的比較，而本論文即以語料庫標記為主要探討目標。我們從「台大台灣南島語多媒體語料庫」的工作經驗出發，以三個賽夏語的多義詞彙為案例，分析語言田野調查過程中，透過發音人的解釋和翻譯了解多義詞的過程。

在田野調查中，確認一個詞彙的意義是相當不容易的，這受到兩個因素的影響：詞彙語意的環境調節(contextual modulation) 以及跨語言的語意切分落差 (mismatches of semantic partitioning)。環境調節為某個詞彙增添或改變語意，因此當該詞彙每次出現在一個新的語境中，就有可能得到不同的解釋。語構對詞彙意義的影響亦不容忽視。此外，語意切分的方式在各個語言中往往有落差，一群概念在賽夏語中被視為相互有關聯，並用同一詞彙表達，在中文中他們之間的關係卻可能被忽略，而用多個不同的詞彙來表達。當一個賽夏語詞彙對應到多個中文翻譯時，我們需要透過分析和比較才能知道哪一個翻譯最能表現該詞彙的核心語意。

我們分析了三個詞彙，發現在呈現語料時，超語言翻譯的影響是很直接的。例如賽夏語的 *nanaw* 可以用於表達數量的限定，接近中文的「只」，也可用於表達對某項事件真實度的肯定，接近中文的「原本」。類似的語意延伸在其他的語言中並不多見，也因此發音人所提供的直接翻譯詞看似彼此並無關聯，容易產生同型異義(homonomy)和多義詞的混淆。另一詞彙 *nahaen* 用於表達事件重複、事件接續以及事件先行，類似的語意延伸較容易在其他語言中發現，其中文翻譯「再」、「又」、「還」等等，大多在中文中都有明顯的功能重疊，*nahaen* 的用法之間的關係也較容易想像。而少數的辭彙，例如 *ma'*，意義接近中文的「也」，其語意延伸有很高的跨語言可預測性，類似的語意網絡常出現在不同的語言中，其中文翻譯一致性較高，不容易產生同型異義的誤解。

分析的結果可以從兩方面來討論。理論面上，跨語言多義詞語意發展的特殊性促使我們重新審視「觀點化」(perspectivization) 現象。一般認為語意變化的途徑是受到主觀化的影響 (Langacker 1990)，意即以說話者的觀點出發。近來則興起互動主觀化的概念 (Traugott 2003)，強調說話者的觀點選定需配合其言談對象。我們的研究支持互動主觀化，並認為互動主觀化的考量對象應包含整個語言社群，方能解釋語言獨特性的產生。

應用面上，我們認為單一詞彙得到多種中文翻譯，是田野調查不可避免的過程。然而經過分析和對照之後，多義詞在語料標記中宜採用一致的超語言注釋 (meta-language gloss) (Lehmann 1982)。其優點有五: 1) 代表語意和語用層面的基本切分，2) 有利語料庫使用者了解多義詞和同型異義的區分，3) 標記的經濟性，4) 利於語料庫內的詞彙搜索，5) 保存瀕危語言的獨特語意切分方式。我們也提出，多義詞的本質為一種分類 (categorization)，而一個多義詞的原型 (prototype) 可作為尋找超語言註解的根據。在一個詞彙的眾多可能意義當中，超語言註解反映的為其原型意義。

　　本文結合了單一語言的多義詞研究和跨語言的多義詞比較，除理論方面的思考外，並探討語意學研究在語言田野調查的進行和語料標記的應用，有助於我們在研究意義的過程中採用更寬廣更活用的觀點。

關鍵詞: 多義詞、比較語言學、分類、原型、語料標記、觀點化

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CODING CONVENTIONS

1SG	first person		INT	interjection
2SG	second person		IPL	hearer-inclusive plural
3SG	third person		LF	locative focus
ACC	accusative		LNK	linker
AF	agent focus		LOC	locative
ALL	allative		NEG	negative
ASP	aspect marker		NEU	neutral
BC	back channel		NMZ	nominalizer
*C*a	*C*a reduplication		NOM	nominative
CAU	causative		OBJ	object
CLF	classifier		PF	patient focus
COM	comitative		PFV	perfective
COMP	complementizer		PL	plural
COND	conditional marker		PN	proper name/place name
CONJ	conjunctor		POSS	possessive
COP	copula		PROG	progressive
CRS	Mandarin *le*		PRTCL	particle
DAT	dative		PST	past
DECL	declarative		Q	question Marker
DET	determiner		RED	reduplication
DIM	diminutive		RF	referential focus
DM	discourse marker		SG	singular
EPL	hearer-inclusive plural		SUB	subject
EXIST	existential		TOP	topic
EXPER	experiential		VOC	vocative
FIL	pause filler			
FS	false start			
FUT	future			
GEN	genitive			

* For markings of discourse features, such as pause, latching, overlapping, etc., please refer to Appendix C.

Chapter 1   Introduction

1.1. Preliminaries

The study of meaning, in cognitive linguistics, has been put in the forefront of linguistic inquiry, with special focus on the exploration of polysemy (Taylor 1990). Since a century ago when Michel Bréal coined the name *polysémie*, investigation into related meanings of the same linguistic form has been made popular (Nerlich and Clarke 2003). Ullmann (1957) even stated that polysemy is the "pivot of semantic analysis." It has been a century since Bréal, but linguists' enthusiasm for polysemy never dies. Today, various journals, books, and conferences are devoted exclusively to the issue of polysemy. One cannot help wondering: Why is polysemy worth studying?

One of its mystiques, as we would like to emphasize in this dissertation, is the fact that polysemy is the crossroads where several research fields intersect. As has been pointed out in Nerlich and Clarke (2003: 16), polysemy is "a phenomenon that exposes the multiple relations and connections between syntax, semantics and pragmatics, and between language, cognition and social interaction." Investigation of polysemy thus has considerable potential to reveal the functioning of human cognition, and also to characterize negotiation of meaning in communication.

It has also been singled out in Nerlich and Clarke (2003) a modern interpretation of polysemy has been facilitated by a revolutionary view of categorization in anthropology and psychology. A categorization view of polysemy offers us a new angle to link

language back with mind, meaning, and society. Particularly, notions of prototype, fuzziness, and family resemblance are found to have insightful manifestations in studies of polysemy. The categorization view of polysemy has nurtured Taylor's (1995) book *Linguistic Categorization* in which he drew our attention to many problems that can be attributed to the categorical nature of language. However, a categorization view of language has also invited many problems that are not empirically dealt with. One of the problems that will be addressed in this dissertation is cross-linguistic comparisons of polysemy. For example, a linguistic form X in one language covers three concepts A-B-C in a conceptual space, yet its counterpart X' in another language covers only B-C, and still another counterpart X'' in a third language may be used to denote concepts B-C-D-E. For a word that is highly polysemous, it becomes very difficult to find a form in another language that entirely equates to it, which can be a source of misunderstanding in cross-linguistic communication.

The prevalence of polysemy demands researchers' special attention in the course of language fieldwork. Vaux and Cooper (1999) remind researchers that if an informant produces a word for a given meaning, this might not be the only meaning of the word. In different contexts, different interpretations are possible. Due to the flexibility of lexical meaning, language glossing, as correctly observed by Samarin (1967), is inevitably inconsistent and imprecise at earlier stages of linguistic investigation. Although Lehmann (1982) suggests polysemes to be glossed by a "cover term" cross-textually, systematic and thorough collation can be done only when abundant linguistic data are accumulated.

And we argue, based on the categorization view of polysemy, that the notion of "prototype" can be exploited in search of a consistent gloss. We propose a set of procedure to find out on the basis of "prototype" the cover term that can be used to gloss a polyseme, which may serve as a reference when systemization of language glossing is attempted.

1.2. Objectives

The aims of our investigation can be many-fold. Empirically, we attempt to explicate three cases of polysemy that have puzzled us in the course of language fieldwork. We find that direct translations offered by the informants can be confusing, and two factors are directly responsible for inconsistent and imprecise translations: contextual elaboration and language-specificity of semantic partitioning. If a researcher wishes to gloss a polyseme by a cover term, we propose a set of procedure that helps us to determine the gloss that can best represent the prototypical meaning of a delimited polyseme. With proper modifications, the procedure may also shed lights on studies of semantics for lexicography, translation, pedagogy, and other cross-linguistic applications.

Theoretically, we consider this as a chance to examine current theories of polysemy. Many current theories polysemy are founded on researcher's intuition, and they can easily fall into three kinds of fallacies: a) vagueness in determining uses of a polysemy, b) inconsistence of meaning representation, and c) vagueness of what the introspective network actually reflect (Dominiek and Rice 1995). A similar position is also held by

3

Croft (1998), who criticizes researchers' introspective data for being too speculative. However, Croft's attitude to introspective linguistic analysis is not entirely pessimistic: He suggests that introspective hypothesis can be validated by three kinds of evidence: psychological experiments, corpus analysis, as well as cross-linguistic comparisons. In addition, in his book *Radical Construction Grammar* (Croft 2001), Croft stipulates that linguists can take an inter-language perspective even for examination of intra-language phenomena. We have been encouraged by Croft's suggestion to carry out the present investigation in a cross-linguistic perspective. Studies of polysemy used to focus mainly on intra-language aspects, and analysis of polysemy from a comparative view may be a chance to re-examine current theoretical frameworks in a new perspective.

Moreover, comparative linguistics inevitably invites the year-long debate on language universality versus relativity. For a very long time, linguists have been ambitious to claim the relation between "speech" and "thought." Some argued that by studying the forms in languages of the world, we may uncover the thinking process of speakers in different speech communities. By studying the verbalization of concepts across languages, we aim to find a way to talk about the relation between language and thought, and also to have a better understanding of cross-linguistic communication.

1.3. A sketch of Saisiyat

To facilitate our discussion, this section presents a brief account on our target language, i.e. Saisiyat, in terms of its geographic distribution, population, and the language family it belongs to, followed by a sketch of Saisiyat grammar.

1.3.1. The Saisiyats

Saisiyat is an Austronesian language spoken primarily in Hsinchu (Wufeng Township and Beipu Township) and Miaoli (Nanjhuang Township and Shihtan Township), the mountainous areas of north-western Taiwan. In two different regions, two dialects are spoken: the Donghe dialect in Hsinchu and the Daai dialect in Miaoli. According to a census conducted by Taiwan Executive Yuan in December 2007, Saisiyat has a population of 5,541. This would be approximately 1.1% of the total indigenous population in Taiwan.

There are different views on the genetic classification of Saisiyat. According to Ho and Yang (2000), Saisiyat belongs to the Atyalic subgroup which includes also Atayal and Pazih. But different views have been proposed in Starosta (1995), Li (1997), and Blust (1999). Constantly adapted to the predominant language communities in Taiwan, Saisiyats are often multilinguals, fluent in Mandarin Chinese, Atayal, Hakka, and sometimes Taiwanese Min. Some elder Saisiyats, ruled by the Japanese government during 1895~1945, also have a basic command of Japanese. As a consequence of cultural adaptation, Saisiyat is among the many aborigines facing the possible death of their own

language. The Taiwanese government in recent years has been working in many aspects to prevent indigenous languages in Taiwan from extinction. Scholars have also endeavored to research and preserve valuable language data. The NTU Corpus of Formosan Languages, as we will introduce soon in Chapter 2, is one of the projects to preserve valuable language and cultural information in view of imminent language loss.[1]

1.3.2. A brief sketch of Saisiyat grammar

This section presents a brief sketch of Saisiyat grammar, including the phonemic inventory, word order, case marking system, personal pronominal system, and focus system.[2]

1.3.2.1. The phonemic inventory

According to the orthography system announced by the Council of Indigenous Peoples, Executive Yuan, there are 16 consonants and 6 vowels in Saisiyat. In addition, Saisiyat makes distinction between long vowels and short vowels. The colon symbol ":" is used to indicate vowels with longer duration. The writing system is shown in Table 1.1 (for consonants) and Table 1.2 (for vowels).[3] In these two tables, the symbols in parentheses represent the corresponding IPA symbols.

---

[1] The Academia Sinica Formosan Language Archive also collects data of Saisiyat in addition to other Formosan languages (cf. Zeitoun and Yu 2005).
[2] Donghe dialect and Daai dialect do not have significant syntactic different. They differ slightly in phonological system and some lexicons (Li 1978b; Yeh 2000).
[3] The spelling system is available on the website of the Council of Indigenous Peoples, Executive Yuan, Taiwan
(http://www.apc.gov.tw/chinese/docDetail/detail_TCA.jsp?docid=PA000000000154&linkRoot=0&linkParent=0&url=)

Table 1.1. Consonants in Saisiyat

Manner ＼ Place	Labial	Alveolar	Palatal	Velar	Glottal
Stop	p	t		k	' (ʔ)
Nasal	m	n		ng (ŋ)	
Fricative (voiceless)		s (θ, s)[4]	S (ʃ)		h
(voiced)	b (β)	z (ð, z)			
Lateral		l (l, ļ)			
Trill			r		
Glide	w		y (j)		

Table 1.2. Vowels in Saisiyat

	Front	Central	Back
High	i		
Mid	oe	e (ə)	o (o/u)[5]
Low	ae	a	
Vowel lengthening		:	

The writing system reflects the phonological system of the Saisiyat language. In Table 1.1, consonants which are of the same place or manner of articulation may share the same symbol if they are not phonologically contrastive. The writing system also takes into consideration easy customization for computer processing. It adopts symbols in the ASCII format; the schwa /ə/ is therefore written as *e*.

---

[4] The sound /s/ is pronounced as [θ] in Donghe dialect but as [s] in Daai dialect. And [z] is pronounced as [ð] in Donghe dialect, but as [z] in Daai dialect. The palatal fricative [ʃ] is not phonologically contrastive to /s/ and /z/, and shares the same symbol with /s/ and /z/ in the orthography system.

[5] The mid-back vowel /o/ is often pronounced as high vowels /u/. They are commonly considered free variations, but there are some debates on their phonological contrastiveness. This issue is, however, beyond the scope of the present dissertation.

1.3.2.2. Word order

Most Austronesian languages spoken in Taiwan are predicate-initial languages, but Saisiyat is unique in that it primarily follows S-V-O order (Yeh 2000a). The predicate (abbreviated as V following conventional tradition) can be a verb, as in (1a), a stative verb,[6] as in (1b), or a nominal predicate, as in (1c).

(1) a. *yao*   *'am*   *s\<om\>i'ael*   *ka*   *'aelaw*
      1SG.NOM   FUT   eat\<AF\>   ACC   fish
      'I want to eat fish.'
  b. *yao*   *kin*   *bali'*
      1SG.NOM   very   thin
      'I am very thin.'
  c. *kizaw*   *'in-sia-a*   *minkoringan*
      PN   POSS-3SG-POSS   wife
      'Kizaw is his wife.'

The word order of a clause is related to its "focus." As we will explain later, Austronesian languages is characterized by a set of verbal marking system that signals the thematic role of the clausal subject (the role marked nominative in a clause).[7] When the event agent is marked as nominative case, the word order follows S-V-O more strictly. On the other hand, when the event patient or benefactive/instrument is marked as nominative case, the word order is relatively free: S-V-O, O-S-V, O-V-S, S-O-V, V-S-O, and V-O-S are all acceptable. The following examples are taken from Yeh (2000a: 71-72) to illustrate the free variation of word order in a patient-focus clause.

---

[6] What are known as "adjectives" in English should be deemed stative verbs in Formosan languages (Zeitoun 2000; Zeitoun and Huang 2000).
[7] The notion of "subject" is controversial, especially for non-Indo-European languages. Since the problem is not the focus of the present study, we will simply adopt this term to refer to the nominative argument in a clause.

(2)a. *pazay   ma'an   si'ael-en*
  rice 1SG.GEN eat-PF
  'I ate the rice.'
 b. *pazay   si'ael-en   ma'an*
 c. *ma'an   pazay   si'ael-en*
 d. *ma'an   si'ael-en   pazay*
 e. *si'ael-en   ma'an   pazay*
 f. *si'ael-en   pazay   ma'an*

Occasionally, predicate-initial utterances can be found in spontaneous speech. We find some instances of predicate-initial utterances, such as (3) and (4), taken from NTU Corpus of Formosan Languages.

(3) Life
144 F: (H)(Hx)(H)(Hx)(H)(Hx) <u>*ititi'an*</u> <u>*ka*</u> <u>*rayhil*</u> *la==*
             V   S
             a.bit NOM money INT
  'Not much money lah.'

(4) Frog5
18 …(2.2) <u>*raiw*</u> <u>*ila*</u> <u>*ka==*</u> <u>*a*</u> <u>*takem*</u>
      V   S
     AF.leave PFV NOM FIL frog
  'The frog escaped.'

Yeh (2000a) mentions that documents in pre-KMT times contain more predicate-initial examples. She thus hypothesizes that constant contact with Mandarin Chinese might have changed Saisiyat word order to S-V-O. Li (1978) argues for V-O-S order in addition to S-V-O. In other words, there can be two canonical word orders in Saisiyat. The S-V-O word order has been statistically attested in Huang, Su and Sung

(2004) in a great majority of transitive agent-focus clauses, but explication of this issue demands further investigations.

1.3.2.3. Case marking system

Nouns in Saisiyat are led by different markers to indicate their case roles in a clause. The case marking system marks six roles: nominative, accusative, genitive, possessive, dative, and locative. The system is further divided into two groups for marking of 1) proper nouns and 2) common nouns. The case markers of Saisiyat are shown in Table 1.3.

Table 1.3. Case markers in Saisiyat (cf. Yeh 2003: 14)

case role / type	nominative	accusative	genitive	dative	possessive[8]	locative
proper nouns	ø/hi	hi	ni	'ini	'an--a	kan/kala
common nouns	ø/ka	ka	noka/no	no	'an noka--a	ray

The case marking system is closely related to the semantic role of the noun and the focus of the sentence. In an agent-focus clause, the agent/actor is marked nominative, and the patient is marked accusative, as shown in (5a). In a patient-focus clause, the patient is marked nominative, and the agent (optionally present) is marked as genitive, as shown in (5b). In a referential-focus clause, the benefactive/instrument role is marked nominative, as shown in (5c).

---

[8] In Donghe dialect, some speakers pronounce the the first vowel of the possessive case marker as /i/. That is, *'an--a* is pronounced as *'in--a*, and *'an noka--a* as *'in noka--a*.

10

(5)a. *ka*   *'ahoe'*   *k<om>aat*   *hi*   *yaba'*
      NOM   dog   bite<AF>   ACC   father
      'The dog bit Father.'
   b. *hi*   *yaba'*   *ka:aS-en*   *noka*   *'ahoe'*
      NOM   father   bite-PF   GEN   dog
      'The dog bit Father.'
   c. *ka*   *kahoey*   *si-Sebet*   *ni*   *yaba'*   *ka*   *'ahoe'*
      NOM   wood   RF-hit   GEN   father   ACC   dog
      'Father used the wood stick to hit the dog.'

The genitive marker, in addition to agent role in patient-focus clause, can also lead

instrument and possessor, as shown in (6).

(6)a. *S<om>bet*   *ka*   *'ahoe'*   ***noka***   ***kahoey***
      hit<AF>   ACC   dog   GEN   wood
      '(Someone) hit the dog with a wood stick.'
   b. *sia*   *ray*   *taw'an*   ***ni***   ***obay***
      3SG.NOM   LOC   house   GEN   PN
      'He is at Obay's house.'

Other case markers are not as multifunctional: The possessive marker encodes the

possessor of a property. The dative marker marks the benefactive role or a reference for

comparison in comparative clauses. And the locative marker introduces a location or a

source.

Table 1.3 follows the proposal of Yeh (2003) that nominative case marker is often

linguistically not realized (ø). However, Hsieh and Huang (2006) find that the "dropping"

of nominative case marker is closely related to the clausal structure and particularly the

discourse pragmatics of language use. Preverbal nominative marker (in SV clause) is

79% dropped, because the SV structure in Saisiyat often elaborates on an existing NP.

11

Postverbal nominative marker (in VS) is only 7% dropped, because VS structure in Saisiyat often introduces a new NP. In other words, the choice between realized *hi/ka* and unrealized ø can be explained by information status of an NP.

1.3.2.4. Personal Pronominal system

The six-way case system is also manifested in the personal pronominal system. According to the grammatical role it plays in a clause, a pronoun changes its form. The personal pronominal system of Saisiyat is summarized in Table 1.4.

Table 1.4. Personal pronominal system in Saisiyat (cf. Yeh 2003: 16)

	Case / Place		nominative	accusative	genitive	dative	possessive[9]	locative
SG	1st		yao/yako	yakin	ma'an	'iniman	'amana'a	kanman
	2nd		So'o	'iSo'on	niSo	'iniSo'	'anSo'o'a	kanSo'
	3rd		sia	hisia	nisia	'inisia	'ansiaa	kansia
PL	1st	H-inclusive	'ita	'inimita	mita'	'inimita'	'anmita'a	kan'ita'
		H-exclusive	yami	'iniya'om	niya'om	'iniya'om	'anya'oma	kanyami
	2nd		moyo	'inimon	nimon	'inimon	'anmoyoa	kanmoyo
	3rd		lasia	hilasia	nasia	'inilasia	'anlasiaa	kanlasia

Like many other Formosan languages and Austronesian languages, Saisiyat distinguishes "hearer-inclusive" and "hearer-exclusive" first person plural pronouns. Also, gender is not distinguished in Saisiyat pronominal system. In addition, Saisiyat does not have pronouns for non-human entities.

---

[9] In Donghe dialect, the first vowel of the possessive pronouns is sometimes pronounced as /i/. For example, *'amana'a* is pronounced as *'imana'a*.

Unlike most of the other Formosan languages, Saisiyat does not have clitics for personal pronouns. All the pronouns in Table 1.4 are free forms.

1.3.2.5. Focus marking system

Focus system is an important linguistic issue specific to Austronesian languages (Ho and Yang 2000). The term "focus" is sometimes used interchangeably with "voice", but recent linguists have attempted to argue that the so-called "focus" or "voice" in Austronesian languages are not identical to the notions that have been used in Indo-European linguistic studies (Himmelmann 2002, Ross and Teng 2005).[10] Further progress of Austronesian language typology may require the explication of this linguistic device. In this dissertation, the term "focus" refers to a set of verbal morphology that signals the semantic role of the grammatical subject in a clause.

Generally, Formosan languages have four foci: agentive, patientive, referential (instrumental/benefactive), and locative. However, in modern Saisiyat, locative focus (-*an*) only occurs in lexical nominalization, and is not used as a clausal focus marker (Yeh 2000a). The loss of locative focus renders three-way distinction in modern Saisiyat, which is summarized in Table 1.5.

---

[10] Terms like "case", "topicalisation", "theme", "recentralisation", and "trigger", all refer to the same morpho-syntactic device commonly known as "voice" and "focus" (Blust 2002). Ross and Teng (2005) regard focus markers as transitive/intransitive markers; agent focus marks intransitive events, whereas non-agent focus marks transitive events.

Table 1.5. The focus marking system of Saisiyat (cf. Yeh 2000a:86)

Focus Set	I	II
Agentive	m-, <om>, ma-, ø[11]	ø
Patientive	-en[12]	-i
Referential	si-/sik-	-ani

As shown in Table 1.5, there are two sets of focus marking in Saisiyat. Set I is used for declarative clauses and negative clauses containing negators *kayni'* and *'okik.* And Set II is used in imperative clauses as well as negative clauses containing negators *'okay, 'izi'* and *'in'ini'*.

The focus markers are affixed to a verb to indicate the semantic role of the clausal subject (the role that is marked as nominative case). Based on Yeh (2003), the roles that are "focused" by using different focus markers are summarized in Table 1.6.

Table 1.6. The relation between focus markers and clausal subject (Yeh 2003)

Focus marking	The semantic role of the subject
Agentive	Agent, Experiencer, Theme
Patientive	Patient
Referential	Instrument, Benefactive, Reason, Transported Patient, Patient-like Roles

Yeh (2003) points out that focus system and case marking system collaborate to indicate the semantic roles and grammatical relations of the event participants. Focus marker indicates the role of the clausal subject, and the roles of other non-subject

---

[11] Formosan languages usually have multiple agent focus markers. Please see Huang (2000) to learn the relation between focus marking and the degree of dynamicity.
[12] The patient focus marker exhibits vowel harmony when influenced by its preceding vowel(s). For example, the patient focus form of *ra'oe* 'drink' is *ra'oe-oen*. Other examples are like *ki:im-in* 'look.for-PF' and *tonrong-on* 'dispatch-PF.'

arguments are indicated by their case markers.

Note that the referential focus marker *si-*, as shown in Table 1.6, introduces a wide variety of thematic roles. The one-to-many mapping between the marker and its multiple functions has attracted some discussions, and an inspiring account on its semantic extension has been postulated in Yeh (2003). However, the roles listed in Table 1.6 appear to be non-exhaustive. According to the Hsieh (2007), Saisiyat referential focus frequently marks the role that perceives a stimulus and is psychologically affected. Example (7) is extracted from our fieldnotes.

(7) Ø   *'ahoe'*   *si-Sebet*   *ka*   *korkoring*   *ra:iw*   *ila*
     NOM   dog   RF-hit   ACC   child   AF.leave   PFV
     'The dog saw the child being beaten, and it ran away (with fear).'
     (Literally: The dog perceived the child being bitten, and it ran away.)

The subject of this clause, i.e. the dog, is the role that visually perceives a stimulus. In other Formosan languages, referential focus typically does not have functions such as in (7), but this function seems to have been conventionally associated with the referential focus in modern Saisiyat.

1.4. Organization

The organization of this dissertation is as follows. Chapter 2 reviews previous studies on polysemy and contrastive linguistics. As our investigation targets language documentation of polyseme, we also introduce relevant conventions of language

documentation. Chapter 3 presents the methodology of our investigation. For analysis of polysemy, we proposed a set of procedure to find out the relation between the meanings of a lexical item.

Chapter 4, 5, and 6 offer analyses on three lexical items. We find that contextual modulation and mismatch of semantic partitioning yields multiple interpretations of a lexical item. Also, "cross-linguistic predictability of semantic hierarchy" has direct impact on the use of direct translation in the course of semantic investigation. For a semantic network that has a low degree of cross-linguistic predictability, similar semantic extensions are rare in other languages, and the relations between the meanings are less conceivable, more likely to be idiosyncratic semantic development manifested in few languages. Direct translations of a polyseme may look irrelevant to each other, and its instances are likely to be taken as homonymous. For some polysemes with intermediate degree of cross-linguistic predictability, direct translation may generate a number of meta-language equivalents that are functionally related to each other in the meta-language. Due to their functional overlaps, we can more easily discover the relations between the translations. If a polyseme shows a high degree of cross-linguistic semantic predictability, similar semantic extension is likely to be found in other genetically-unrelated languages. However, such cases are rare on account of language-specificity of semantic partitioning.

A cross-linguistic study of polysemy invites us to rethink the notion of "perspectivization." In Chapter 7, we show how the development of a semantic network

reflects the speaker's desire to express his subjective epistemic stance, i.e. subjectification, and also to appeal to his hearer, i.e. intersubjectification. Cross-linguistic comparison further suggests that the notion of intersubjectification should be expanded to the speaker's adjustment of his viewpoint to a collective perspective — to align with the conventional way of perspectivization in the speech community.

Chapter 8 is devoted to the possible application of our study to language documentation. Direct translation often results in inconsistent meta-language glossing. If a polyseme, as postulated in Lehmann (1982), should ideally be glossed with a cover term cross-textually, the cover term of a polyseme, we suggest, should be sought on the basis of its prototype. In Chapter 8, we, according to the categorization view of polysemy, exploit the notion of "prototype" to look for the Mandarin cover terms of three Saisiyat polysemous lexical items.

Finally, a comparative study of polysemy gives rise to many points worth further investigation. Conclusions and suggestions for further study are given in Chapter 9.

Chapter 2   Literature Review

2.1. Preliminaries

In his work *The Meaning of a Word*, Austin (1961) raised a crucial question: Why do we use the same term to name things of different kinds? In fact, almost all lexical items are polysemous (Deane 1988). Given that one linguistic form typically has more than one meaning, we may intuitively assume that polysemy should be a major source of miscommunication. Contrary to this intuition, in everyday communication, language users typically generate appropriate interpretations of a polyseme with no difficulty, and are generally not even aware of its multiple potential readings (Taylor 2003). Polysemy is found to cause problems mainly in cross-linguistic contexts. In Section 2.2, we will talk about the most noticeable problems triggered by polysemy. Being an important theoretical and empirical issue, polysemy has received a great much attention from different fields in the past century. In Section 2.3, we review some best-known approaches to polysemy, including lexical, cognitive, functional, and cross-linguistic approaches. And these approaches share the same view — polysemy is an effect of conceptual categorization. The categorization view of polysemy will be addressed in Section 2.4.

Motivated by empirical applications to language documentation, we also introduce the convention of language glossing in Section 2.5. Notions such as "equivalent" and "translation" are important for glossing of polysemy, yet they are usually taken for

granted in corpus documentation. Insights from contrastive linguistics are deemed valuable in this respect. Specifically, the levels of contrastive analysis and the relation between translation and prototype are highly relevant to the design of interlinear language documentation. Relevant issues are introduced in Section 2.6.

2.2. Polysemy in cross-linguistic contexts

In intra-language communication, polysemy rarely results in misunderstanding (Taylor 2003). The association of one form to multiple meanings becomes confusing primarily when we need to use one language to explain the meaning of the other. As pointed out in Riemer (2001), we often do not realize a linguistic form in our native tongue as polysemous until we have to translate it into a second language.

Forms deemed "equivalents" in two languages often differ in terms of the meanings they can denote. *Spiel* in German is usually translated as "game" in English. However, German *Spiel* is not entirely identical to English *game* — A boy playing sand alone in the park can be denoted by German *Spiel*, but not by English *game* (Wierzbicka 1996). Development of any bilingual dictionary inevitably faces this problem. Translation is another area that requires careful treatment of polysemy. To translate a linguistic form accurately and appropriately, one has to appreciate how its semantic extensions differ from those of its crosslinguistic equivalent(s). Moreover, for pedagogical applications, recognition of polysemy is also crucial. Understanding the semantic network and the cognitive-pragmatic links between the meanings can help language learners appreciate

the various usages of a linguistic form in a more coherent and insightful manner in stead of rote memorization (Tyler and Evans 2004). On the other hand, a student may use a foreign word inaccurately if he transfers the entire semantic network from the meta-language. For example, a Chinese student, by producing an utterance *Move your mouse to **point** this icon*, fails to recognize that Chinese *diăn* means 'to point (at something)' most of the time, but it also means 'to click (the mouse)' as a semantic extension. The English equivalent *point* does not yield this peripheral extension.[1]

2.3. Approaches to polysemy

Studies of polysemy are not only crucial for empirical purposes, they are also promising to reveal the richness and flexibility of human cognition, and have thus drawn the attention of many linguists. This section introduces several well-known approaches to polysemy. We will evaluate their advantages and shortcomings in due course.

2.3.1. Lexical approach

Lexical approach to polysemy is generally concerned with the meaning of words and the meaning relations among words. Lexical semanticists explore semantic properties and semantic relations in an attempt to discover the structure of mental lexicon as a part of our linguistic knowledge. Some representative works are Weinreich (1962), Ullmann (1972), Palmer (1976), Lyons (1977), Cruse (1986), among others.

---

[1] This example is drawn from the speech of Lily I-wen Su "Metaphor as a way of categorization: Evidence from metalanguage glosses" delivered in Academia Sinica, May 16, 2005.

One of the foci of lexical semantics is the study of semantic boundaries. Polysemy causes few troubles in daily communications (Taylor 2003), but "semantic boundaries" are psychologically real. Language users often play on the confusion between two senses of a linguistic form to deliberately create anomalies for the purpose of amusement, such as punning. A discrete semantic unit psychologically contrastive to other meanings is called a "sense," and ways to look for senses of a word are of particular value to lexicographic works.

Nevertheless, the nature of full sense units is difficult to state. Lexical semanticists rely on "boundary effects" to detect the discreteness between meanings of a linguistic form. Many tests for ambiguity have been proposed: identity test, independent truth conditions, and independent sense relations. Nevertheless, few of them have been uncontroversially successful (Cruse 1986, 2000). Cruse (2000) suggests that "discreteness" may not satisfactorily define sense boundaries. Instead, "antagonism," also known as "attentional autonomy," is the decisive aspect. In view of this, a sense is defined in terms of its autonomy and its mutual exclusiveness with other meanings.

And as polysemy is commonly defined as the "association of two or more related senses with a single linguistic form" (Taylor 1995: 99), another issue regards how "relatedness" is to be rated. The question may have an absolute answer for diachronic semantic research, but in synchronic studies, "relatedness" is a matter of degree. Palmer (1976) has proposed four criteria to distinguish homonymy from polysemy: etymology, relations, existence of a central core, and independent lexical relations. The first criterion

pertains to diachronic evidence. For languages with abundant historical documents, relatedness between senses is a yes/no question, depending on whether their common origin can be found (Croft and Cruse 2004). However, when we encounter languages with few historical data, the distinction can only be judged in terms of the latter three criteria which are nevertheless based mainly on speakers' intuition. Speakers constantly differ in their intuitions; and even worse, in cases where all four criteria can apply, historical fact and speaker intuitions may contradict each other (Saeed 1997).

There is a growing awareness of the fuzzy boundary between vagueness and polysemy (cf. Tuggy 1993; Cruse 2000; Croft and Cruse 2004), and between polysemy and homonymy (cf. Taylor 2003). The difficulty to find a consistent way of sense judgment is well reflected in lexicographic works — Different dictionaries are diverged in terms of their ways to list entries of a lexical item and to cluster them (Taylor 2003).

2.3.2. Cognitive approach

Cognitive approach to polysemy attempts to understand the use of language on the grounds of human cognitive ability. Cognitive Grammar (Langacker 1987a, 1991) is typical of this approach. Langacker (1987b) proposes that a linguistic form has a canonical language use, generally known as the "prototype." The prototype is potential to develop, and the development includes not merely "outward" extension, but also "upward" schematization. This can be sketched in Figure 2.1.

Figure 2.1. Extension and schematization (Langacker 1987b: 140)
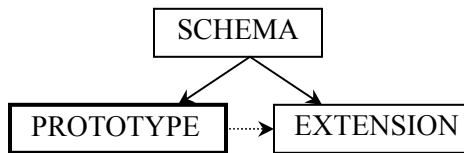
Extension and schematization yield collaboratively a complex network through constant use of a linguistic item, and the development may look like the sketch shown in Figure 2.2. In this figure, the sequence of the alphabets represents chronological order.

Figure 2.2. Development of a linguistic network (Langacker 1987b: 140)
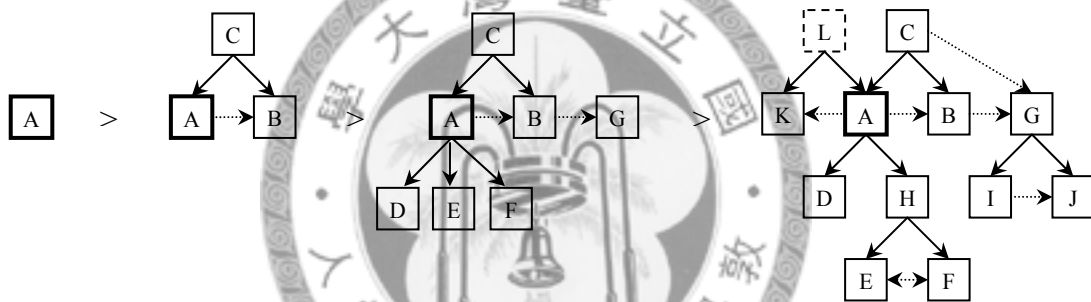
The development involves: (a) extension: with a schema (such as A➔B) or without a schema (such as B➔G), (b) downward articulation: finer distinctions being made (such as A➔D, E, F), (c) sub-schematization: sub-schemas extracted and interpolated for nodes already present (such as H), (d) incorporation of additional categorizing relationships (such as C➔G), and (e) constant adjustment: determined by usage and experience. The development of a network generally shows a tendency of "subjectification"— the speaker tends to include his own epistemic attitude and personal evaluation when using an expression (Langacker 1990).

Polysemy can be viewed as relationships whereby the instantiations are strongly entrenched and the schema does not exist or is only weakly felt, such as the relation between A and K, under a non-entrenched schema L. Vagueness, or ambiguity, is the case of further instantiations subsumed under an entrenched schema, such as A and B (under C). Tuggy (1993) proposes a framework to treat the continuum between ambiguity and vagueness, which is very much like Langacker's framework, but one additional parameter is added: the elaborative distance between the schema and the instantiations.[2]

Then what is essential of being subsumed under the same schema? This question pertains to the notion of "frames" or "domains." It has been suggested that **concepts are not stored separately**. When encountering new experience, we look for its connections with those that we have already known, and we solve a problem according to its typical situation we expect (Tannen and Wallet 1993; Croft and Cruse 2004). The "abstract structure of expectations with roles, purposes, sequences of events" is called a "frame" (Fillmore 1982: 117) or a "domain" (Langacker 1987a).[3] Polysemy is an effect of different "perspectivization" of the elements in the same frame. For example, the English word *safe* can be used in *safe child*, *safe razor*, and *safe helmet*. They respectively mean

---

[2] Ruhl's (1989, 2002) and Biwerwisch's (1983, cited in Taylor 1995b) model of monosemy reflects the same view of language networking. However, they take a very different methodology by treating all instances in the same network as non-discrete entries.

[3] According to Evans and Green (2006), Langacker refers to "domain" as distinguished from "frame" in the following aspects: 1) the proposal of "domain matrix" by which the typical structure of our experience is dependent on multiple domains and their interactions, 2) the distinction between "basic domains" such as TIME and SPACE that are based on our embodied experience and "abstract domains" such as MARRIAGE that may involve social activities, 3) The hierarchical organization of domains as chains of profile-base relation by which ARC triggers the "base domain" CIRCLE which is in turn the profile of another base SPACE, and 4) the emphasis on the conceptual relation between knowledge rather than on grammatical manifestations.

"to be free from harm," "not to cause harm" and "to be able to resist harm." The term *safe* activates a frame of "harming" and the speaker can use this term to highlight the harmer, the harmed, and the harm defender (Fauconnier and Turner 2002). There are many other researchers who employ different terminologies with different emphases: frame (Fillmore 1982, 1992), scenario (Sanford and Garrod 1981), domain (Langacker 1987a; Lakoff and Johnson 1980), cognitive model (Ungerer and Schmid 1996), idealized cognitive model (Lakoff 1987), protoscene (Tyler and Evan 2001), script (Schank and Abelson 1977), schema (Langacker 1987a; Johnson 1990), just to name some of them. But essentially, they all highlight the relation between embodied experience and conceptualization, and also how the relation is reflected in linguistic structures.

Frames or domains elucidate the cognitive process we often come across in language use. The use of a lexical item activates a concept in a frame of related notions. When it is used to highlight another concept in the same frame, it creates metonymic effects. When a lexical item is used in an atypical frame, new meanings will emerge, which is a metaphorical effect. Verbalization may also involve concepts of more than one frame; the integration of frames is called blending (Fauconnier and Turner 2002).

The radial-center approach springs from the notion of frames and domains (cf. Lakoff 1987; Brugman 1988). Semantic extensions are constrained by our knowledge structure: A linguistic unit is used to refer to concepts that are of metaphoric, metonymic, propositional, or image-schematic relation to an "idealized" member of a frame.

Following the spirits of Wittgenstein (1963), the semantic network of a linguistic form is considered manifesting "family resemblance", i.e. sharing similarities not by all members, but rather in a crisscross manner.[4] Principled Polysemy (Tyler and Evans 2001, 2003, 2004; Evans 2003) is also one of the offshoots that gracefully picture the development of a linguistic family from the prototype (center) to peripheral members (radial). In addition, Principled Polysemy has suggested a set of concrete criteria to (a) determine the discrete senses of a lexical item, (b) construct the links between those senses, and (c) look for the dynamic adjustments of the protoscene to meet various cognitive and pragmatic considerations in language use.

The cognitive approach to polysemy links linguistic phenomena back with their underlying cognitive mechanisms. The division between vagueness and ambiguity is not as important. Instead, the central aim of the cognitive approach is to explain WHY multiple meanings are associated to one linguistics form.

2.3.3. Functional approach

The functional approach is interested in how speaker-hearer interaction, particularly the negotiation of meaning, contributes to the extension of a semantic network. Polysemy, as suggested by Gyori (2002), is the product of conversational demands. On the one hand, the speaker has to choose the form that best communicates the "information." On the other hand, he has to choose a way that best captures his "intention." As has been pointed

---

[4] It is called "chaining" in Austin (1961).

26

out in Sweetser (1990: 75), "an utterance is content, epistemic object, and speech act all at once." Polysemy is a way to be economical and informative at the same time, being a problem-solving strategy under the pressure of communicative interaction to meet the demands from semantic pole and pragmatic pole.

Conventionalization is another focus of functional approaches to polysemy. When using a linguistic form, the speaker usually avoids innovations unless the current linguistic form fails to fully convey the speaker's information/intention. The priority rule of language use is to follow the convention "ritualized" in the conversation ceremony (Goffman 1981). Even when innovations are necessary, the speaker has to "invite" his audience to make appropriate interpretation (Traugott and Dasher 2002; Levinson 2000). In this case, relevance is the key. In a normal cooperative conversation, the speaker is expected to produce utterances that are relevant to the context (Sperber and Wilson 1995). Meanings that are implicated, or less canonical to a linguistic form, can be well-understood if relevant to the linguistic or non-linguistic contexts. A similar view is also advocated by the "emergent approach" which believes that semantics is the study of semanticization, i.e. of negotiated pragmatic meanings which become conventionalized in collective discourse (Huang 1998). In other words, what we agree to be the meaning of a linguistic form is in fact the reflection of sense-frequency in cross-textual consistency. To make himself understood by the hearer(s), the speaker has to simulate the knowledge background and the viewpoint of his hearer(s), a process known as "intersubjectification" (Traugott 2003; Traugott and Dasher 2002). When the polysemy network of a lexical item

is established, it would become the solution to the future communicative problems, and serve as the foundation of further semantic extensions (Campbell 1998).

The functional approach to polysemy, like the cognitive approach, is also interested in WHY we use the same term to name things of different kinds. But the functional approach expands its scope beyond solely speaker-internal cognition, and aims to restate the essence of polysemy as a tool of communication in wider social-interactive contexts.

2.3.4. Cross-linguistic approach

Cross-linguistic approach aims to explore the universal way of human conceptualization by comparing semantic extensions of crosslinguistic equivalents. The units of comparison are not limited to "words," but include grammatical constructions and expressions of concepts. Abundant works have been devoted to crosslinguistic comparisons of linguistic expressions, such as emotions (Wierzbicka 1999), space (Bowerman 1996; Talmy 1983), middle voice (Kemmer 1993), indefinite pronouns (Haspelmath 1997), modality (van der Auwera and Plungian 1998), classifiers (Aikhenvald 2000), grammatical relation (Croft 1991), among others.

Natural Semantic Meta-language (NSM) breaks down a given concept into "semantic primitives" or "primes" which are believed to be fundamental building blocks of meaning in all human languages (Goddard and Wierzbicka 2002, 2004; Wierzbicka 1972, 1996, 2002). By decomposing a linguistic form into more basic elements, a researcher can work out cultural scripts to look for similarities and differences of similar

concepts in different languages. However, NSM has difficulty deciding which primes should be included. Wierzbicka (1972) started the list with 14 primes, but the list has continued to grow, and is not yet exhaustive. Also, from a philosophical perspective, it is doubtful whether a decompositional script can faithfully reflect speaker's conceptualization. As suggested in Jackendoff (1987), some concepts or words, e.g. such as *purple* or *atonal*, can hardly be broken down into smaller elements. Lehrer (1992) also directly warns that there may be no definition that can fully capture the meaning of a word, even the elaborate definitions proposed by Wierzbicka.

Many cross-linguistic studies adopt directly or indirectly a "Semantic Map" (SM) approach, for example, Kemmer (1993), Haspelmath (1997), van der Auwera and Plungian (1998), Bowerman and Choi (2003, indirectly), and so on. SM metaphorically conceptualizes a group of related concepts as a "space," and related functions of a linguistic form are imagined to be "regions" interconnected in this space. Each language has a specific way of partitioning this space, constituting a "map" of language-specific grammars. For example, in Figure 2.3, the two squares enclose two function-sets that are represented by two linguistic forms.[5]

---

[5] One advantage of Semantic Map approach is that is can be used for intra- as well as inter-language comparison: the two enclosures can be linguistic items in a language or counterparts in two different languages. With proper methodological adjustment, a semantic map is also potential to be used for diachronic changes of a linguistic form across time.

Figure 2.3. Partitioning of a conceptual space by linguistic forms
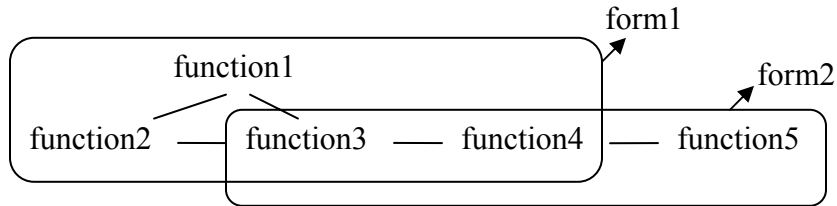
Each function is given a semantic label according to crosslinguistic comparisons: When any of the languages under investigation make distinction of any two functions by different ways of verbalization, they have to be labeled as two separate functions on the map. The connections between the functions are also verified according to crosslinguistic comparisons: When a group of concepts is denoted by the same linguistic form in a dozen of languages, its members are said to be conceptually related.

Croft's (2001) Radical Construction Grammar (RCG) also exploits the notion of "semantic map" to show the functional distribution of particular contrastive linguistic forms. The Semantic Map Connectivity Hypothesis is proposed: "Any relevant language-specific and construction-specific category should map onto a CONNECTED REGION in conceptual space" Croft's (2001: 96). Those functions sharing the same form can be monosemous (uses of one single sense) or polysemous (related senses of a linguistic form), subject to linguists' interpretation.[6] RCG also aims to search for common tendencies of functional extension and its underlying cognitive mechanisms. By connecting the semantic map of more languages in the world, we are likely to see the

---

[6] "Homonymy" is theoretically excluded from semantic map approach. However, the subtle differences between polysemy and homonymy and the difficulties in finding an "idealized" homonymy are issues beyond the scope of the present study.

semantic subdivisions of a conceptual space, in an attempt to construct a universal semantic map which human thinking is based upon. In general, the ultimate goal of SM-based approaches is to rebuild the blueprint of universal human conceptualization (Haspelmath 2003).

2.4. A categorization view of polysemy

The above-mentioned four approaches to polysemy are not mutually-exclusive. Lexical approaches such as Lyons (1977) and Cruse (2000) pay much attention to the functional basis of polysemy, and cross-linguistic approaches such as Radical Construction Grammar (Croft 2001) draw a lot of insights from cognitive semantics and functionalism. As we will show in the next chapter, they are equally insightful in our investigation. In fact, with the increase of studies on polysemy, a categorization view of polysemy gradually takes shape. **From both the processing and the production aspects, it is uneconomical to give everything a "name," and the solution is to group relevant concepts into a "category" by the use of the same linguistic label. This process will be referred to as "conceptual categorization" in the present dissertation**. Taylor's (2002:10) statement can gracefully conclude the categorization view of polysemy:

> *... words can be regarded as names for categories. To know the word tree means, among other things, being able to apply the word to anything that can be categorized as a tree. Secondly, language itself is an object of categorization. Acoustically different sound signals get categorized as instances of the same linguistic expression, diverse*

linguistic expressions get categorized as examples of the same lexical or

syntactic category, such as 'noun', 'verb', 'transitive clause', and so on.

The lexical approach to polysemy pays more attention to boundary effects of categories; the cognitive and functional approaches focus on the motivations of category extension; and the cross-linguistic approach aims to find out similarities and differences of conceptual categorization in different languages. Many linguists have also made insightful explorations on linguistic categorization, although they do not propose a systematic framework, e.g. Geeraerts (1989), MacLaury (1991), Taylor (1995), etc. Fruitful findings are collected in the book *Meaning and Prototypes* edited by Tsohatzidis (1990).

The categorization view of polysemy is closely related to philosophical and psychological observations on human categorization, including Austin (1961), Wittgenstein (1963), Rosch (1978, 1983), Coleman and Kay (1981), Labov (1973), etc. According to categorization theories, concepts are grouped and stored according to their similarity, and in a group there is a central core that represents the most typical referent, known as the "prototype." In studies of polysemy, "prototypicality" can have two readings. The first one regards prototypicality as the most salient member in a category, and the second one views prototypicality as an abstracted generalization of its category members (Cruse 1990). Although Cruse (1990) speculates that the abstraction view might have more explanatory power for linguistic phenomena, for studies of polysemy, Sweetser (1986) postulates that the two possibilities may apply to different types of

polysemy studies: Some categories, such as "negation" and "coordination," are more likely to have an abstract schema, whereas prepositions and modalities might have many unpredictable and idiosyncratic extensions, and one of the senses may be saliently perceived as the prototype. The distinction is nevertheless blurred in Langacker (1987a: 371) and Taylor (1990), who hold that the two are not conflictive. A category DOG, for example, could have an abstract gestalt image generalized from the dogs we have experienced. It could also activate a salient type of dogs, say, a golden retriever, due to frequency, goodness, or other factors.

The semantic extension from the core to the peripheries can also be accounted for by the categorization view. It has been suggested that the use of language is a kind of meta-representation (Vervaeke and Green 1997; Levinson 1997; Wilson 2000). It does not reflect reality of the physical world, but rather represents our subjective reconfiguring of a perceived stimulus, also known as "construals" (Langacker 1987a).[7] There was a time when linguists generally gave up semantic studies, because the paths of extension appeared to be irregular and idiosyncratic, and thus, following Chomskian approaches, fail to reflect the universality of human thinking. Over the years, this view has been overthrown by the discovery of the principled ways of semantic extensions (Langacker 1987a; Lakoff 1987; Levin 1993; Pustejovsky 1998; Tyler and Evans 2001). Four types of construal changes are concluded in Croft and Wood (2000) and Croft and Cruse (2004): Attention (salience), Judgment (comparison), Situatedness (perspective), and Constitution

---

[7] They are similar to Talmy's notion of "imaging system" (Talmy 1977, 1978).

(gestalt). They will be introduced in detail in Chapter 3. When we change our typical way of conceptualization, a different aspects of the same scene will come into view, which explains how additional meanings are derived in everyday use of language. And through pragmatic strengthening and constant use, the meaning implied in specific contexts may become conventionalized.

Basically, a categorization view of language use is underlying many different theoretical frameworks. Linguists, nevertheless, feel the need to adopt different terms to highlight their different emphases. Table 2.1 summarizes the terms used by different linguists.
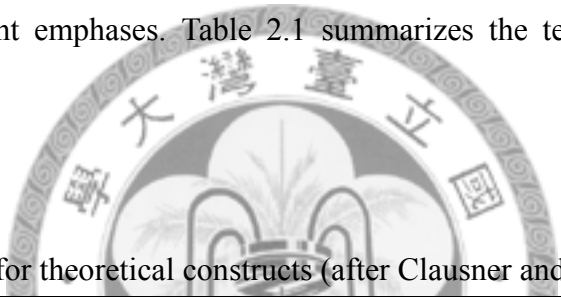
Table 2.1. Terms used for theoretical constructs (after Clausner and Croft 1999: 4)

Used by Notions	Fillmore	Langacker	Lakoff	Clausner & Croft	Tyler & Evans
focus	concept	profile	concept	concept	profile
knowledge base	frame	base, domain	ICM, domain	domain	protoscene
extensions	perspective	focal adjustment, construal, conceptualization	metaphor, metonymy, image-schemata, proposition	construal	construal

This table is after Clausner and Croft (1999), and we add Tyler and Evans' (2001) terminology as a comparison. Though different terms are employed, linguists of the cognitive camp consistently regard language use as a "category" by acknowledging an idealized or schematized knowledge base, upon which language users can manipulate ways to focalize an element when motivated by cognitive or interpersonal mechanisms.

2.5. Polysemy and language documentation

Polysemy is the product of conceptual categorization, and because different motivations constantly compete with each other, categorization of concepts is bound to vary in different languages (Croft 2001). For example, Bowerman and Choi (2003) illustrate that English phrasal verbs *put on* and *put in* constitute a group of similar spatial configurations of placing one thing on top of the other, but the scenes denoted by *put on* and *put in* are expressed in Korean by five linguistic forms *nohta*, *ssuta*, *pwuchita*, *kkita*, and *nehta*. Their study gracefully shows how verbalization reflects categorization of concepts, and the boundary of each conceptual category may differ from one language to another.

As the present study attempts to explore polysemy manifested in language documentation, this section addresses relevant assumptions and conventions of language documentation as well as the way of documenting polysemes.

2.5.1. Development of language documentation

Speech is transient. In conversations, speakers produce sound strings that vanish in a few seconds. If a language gradually loses its speakers, as what is happening to 50% of the world's more than 6000 languages,[8] or if it changes with time, we cannot understand what the language is like in terms of its grammatical structure, semantic system, and

---

[8] The number is based on a report "Language vitality and endangerment" announced by UNESCO Intangible Cultural Heritage Section's Ad Hoc Expert Group on Endangered Languages (http://www.unesco.org/culture/ich/index.php?pg=00142), retrieved on 14 April, 2008.

ways of interaction. The death of a language is accompanied by permanent loss of its cultural heritage and a unique way of looking at the world. Speakers of an endangered language may also experience loss of ethnic identity. Language documentation is one way to preserve the precious information for those who seek to prosper or revitalize a language.

To document a language means to "to provide a comprehensive record of the linguistic practices characteristic of a given speech community" (Himmelmann 1998:166). It typically involves recording (audio and/or audio-visual), transcribing, analyzing, glossing, and archiving. As language documentation is imperative, especially to minority and endangered languages, many linguists and organizations have been working on the methods of documenting recorded linguistic data. IPA (International Phonetic Association) has suggested a system of phonetic notation that is commonly employed for transcription of sound strings into written forms. When there is no available orthographic system in a language, the IPA system can be of great use for notation of languages all over the world. The SIL International (originally Summer Institute of Linguistics) has been encouraging and instructing gathering of language data in authentic fieldwork. It aims to expand our current knowledge on the structure of human languages and to promote greater appreciation of cultural differences. The SIL International has developed many tools of data management for language fieldwork, including Shoebox and Speech Analyzer.[9]

---

[9] Please visit SIL's website (http://www.sil.org/computing/catalog/index.asp) to learn more details.

The format of language documentation has been greatly diverse since the nineteenth century. Lehmann's (1982) "interlinear morphemic translation" has had a great impact on the unification of notations for management of language data. One of the well-known schemes that inherit the tenet from Lehmann is the Leipzig Glossing Rules.[10] The Leipzig Glossing Rules adopt Lehmann's three-line interlinear layout that contains transcription, morph-to-morph gloss, and free translation. The Rules may serve as a basic standard set of documentation, and modifications can be advanced for different demands of the analysis. For example, some researchers may need glossing at textual, phrasal, and word levels in addition to the morphemic level.

2.5.2. Glossing of polysemy

Lehmann (1982) advances the interlinear morphemic translation (IMT), and according to his guideline, the format of language documentation basically contains three lines: the phonetic/orthographic transcription, the morphemic gloss, and the translation. A piece of example is presented in (8), a case of Latin as the target language (L1) and English as the meta-language (L2).

(8) Latin (Lehmann 1982: 204)
Time-o   ne   veni-a-t.   **Transcription**
fear-1.SG   NEG.VOL   come-SBJV.PRS-3.SG   **Morphemic gloss**
'I am afraid he might come.'   **Translation**

---

[10] The Leipzig Glossing Rules are developed jointly by the Department of Linguistics of the University of Leipzig and the Department of Linguistics of the Max Planck Institute for Evolutionary Anthropology. It is available on http://www.eva.mpg.de/lingua/resources/glossing-rules.php.

The second line, morphemic glossing, should contain translation of semantic elements and grammatical elements. Semantic elements are glossed with regular fonts, and grammatical elements are coded in small capitalized fonts such as NEG for negative marker.

It is important that the boundaries of units should be made clear. Words are separated by blank spaces, and the gloss should align with the transcription word-by-word. Morphemes are separated by hyphens, and each separated morpheme should have a corresponding gloss. One unit that corresponds to many elements is glossed with the elements separated by a period. The clear boundaries can facilitate comparison of L1 units with L2.

Glossing, as singled out in the Leipzig Glossing Rules, is the result of analysis, not part of the data. In this dissertation, the focus will be placed on the glossing of polysemes. Specifically, Lehmann (1982: 221) advances that "**a polysemous L1 morpheme is constantly rendered by its nearest context-independent L2 equivalent**." The rationale is illustrated by the following quote from Lehmann (1982: 203):

> *It is advisable to regard the lexical morphemes of L1 as context-independent and translate them in the IMT by a cover term…First, it contributes to making the grammatical structure of the L1 text explicit because it lets the reader see what meaning elements there are and which aspects of the total meaning of the text are to be attributed to the grammatical construction. Second, it allows one to give an identical translation for repeated recurrences of the same lexeme in different contexts, which facilitates the reader's task of identifying the elements.*

In view of Lehmann, coding a polyseme by a cover term is helpful for the purpose of clarity and consistency. However, determining the semantic and grammatical functions of a linguistic form often involves comparison of a form in all its occurrences, and it is particularly difficult if the researcher is not a native speaker of the language. Language documentation is a long-term labor. Inconsistency and inaccuracy is an inevitable outcome at earlier stages of language investigation (Samarin 1967), and we need to accumulate abundant examples before we can actually approach any case of lexical polyseme.

2.6. "Prototype" and "equivalent" in contrastive linguistics

Lehmann (1982) suggests glossing a polyseme with a "cover term" or "nearest context-independent equivalent." It is nevertheless a pity that he does not specify how the "cover term" can be ascertained. In this section, we employ insights from contrastive linguistics to address this problem. We first introduce the notion of *tertium comparationist* which helps us to rethink the design of interlinear language documentation. We then explain how the notion of "prototype" can be exploited when a researcher considers looking for a cover term as the gloss of a polysemy.

2.6.1. *Tertium comparationis*

In comparative linguistics, it is very important to determine the "level" at which counterparts are compared. The term *tertium comparationis* is used by linguists to refer to

the level of comparison. *Tertium comparationis* comes from Latin 'the third place of comparison.' It is the ground where two inputs are compared in order to look for commonalities (Pan and Tham 2007). A linguistic form in one language may share a lot of commonalities with many candidate-equivalents in another language. As a researcher aims at different *tertium comparationis* (hence abbreviated as TC), different ways of contrasting and different results will come into view. For example, when responding to complements on one's appearance, (10) may be a proper Polish translation of (9) at the pragmatic level. Although they are apparently of different syntactic structure and semantic content, they grasp the convention of responding to complement in different speech communities: thanking in English, and self-denigrating in Polish (Jaszczolt 2003).

(9) Thank you.
(10) To tylko stara sukienka. 'It's only an old address.'

The level of comparison is a key concept that was brought into discussion firstly by Tomasz Krzeszowski. Seven TCs are advanced in Krzeszowski (1984), as listed below. A figure to illustrate the typology of TCs can be found in Appendix A.

(11) Typology of *tertium comparationis*
   a) **statistical equivalence**: quantitative comparison of translationally non-equivalent texts
   b) **translational equivalence**: qualitative comparison of translationally equivalent texts
   c) **system equivalence**: systematic corresponding of paradigms in two languages
   d) **semanto-syntactic equivalence**: systematic corresponding of constructions in two languages

e) **rule equivalence**: systematic corresponding of deep structure rules in two languages

f) **pragmatic equivalence**: systematic corresponding of pragmatic functions in two languages

g) **substantial equivalence:** systematic corresponding of phonological or lexical elements

Among the TCs, Krzeszowski (1990a: 31) suggests that "semantics at sentential level" should be considered the "universal, language-neutral, semantic representation." The **semanto-syntactic equivalence** is deemed the foundation of contrastive linguistics. Krzeszowski's argumentation is largely driven by generative theories, demonstrating semanto-syntactic level as where language competence resides, and equivalence at this level should facilitate contrasts of deep structure in different languages. But perhaps Lakoff's interpretation of "deep structure" can more adequately justify the significance of semanto-syntactic equivalence (Lakoff 1968, cited from Krzeszowski 1990a). According to Lakoff (1968), deep structure is where a) basic grammatical relations between fundamental grammatical categories are defined, b) selectional restrictions and co-occurrence relations are stated, c) appropriate grammatical categories receive lexical representations, and d) inputs to transformation rules are provided. The semanto-syntactic level is the level where these features reside.

Studies of contrastive linguistics are not necessarily interested in transformational rules and generative grammar, but the semanto-syntactic level always receives more attention than other TCs. Many studies in contrastive linguistics choose to start from "meaning," e.g. Chesterman (1998) and Wierzbicka (1996, 2002). It is noteworthy that semantic relations are often accompanied with formal resemblances. Chesterman (1998)

points out that when we look at meaning as manifested in translation, we tend to neglect data that appear "too freely" translated — differ too much from the original formal structure. In other words, semanto-syntactic equivalence indeed plays a significant role in the users' mind.

We find that among many TCs, Lehmann's interlinear morphemic translation model (IMT) focuses on two of them: the semanto-syntactic equivalent and the pragmatic equivalent. Morphemic glossing is to look for context-independent semantic or grammatical essence of a linguistic unit, which corresponds to the aim of semanto-syntactic equivalent. The third line in IMT is "free translation" which looks for the meaning that makes sense in the context in terms of the speaker's illocutionary act, perlocutionary act, and appropriateness. This line corresponds to the pragmatic equivalent in Contrastive Linguistics. The correspondences of IMT and TC are shown in Table 2.2.

Table 2.2. Language glossing and *tertium comparationis*

Terminology in IMT	Levels in TCs
Morphemic gloss	Semanto-syntactic equivalent
Translation	Pragmatic equivalent

Semanto-syntactic and pragmatic equivalents are two TCs that are most frequently investigated in comparative linguistics. It is thus not surprising to find that language documentation reflects these two TCs. They contain the most essential information a non-native researcher wishes to know when encountering a word or utterance in an unfamiliar language.

2.6.2. Prototype and equivalent

By IMT, Lehmann (1982) suggests using a "nearest context-independent" equivalent" to gloss a polysemous lexical item, but there is one crucial question that has been taken for granted: How do we know a form is an equivalent of another form in a different language? In Contrastive Linguistics, it is suggested that equivalence is be judged on degrees of correspondence of pragmatic, semantic and syntactic properties (Krzeszowski 1990b). According to the degree of correspondence, one form in a language might be full matching, partial matching, or no matching with a target linguistic form in another language. Full matching is an idealized situation; practically one should look for the equivalent that is of the higher partial matching. For example, Krzeszowski (1990b: 32-33) employs this example of pragmatic equivalent:

(12) All visitors are kindly requested to leave the boat immediately.

(13) Proszę siadać. 'Please sit down.'
(14) Spieprzajcie stąd. 'Get the hell out of here.'
(15) Statek zaraz odpływa. 'The boat departs in a moment.'

As Polish equivalents to (12), sentence (13) is of the same speech act 'request,' (14) carries the same perlocutionary act 'visitors leaving the boat,' and (15) is of the same modality (degree of politeness). Among them, (15) is generally considered more acceptable than (13) and (14), which suggests that some parameters are more important than others when we are looking for equivalents, i.e. modality > perlocutionary/speech

act. Krzeszowski (1990b) suggests that properties of an utterance constitute a **gestalt.** In this case, (12) has a gestalt of POLITENESS REQUEST TO LEAVE THE BOAT. Bilingual speakers more or less know that some linguistic expressions are better equivalents of a target form than others according to their degree of matching to this gestalt.

Nevertheless, (15) is structurally very different from (12). Other things being equal, we are still prone to look for an equivalent that is structurally similar. Sentence (16) is the best equivalent among (13) ~ (16).

(16) Uprasza się gości o natychmiastowe opuszczenie statku.
   'Visitors are requested to leave the boat immediately.'

Semanto-syntactic equivalent may also be ascertained by matching of properties. Krzeszowski (1990b) concludes that the best Polish equivalent of English *over* is *nad*, because a) the most prototypical four meanings of *over* is translated into Polish *nad* by Polish students advanced in English, and b) the four most prototypical meaning of *over* matches the most prototypical three meanings of *nad*. In his study, we can see that the best equivalent is judged on two factors: **a) how well the prototype of the target form is depicted**, and **b) how the prototypes of the two linguistic forms match each other**. The two factors both emphasize the significance of identifying the "center" or the prototype of a linguistic form.

2.7. Summary

Polysemy is an important issue of semantic investigation and has attracted studies of different approaches. Insights from philosophy and psychology have also fostered the categorization view of language use. In this chapter, we proposed a set of procedure to construct the semantic network of a polyseme and to identify its prototype. Although Lehmann (1982) did not specify how the "nearest context-independent counterpart" can be ascertained for the purpose documenting a polyseme, we suggest that the notion of "prototype" can be exploited in this regard, and the so-called "nearest context-independent counterpart" could be the one that best fulfils prototype matching with the delimited polyseme.

Chapter 3   Data and Methodology

3.1. Preliminaries

In the progress of our language elicitation, the firsthand linguistic data are collected, transcribed, and glossed. Glossing of linguistic data is a result of analysis, and as non-native speakers of Saisiyat, we rely greatly upon our native informants to determine the meaning of a linguistic form. The quickest way is by translating a delimited form into a meta-language that is familiar to both the informants and the researcher. In our case, the meta-language is Mandarin Chinese. However, direct translation, being an effective way of field elicitation, is in some aspects limited. Because a polyseme is often given different interpretations in different contexts, we sometimes fail to discover the relations between its different instances. In addition, mismatches of semantic partitioning in different languages are prevalent, and the use of meta-language gloss sometimes fails to reflect the way of semantic partitioning of the target language. This chapter introduces the procedure we take to construct the lost link between instances of a linguistic form, and the suggested criteria to find out the core meaning of a polyseme.

In this chapter, we specify the types of data that are employed in Section 3.2. The best way to illustrate the problems and our points is by presenting empirical cases. We will concentrate on three single lexical items in the present dissertation. In Section 3.3, we introduce the selection of the three cases. The specific procedures taken in our semantic comparative investigation will be introduced in Section 3.4. We propose a

standard set of procedure to construct the semantic network of a delimited polyseme, including classification of its "senses" (discrete semantic units), identification of its prototype, and examination of the relations between the senses. Following Lehmann (1982), a polyseme is advisable to be glossed consistently by a cover term. Lehman nevertheless did not specify the way to find out a proper cover term to gloss a polyseme. By the notions of "prototype equivalent" suggested in Krzeszowski (1990b), we suggest discovering the cover term of a polyseme on the basis of its prototype.

3.2. The database

The target language of our investigation is Saisiyat. Two types of linguistic data are employed: a) linguistic data in NTU Corpus of Formosan Languages, and b) direct elicitation.

NTU Corpus of Formosan Languages (hence abbreviated as NTU-Formosan) compiles data of Austronesian language spoken in Taiwan.[1] The corpus currently contains data on Saysiyat, Kavalan, Amis, Tsou, and Sakizaya. The current team has been continuously taking efforts to improve and expand the online corpus.

---

[1] In early 2000, Prof. Lily I-wen Su got the funding for the Multimedia Laboratory by joining the multi-disciplinary project initiated by the Center for Information and Electronics Technologies established for the Graduate Institute of Linguistics, National Taiwan University. The disciplines who participated in this project include several institutes of National Taiwan University: Graduate Institute of Linguistics, Department of Information Management, Department of Library and Information Science, Department of Computer Science and Information Engineering, Department of Electrical Engineering, Department of Journalism, and Department of Drama and Theater. NTU-Formosan was created based on the Formosan corpus data collected by the many NSC projects (such as NSC 90- 2411-H-002-050, NSC 91-2411-H-002-088, and NSC92-2411-H-002-078) undertaken by Prof. Shuanfan Huang, Prof. Lily I-wen Su and Prof. Li-May Sung, and is a joint effort again by all three of them. Since March 2006, NTU-Formosan (http://corpus.linguistics.ntu.edu.tw/) is supervised by Prof. Li-May made possible by grants from the Center for Humanities Research, National Science Council. An introduction of the Corpus will appear in *Corpus Linguistics and Linguistic Theory* (Su et al. forthcoming).

NTU-Formosan encodes semantic and grammatical elements following in principle the Leipzig Glossing Rules. In addition to semantic and grammatical glosses, we also include transcription of discourse elements (e.g. intonation unit, pause, overlapping, etc.) according to the convention suggested in DuBois et al. (1993). Please refer to Appendix B for grammatical coding conventions of NTU-Formosan, and Appendix C, for an inventory of its discourse coding.

Among the languages available in the corpus, Saisiyat data amount to roughly two hours, being the largest corpus in NTU-Formosan, making it a better target for our investigation. The types of data in the Saisiyat corpus include face-to-face conversation, spontaneous narration of folklores, pear stories (five pieces), and frog stories (eight pieces).[2]

When citing examples from NTU-Formosan, we as current and past members who have worked on the creation and maintenance of the corpus try to follow faithfully the authentic transcription and glossing as they are presented. We understand that the information presented is far from perfect and we understand that the Corpus is constantly revised and updated.

Many examples are also taken from our own fieldnotes that record not only the trial and error of meta-language translation, but also the interactions with Saisiyat informants

---

[2] Pear stories are developed by Wallace Chafe. The speaker is asked to watch a videotaped movie in which a boy stole the fruits picked by an old man. The speaker is asked to narrate the story after the movie ends. The frog story is developed by Dan Isaac Slobin. The speaker is asked to watch a cartoon book in which a boy and his dog tried to look for a frog in a forest. The speaker must look at the cartoon book and meanwhile make an on-line narration. The frog story is found to be helpful for comparative studies on how motion events are expressed in different languages.

that are at later stages of language investigation found to be crucial to the interpretation of our linguistic data. Table 3.1 lists the major informants who generously cooperate and supply us with linguistic data. Most data concerning grammatical judgment are from the two eldest informants.[3]

Table 3.1. Major informants of direct elicitation

Saisiyat name	Mandarin name	Gender	Year of birth	Dialect	Language repertoire (in order of proficiency)
Kalaeh a 'oemaw	A-liang Zhu	M	1928 (80yrs)	Donghe	Saisiyat, Mandarin, Hakka, Japanese
Parain a 'oemaw	De-sheng Gao	M	1928 (80yrs)	Donghe	Saisiyat, Japanese, Mandarin, Hakka
Bownay a Tahes	De-hui Feng	M	1933 (75yrs)	Donghe	Saisiyat, Japanese, Hakka, Mandarin
'Oemaw a 'obay	Shan-he Zhao	M	1939 (69yrs)	Daai	Mandarin, Hakka, Saisiyat

If the examples are taken from my own fieldnotes, the transcription follows the orthography system adopted by Executive Yuan in November, 2005 (shown in Table 1.1 and Table 1.2 in Chapter 1).

3.3. Selection of the case studies

Three case studies investigated in the present dissertation are *nanaw*, *nahaen,* and *ma'*. They have a meaning close to English 'only,' 'again,' and 'also' respectively, but their semantic extensions are so rich that many of their meanings in the contexts have an opaque relation to each other. For *nanaw*, the informants offers Mandarin translations such as *réngrán* 'still,' *zhǐyǒu* 'only,' *hái* 'still,' and *yúanběn* 'originally.' The lexical item *nahaen* has thirteen Chinese translation, accompanied inconsistently by six English

---

[3] The collection of data is supported by NSC grants (NSC 95-2411-H-002-042-MY3).

translations: *háiyŏu* 'another,' *réngrán* 'still,' *réng* 'still,' *lìngwài* 'another,' *lìng yigè* 'another,' *zài* 'again' and 'later,' *zàidù* 'again,' *xian* 'first,' *háiyào* 'still,' *hái* 'still,' *shŏuxian* 'first,' *yòu* 'again' and 'later,' and *yi-hŭi-r* 'a while.' The lexical item *ma'* has many abstract meanings associated with speech interactions, and it receives multiple Mandarin glosses by means of direct translation: *yĕ* 'also', *ránhòu* 'then,' *jìu* 'then,' *nàyàng* 'that,' *nà* 'that,' DM (discourse marker), FIL (filler). Many instances of *ma'* are glossed with "??," meaning uncertain functions.

The three lexical items are chosen for three reasons. First, when the informants are asked to interpret the meaning of these lexical items, multiple Mandarin translations are provided. Investigation of these lexical items is thus potential to reveal the principles that govern changes of meaning in contexts. Investigation of the three lexical items may thus help us to find a consistent way of glossing for language documentation in the future.

We are also driven by Wang's (1947) observation that lexical items involving logical reasoning in Mandarin, such as *cái* 'just; only,' *jìu* 'then,' *yòu* 'again,' *yĕ* 'also,' *dou* 'all,' are prone to be polysemous in different contexts. Related lexical items in Saisiyat also appear to show complex semantic extensions. We assume that cross-linguistic counterparts do not necessarily develop the same semantic network due to competing motivations in different languages (Croft 2001), and comparison of highly polysemous counterparts, instead of less polysemous ones, will be more interesting. The search for an equivalent of those lexical items in the meta-language, i.e. Mandarin Chinese, would help us to understand the relation between semantic mismatches and human conceptualization.

In addition to the above-mentioned two reasons, our selection of case studies is also driven by the fact that those lexical items retain some semantic elements, yet have abstract meanings that are closely intertwined with linguistic structures and the immediate contexts they occurs in. The three lexical items we have chosen have an intermediate position between semantics and syntax, and each case study exhibits different degree of abstractness. *Nanaw* is more semantically-oriented, and *ma'* is more syntactically-oriented, with *nahaen* in between them. Also, their meanings are strongly related to pragmatic considerations. Understanding their manifestation in different contexts will help to explicate the relation between semantics, pragmatics, and syntax.

3.4. Procedures

Our analysis of each case study will follow a standardized procedure. First, the "senses" of the linguistic form have to be identified. Second, the relations between the senses are constructed if they are "members of a family." This task involves two stages: a) identifying the prototype, and b) associating the peripheral extensions with the central prototype. After that, we will search for the equivalent that best matches this prototype for the purpose of language documentation. Specific strategies of each task are explicated in this section.

3.4.1. Identifying the senses

In vein of lexical semantics, we assume that "sense" is the basic units of conceptual

discreteness. Following Cruse (2000) and Croft and Cruse (2004), we hold that "antagonism" is the defining criterion of "sense boundaries." "Antagonism" is also called "attentional autonomy," referring to the exclusiveness of two units as foci of attention. Some tests proposed by previous studies are directly related to this, as listed below. When one reading passes any of the tests, it will be considered antagonist to other meanings, and gets listed as a discrete sense.

(17) Tests of antagonism (attentional autonomy)

(a) identity test (Cruse 2000)

E.g. Mary is wearing a *light* coat, and so was Jane.

There are two situations. First, Mary and Jane both wear a coat light in color. Second, they both wear a coat light in weight. No crisscross reading is possible. 'Light in color' and 'light in weight' are two antagonist meanings.

(b) truth-condition test (Cruse 2000)

E.g. Are you wearing a *light* coat?

Two answers "Yes and No" can exist simultaneously. For example, if the hearer is wearing a light-colored coat, he can say "Yes, I am wearing a light coat (in color)" or "No, I am not wearing a light coat (in weight)."

(c) exclusiveness test (Cruse 2000)

E.g. We finally reach the *bank*.

The speaker may either reach the bank of a river or a financial institution, but he cannot be at the two places at the same time.

(d) punning (Tuggy 1993)

E.g. John and his driving license *expired* last Thursday.

The co-existence of two readings 'to die' and 'to pass a fixed date' in the same utterance creates a comical effect.

The above mentioned criteria can only be applied in situations where two meanings share the same grammatical structure. **Any two senses that can be distinguished by the autonomy tests listed in (17) will be called "senses" (or "full senses").**

According to Croft and Cruse (2004: 113), attentional autonomy is the most marked level of autonomy. There are, however, other symptoms of boundary effects that "appear at lower levels of autonomy and are therefore not diagnostic for full sense boundaries." They may manifest relational autonomy and compositional autonomy, and the criteria are listed in (18) and (19).

(18) Tests of relational autonomy
E.g. Mary is wearing a *light* coat, but Jane is wearing a *dark* one.
Mary is wearing a *light* coat, but Jane is wearing a *heavy* one.
'Light in color' and 'light in weight' have distinct antonyms, meronyms, hypernyms, hyperonyms, etc.
(19) Tests of compositional autonomy
There are cases of units delimited by full sense boundaries. For instance, in "a steep *bank*," the 'financial institution' meaning is ruled out; in "a high-street *bank*," the 'edge of river' meaning is cut off as if it did not play a part.

Principled Polysemy (Evans 2003, 2005) also suggests two criteria of sense judgment that are in essence tests of compositional autonomy. They are listed in (20). [4]

---

[4] Tyler and Evans (2001) suggest two criteria in their early work of Principled Polysemy:
  i) additional sense: a sense has additional semantic elements
  ii) contextual independence: a sense should be contextually independent
But in Evans (2003, 2005), the second criterion is replaced by two contextually-oriented criteria: concept elaboration criterion and grammatical criterion. The former pertains to selectional or collocational patterning of a specific linguistic form. The latter refers to the nature of the grammatical profile adopted by the linguistic form (i.e. in what grammatical constructions it can appear). The substitution of the criteria shows Principled Polysemy holds an ambivalent attitude to contextual effects.

(20) Sense judgment criteria in Principled Polysemy (Evans 2003, 2005)

    (a) Concept elaboration criterion

        E.g. a <u>steep</u> bank

            an <u>international</u> bank
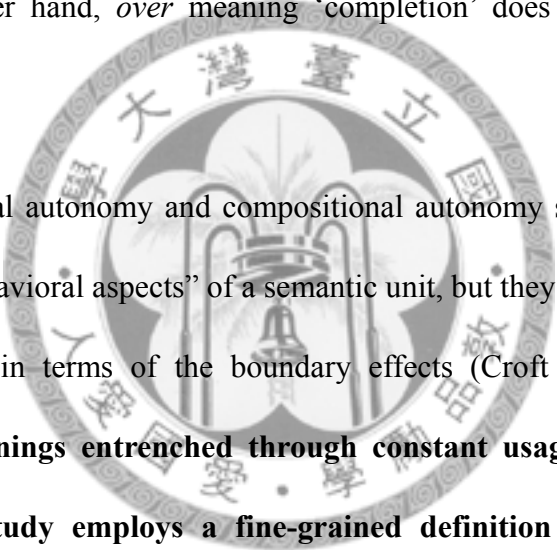
        'River bank' and 'money bank' select different collocated words.

    (b) Grammatical criterion

        E.g. Prepositional structure: The picture is *over* the sofa.

        E.g. Predicative structure: The story is *over*.

        Over meaning 'on top of' is followed by a noun that indicates location. On the other hand, *over* meaning 'completion' does not have a noun that follows.

Tests for relational autonomy and compositional autonomy such as (18), (19), and (20) highlight the "behavioral aspects" of a semantic unit, but they are not as indicative as attentional autonomy in terms of the boundary effects (Croft and Cruse 2004). **To account for the meanings entrenched through constant usage in contexts and in constructions, this study employs a fine-grained definition of "sense." That is, meanings that are distinguished from each other by any of the criteria listed in (18), (19), and (20) will also be listed as "senses."** They may include "facets" or "subsenses" (micro-senses) according to Croft and Cruse (2004). By including relational and compositional autonomy in polysemy studies, we assume that studies of meaning cannot be entirely independent from contextual clues, following the functional approaches to polysemy. Although some linguists attempt to single out senses that are "independent from the contexts" (such as Tyler and Evans 2001), for many linguistic items that are

functional or with a higher degree of abstractness, dependence on contexts is inevitable (semantically as well as syntactically).

Sometimes, for illustration or explanation of subtle functions of a linguistic form, we need to talk about meanings that are at very low levels of usage, i.e. pragmatic elaboration. **Those subtle meanings will be termed as "forces," and will be attributed to relevant senses.** [5] They may include perspectives, active zones, contextual modulations, etc. Please see Appendix D for their definitions in Croft and Cruse (2004). It should be made clear that although meanings may be relevant to contexts, it is not possible to enlist every meaning whenever a lexical item is used in a new context. Comparison of semantic and syntactic structures occurring cross-textually helps us to determine the systematic patterning of a specific linguistic form in use.

In the course of our investigation, we also find that it is necessary to make a distinction between meanings dependent upon "constructional patterns" and meanings that are independent from "constructional patterns." Some meanings are only found in highly fixed linguistic patterns, in the forms of **idioms**, including **substantive idioms** (lexically fixed) and **formal idioms** (with some open lexical slots). The meanings of an idiom is partly dependent upon the structure itself, rather than from its lexical components (Fillmore et al. 1988). Due to its cognitive autonomy from the meaning of its lexical components, an idiom will be listed independently as one single entry.

In sum, we make distinction between senses, forces, and idiomatic senses. In our

---

[5] The term "force" is adopted from Löbner (2002).

analysis, they will be represented with different fonts, following the hierarchy below:

(21) Hierarchy of sense listing convention in this dissertation
    SENSE_1
        [force_1]
        [force_2]
    SENSE_2
    (IDIOMATIC_SENSE_1)
    (IDIOMATIC_SENSE_2)

Senses are written in small capitals, forces are in regular fonts placed within brackets, and idiomatic senses are in small capitals surrounded by parentheses.

3.4.2. Constructing the semantic network

Identification of senses does not tell us whether the senses are principled extensions from a prototype or accidental sharing of the same linguistic form. To verify the relations between the senses, the central core, the 'prototype,' needs to be identified, and then the membership of other senses are judged according to their resemblance to this core.

3.4.2.1. Identifying the prototype

Principled Polysemy (Tyler and Evans 2001, 2003, 2004; Evans 2003, 2005) postulates that a complex linguistic category is organized by an idealized protoscene as well as the sense that best instantiates this protoscene (called the "primary sense in Tyler

and Evans 2001, and known as the "sanctioning sense" in Langacker 1987a) [6]. The primary sense may have the following features:

(22) Possible features of the primary sense

(a) **Reciting salience**:

When a speaker is asked "What does the word X mean?" the primary sense is earliest triggered.

(b) ***Earliest attested sense (also called "origination sense")**:

The historically earliest sense is likely to be the primary sense.

(c) ***Predominance in the semantic network**:

The primary sense is likely to be the one whose semantic components are found among other distinct senses.

(d) ***Relations to other [prepositions]**:

There is the so-called "compositional set." For example, over, above, under and below divide various spatial dimensions. The meaning used to contrast one another is more likely to be the primary one.

(e) **Naturalness of predication:**

The primary sense can most gracefully explain the entire semantic network. That is, other senses are predictable on the basis of this sense.

(f) **Socio-phenomenological experiences:**

The primary sense is the one that is most close to our everyday experience of a specific notion.

Three of the criteria are marked by asterisk because they are found to be NOT applicable to our investigation. Diachronic evidence (22b) is not available because Saisiyat (as well as many other Formosan languages) does not have historical linguistic

---

[6] According to Tyler and Evans (2001), many of these criteria suggested in Principled Polysemy are inspired by Langacker (1987a).

documents. Predominance of semantic network (22c) is more applicable when the number of the discrete senses is high, particularly for prepositions and modalities. Let us first see how Tyler and Evans (2001: 108) employ this criterion in their investigation of English *over*.

> *We have identified 14 distinct senses associated with* over. *Of these, eight directly involve <u>the TR being located higher than LM</u>; four involve <u>a TR located on the other side of the LM vis-à-vis the vantage point</u>; and three — covering, reflexive, and repetition—involve multiple TR-LM configurations. Thus, the criterion of predominance suggested that the primary sense of* over *involves a TR being located higher than the LM.*[7]

We underlined some of the phrases which represent schemata that are generated from the senses, and senses sharing the same schema are said to form a "cluster." For spatial relations, it might be much easier to generate a schema due to a large number of senses. For categories with a low number of senses, the judgment of predominance can be very subjective. The other criterion suspended in the present study is compositional sets (22d). This criterion is found to be applicable mainly for adjective, verbs, or prepositions, which have highlighted the "relation" among event components, and are easy to make contrasts with other lexical items. For presupposition-related lexemes, compositional sets have less implication.

We will therefore rely on **Reciting salience**, **Naturalness of predication**, and

---

[7] TR: trajectory; LM: landmark.

Socio-phenomenological experiences to look for the primary sense. **And we maintain that the saliency of a meaning in the native speakers' mental lexicon is a necessary criterion for a prototypical meaning.** The prototype must fulfill the first criterion. Specific procedures taken to find the most salient meaning are:

(23) **Reciting salience**:
    (a) In de-contextualized situations, ask the informants "What does the word X mean?"
    (b) Before showing them any instances, ask the informants "Can you make up an utterance that contains X?"

The other two criteria are optional. They will only be used when we obtain vague or conflicting answers via (23). Specific procedures to look for the meaning with **Naturalness of predication** and **Socio-phenomenological basicness** are listed in (24).

(24) Two optional criteria for the primary sense
(a) **Naturalness of predication**
    Grammatically, for any sense directly links to the primary sense, we should be able to find a context in which the implicature gives rise to the meaning. For any sense that does not link directly to the primary sense, we should be able to trace it to a sense derivable from the primary sense.
(b) **Socio-phenomenological experiences**[8]
    (I) If there is metaphorical relation involves, consider the followings:
        (i) Heine et al. (1991): PERSON>OBJECT>PROCESS>SPACE>TIME>QUALITY
        (ii) Langacker (1987a): PHYSICAL INTERACTION > SOCIAL/EXPERIENCIAL

---

[8] Tyler and Evans (2001) in fact did not provide details in terms of this criterion. We rely on other linguists' works to make this criterion operative.

(II) If the senses are within the same domain, consider the followings (based on Langacker (1987a):

(i) change in status: from actual to potential, or from specific to generic

(ii) change in focus: a particular elements stand out as focus of attention

(iii)change in the locus of activity or potency: from a focused onstage participant to an offstage one, or from a specific move to a non-specific, generalized one.

Above-mentioned are the working criteria for identification of the prototypical meaning of a lexical item. We will pay special attention to any instance that cannot be solved by these criteria, and view it as hinting at the possibility that the previous studies are far from perfect.

3.4.2.2. Networking the senses

After the prototype of a linguistic category is identified, we can then proceed to the next step: looking for the relations between the senses. Any sense in a category must have direct or indirect relation with the prototype. However, "relatedness" is a vague notion. To avoid subjective or introspective speculation on the relations between the senses, we employ Croft and Cruse (2004) in which the notion of "construal change" can help us determine the kinds of association between two senses. The types of construal changes suggested in Croft and Cruse (2004) are listed in (25). Two senses are considered related if they exhibit any of the following changes of construal.

(25) Types of construal change

    a) Attention (Salience)

    b) Judgment (Comparison)

    c) Situatedness (Perspective)

    d) Constitution (Gestalt)

Attention (Salience) is also known as the "focus of consciousness" (Chafe 1994). According to Croft and Cruse (2004), it refers to not only the human ability to relate a scene to an existing experience, but also the tendency to 1) select some relevant elements and ignore the others, 2) broaden or narrow the scope of attention in appropriate contexts, 3) adjust the granularity to specific information or to general ones, and 4) to move one's attention across the time to create stative/dynamic contrasts. Typical changes of construals operated upon "attention" are: metonymy, profiling, abstraction, schematization, summary/sequential scanning, fictive motion, etc.

Judgment (Comparison) is a fundamental cognitive ability to understand a "particular" in relation to a "universal." This usually involves alignment of a novel experience to a prior experience (Croft and Cruse 2004). Typical changes of construal operated upon "judgment" are: categorization (in the sense of framing), figure/ground alignment, metaphor, etc.

Situatedness (Perspective), according to Croft and Cruse (2004), refers to the conceptualizer's awareness of his location in the temporal, spatial, epistemic, and cultural contexts. Typical changes of construals operated upon "situatedness" are: vantage point, orientation, subjectivity/objectivity, deixis (epistemic, spatio-temporal, and empathic),

etc.

Constitution (Gestalt) refers to the most basic structure of experience, and usually involves the construction of fragmental sensory experiences into a coherent, meaningful, united one. It includes most image-schemata suggested in Johnson (1990), Lakoff and Turner (1989) and Clausner and Croft (1999), and the so-called force dynamics suggested by Talmy (1988). According to Croft and Cruse (2004), image schemata and force dynamics actually exhibit the characteristics of "domains." In other words, they are the base of operation rather than the mechanisms of operation. However, there are some changes of construals that are related to our choice of gestalt, e.g. construing a bowel as a container (*milk in the bowel*) or a surface (*dust on the bowel*).

In our investigation, we will continuously check the types of relations between senses. Since Croft and Cruse (2004) did not claim their list of construal change to be exhaustive, we also take this chance to see whether there are associations in addition to the above-mentioned four types. When two senses do not have justifiable relation between them, they are very likely to share the same form accidentally, i.e. what we have is a case of homonymy.

3.4.3. Looking for the cover term

As correctly observed by Matthewson (2004) and Lehrer (1992), the use of a natural language for semantic glossing is inevitable. Though we agree that Chindali has nine parameters in constructing their verbs of motion which cannot simply be translated into

'come' and 'go' (Botne 2005), we nevertheless find the texts hardly conceivable when those verbs are glossed by schematic symbols that are not natural to speakers of any language. We hold that language documentation is designed to be readable to its users, and the problem is how we can more adequately represent the meaning of a lexical item by use of a meta-language.

As polysemy is common for all natural languages, using one natural language to gloss the other natural language nevertheless leads to coding a polyseme with a counterpart that is itself multi-functional. Although Lehmann (1982) suggests that a polysemy should be glossed by a cover term in the line of morphemic gloss, the way to look for the cover term is not specified. According to the notion of "prototype equivalent," the best equivalent that can be used as the cover term should exhibit "prototype matching" with the target linguistic form. "Prototype matching," based on Kreszowski (1990b), should at least fulfill the following two features:

(26) Criteria for prototype matching

    (a) Wide coverage of the prototype of the target

        The target form (F1) in the target language (L1) should be covered as extensively as possible by the equivalent (F2) in the meta-language (L2). The degree of coverage can be verified by whether the instances of F1's prototypical sense are preferably translated into F2.

    (b) Mutual matching in their prototype area

        The parts of F2 that match F1's prototype should be F2's prototype. After we get the candidate Chinese equivalents via the steps in (26a), we look into three dictionaries of Modern Mandarin to learn the meanings of the F2s. The three dictionaries are Jìaoyùbù Gúoyǔ Xǐao Zìdǐan (Miniature

Dictionary of the Ministry of Education), Gúoyǔrìbào Cídiǎn (Mandarin Daily News Dictionary), and Shìyi Xúesheng Cídiǎn (Student Dictionary published by Shiyi Bookstore). The order of the senses listed under an entry should have shown their prototypicality. We choose these dictionaries because their intended readers are elementary or junior high students. Their order of sense listing is more likely to be usage-oriented rather than historically-based. [9]

Above mentioned are two criteria to look for the equivalent of F1 in the meta-language on the basis of prototype matching. There are, however, cases wherein F1 and F2 exhibit prototype matching, yet their boundaries differ. For example, Croft and Cruse (2004) suggest that French *corde* and English *rope* have similar prototype; they prototypically refer to a length of strong line made of twisted fibers or other materials. Their boundaries are, nevertheless, different from each other. French *corde* can refer to a thin line of fibers usually encoded in English as *string*, but English *rope* cannot refer to thin strings. Based on the notion of prototype matching, we suggest that when F1 does not have an equivalent that matches it both in prototype and in boundaries, a researcher should be content with prototype matching and neglect the boundary mismatch, accepting *rope* as the best equivalent of *corde*. If a researcher wants to make further testing of boundary matching, we have the following two suggested criteria:

---

[9] We tried using the Chinese WordNet developed by Academia Sinica (http://cwn.ling.sinica.edu.tw/). There are some problems that make us give up using it. Many of the common lexemes are not included, such as Chinese *zài* 'again' and *yě* 'also.' Also, the ordering of senses, although claimed to be from the primary to the derived, is sometimes contradictory to native intuition, and conflictive to other dictionaries. For example, the second sense listed under the entry of *yòu* is 'the second activity in a successive event,' and the third sense 'repetition of an activity' occurs to many native speakers as more prototypical. The Chinese WorNet states that "frequency of a sense" is employed to determine prototypicality; however, whether frequency can represent prototypicality is a controversial issue especially for highly grammaticalized linguistic elements.

(27) Criteria for periphery matching

    (a)   More coverage of F1's extensions

         Most F1's extensions are also F2's senses.

    (b)   Less extensions that are not covered by F1

         F2 may have many extensions of itself, and the extensions should better be

         the senses of F1.

These two optional criteria may be employed when a lexical item has more than one candidate equivalents in the meta-language, and the researcher wishes to use just one of them for the economical consideration of corpus documentation. Sometimes, when there is NO equivalent in the meta-language that reflects prototype-matching, the criteria of periphery-matching can serve as a reference. To learn the degree of periphery matching, we can look up the candidate counterparts in the dictionaries to look for the one that has the most functions similar to that of L1, fewer functions that are not covered by L1.

3.5. Summary

Direct translation of a polyseme is bound to be inconsistent, and the inconsistent translations sometimes hinder us from noticing the links between the uses of the polyseme. According to the categorization view of polysemy, there is usually a central meaning, i.e. the prototype, in a polysemy network. Other uses of the polyseme are associated directly or indirectly with the prototype. To construct the lost links between instances of a polyseme, we suggest a set of procedures to identify its prototype and to

network the derived senses back to this prototype. As it is advisable to gloss a polyseme consistently with one cover term in a corpus, our identification of the prototype has one additional advantage: When a researcher considers using a cover term to gloss a polyseme, the cover term can be sought on the basis of its prototypical meaning.

Chapter 4   The Case Study of Saisiyat *Nanaw*

4.1. Preliminary

Language documentation involves not only collection of linguistic data, but also semantic and structural analyses. In the course of eliciting Saisiyat data, we rely greatly upon the informants' translation to make our analyses: The transcribed Saisiyat texts are translated paragraph by paragraph and then word by word into the meta-language, i.e. Mandarin Chinese. Being non-native-speakers of Saisiyat, we believe direct translation is an inevitable way to learn the functions of unfamiliar linguistic forms. Because most of our major informants are proficient in Mandarin, direct translation has been proved to be very effective in our investigation of Saisiyat.

However, as suggested in Matthewson (2004), direct translation is limited in some aspects; one of the shortcomings of direct translation is the mixture of a linguistic form's senses with its contextual elaborations. In our course of fieldwork, the informants' direct translation of the Saisiyat texts indeed reflects the integration of lexical semantics with contextual considerations. It is common for one lexical item to get multiple Mandarin glosses that are used inconsistently cross-textually. Although inconsistent glossing manifests the richness of meaning inference, context-situated interpretation, as has been pointed out by Riemer (2001), sometimes leads to the polysemous reading of a lexical item that even the native speakers are not aware of.

In this chapter, by focusing on a Saisiyat lexical item, *nanaw*, we show the

difficulties of determining the meaning of a polyseme in the course of language elicitation. This chapter is organized as follows. In Section 4.2, a brief grammatical sketch of *nanaw* is presented. In Section 4.3, we follow the autonomy tests proposed in Section 3.4.1 to identify the senses of *nanaw* in all of the instances we can found in NTU-Formosan and our fieldnotes. Then we identify the prototypical meaning of *nanaw*, and construct the semantic network of *nanaw* according to the relatedness of its senses; the result will be presented in Section 4.4. Via the analysis, we find that the meanings of *nanaw* are highly dependent upon its contexts, and the Saisiyat speakers rely greatly upon contextual clues to disambiguate. Some of the linguistic contexts have become conventionalized with a fixed formal structure, and studies of meaning inevitably have to address the relation between lexical meaning and constructional meaning. Discussions on contextualization and construction formation will be addressed in Section 4.5. Finally, in Section 4.6, we will talk about the unique development of *nanaw*'s semantic network, which leads to seemingly irrelevance of its Mandarin translations.

4.2. A sketch of Saisiyat *nanaw*

This section presents a brief syntactic sketch of Saisiyat *nanaw*. In NTU-Formosan, it has two transcriptions: *nanaw* and *nanao*. When carefully pronounced, it ends with a diphthong /aw/. Throughout the study, this lexical item will be referred to as *nanaw*, but if examples are extracted from the corpus, the original transcription is kept.

Typically, *nanaw* is placed after a noun or a verb to modify its preceding elements,

as shown in (28) and (29) respectively. It cannot be placed in sentence-initial position as shown in (30).

(28)   *sia*   ***nanaw***   *'am*   *rima'*
          3SG.NOM   **NANAW**   FUT   AF.go
          3SG. NOM   **NANAW**   FUT   AF.qù
          'He is the only one who is going there.'
          'Zhǐyǒu ta yào qù.'

(29)   *ko'hael*   *'am*   *rima'*   ***nanaw***
          next.year   FUT   AF.go   **NANAW**
          míngnían   FUT   AF.qù   **NANAW**
          'Next year, (someone is) going (there) in any case.'
          'Míngnian yidìng yào qù.'

(30)*   ***nanaw***   *ko'hael*   *'am*   *rima'*
          NANAW   next.year   FUT   AF.go
          NANAW   míngnían   FUT   AF.qù

Saisiyat verbs carry a "focus marker" to indicate the semantic role of the clausal subject.[1] Three kinds of semantic roles can be in focus: agent, patient, or instrument/benefactive. And *nanaw* can be used in clauses of all these focus patterns, as shown in (31) ~ (33).

(31)   *sia*   *m-a'erem*   ***nanaw***
          3SG.NOM   AF-sleep   **NANAW**
          3SG. NOM   AF-shùe   **NANAW**
          Reading 1: 'He kept sleeping.'
                        'Ta yizhí shùi.'
          Reading 2: 'He only slept (and do nothing else).'
                        'Ta zhǐyǒu shùi (Méi zùo bíedė shì).'

---

[1] Please see Section 1.3.2.5 for details of Saisiyat focus system.

(32) *'ahoe'* ***nanaw*** *ni* *'obay* *tebok-on*
     dog **NANAW** GEN PN kill-PF
     gǒu **NANAW** GEN PN sha-PF
     Reading 1: ''Obay killed only dogs.'
          ''Obay zhǐyǒu shasǐ gǒu (Méi sha bíedė dòngwù).'
     Reading 2: ''Obay kept killing dogs.'
          ''Obay yizhí shasǐ gǒu.'

(33) *rayhil* *si-bay* ***nanaw*** *hi* *'obay*
     money RF-give **NANAW** NOM PN
     qían RF-gěi **NANAW** NOM PN
     Reading 1: ''Obay is the only one that the money is given to.'
          'Qían zhǐyǒu gěi Obay.'
     Reading 2: 'Money is always given to 'Obay.'
          'Qían yizhí gěi 'Obay.'
     Reading 3: 'Merely money is given to 'Obay.'
          'Zhǐyǒu gěi 'Obay qían (Méiyǒu qíta dongxi).'

It has been found that the so-called adverbials in Formosan languages should be understood as verbal predicates (Chang 2004). They exhibit some verbal properties. For example, a verb-like adverbial may carry focus marking or tense-aspect marking, and it can stand alone in a clause. Example (34) illustrates an instance of typical verb-like adverbials in Saisiyat. In the answer of speaker B, the adverbial *maybalblay* 'slowly' alone can constitute a clause. In a negative clause, a verb is realized in its nonfinite root form. If we compare (34) with (35), we can be certain that *maybalbalay* in fact carries an agent focus *m-*.

70

(34) *A:*   *sia*   ***m-aybalbalay***   *ay*   *manraan?*
      3SG.NOM   **AF-slow**   Q   AF.walk
      3SG. NOM   **AF-màn**   Q   AF.zǒu
  B: *'ihi'.*   ***maybalblay.***
      BC   **AF.slow**
      BC   AF.màn
  A: Did he walk slowly?
  B: Yes. Slowly.
  A: Ta mànmàn zǒu guòlái ma?
  B: Dùi. Mànmàn dė.

(35) *sia*   *'okay*   ***paybalbalay***   *manraan*
     3SG.NOM   NEG   **slow**   AF.walk
     3SG. NOM   NEG   màn   AF.zǒu
  'He did not walk slowly.'
  'Ta méiyǒu mànmàn zǒu.'

On the other hand, *nanaw* does not have AF, PF, or RF marking. One might suspect that it is a Ø-marked verb in agent focus. However, it differs from verb-like adverbials in that *nanaw* alone cannot constitute a clause, and this is evident by the contrast between (34) and (36).

(36) *A:*   *'obay*   *s<om>i'ael*   ***nanaw***   *ka*   *'aelaw*   *ay?*
      PN   eat<AF>   **NANAW**   ACC   fish   Q
      PN   chi<AF>   **NANAW**   ACC   yú   Q
  *B: *'ihi'.*   ***nanaw.***
      BC   **NANAW**
      BC   **NANAW**
  A: Does 'Obay eat only fish?
  *B: Yes. Only.
  A: 'Obay zhǐ chi yúma?
  *B: Dùi. Zhǐyǒu.'

Although many adverbials in Saisiyat have verbal properties, *nanaw* as an adverbial does not behave like a verb. It cannot be affixed with focus markers, and is more dependent upon its co-occurring verb than verb-like adverbials are.

4.3. Senses of *nanaw*

Following the criteria of sense judgment proposed in Chapter 3 (§ 3.4.1), we analyze all instances of *nanaw* in NTU-Formosan and in direct elicitation, classifying them into senses. The senses we have identified are:

LIMITATION

CONTINUATION

    [persistence]

    [inherence]

(NEG_EXTREME)

Two senses pass the autonomy tests: LIMITATION and CONTINUATION. Two forces, [persistence] and [inherence], are subsumed under CONTINUATION. Besides, there is one constructional pattern that occurs as an idiom: (NEG_EXTREME).

4.3.1. LIMITATION

The LIMITATION sense includes cases in which "**the scope of predication is limited to a specific entity/activity/state/quantity as opposed to other possibilities, usually less than expected.**" An example is given in (37), which is extracted from a Saisiyat folktale. A mysterious woman came from water and married a Saisiyat man; they had a child. After a quarrel with Saisiyats, the lady decided to go back home. She cut the child into two; she got the head, and her husband got the body.

(37) kathethel

184. ... | *hiza* | *noka* | *kamamanra:an* |
| --- | --- | --- |
| that | GEN | male |
| nàgė | GEN | nánrén |

185. .. | *p\<in\>a-'apol* | *ila* | *naw²* | *basang* | **nanaw** |
| --- | --- | --- | --- | --- |
| CAU-divide\<PFV\> | PFV | only | body | **only** |
| CAU-fen\<PFV\> | PFV | zhĭ | shentĭ | **zhĭ** |

'\<in\>oka'	*ka*	*ta'oeloeh*
NEG\<PFV\>	ACC	head
NEG\<PFV\>	ACC	tóu

'The man was given the child's body without a head.'
'Gěi nánrén dė xĭaohái shentĭ méiyŏu tóu.'

In (37), *nanaw* is placed after a noun *basang* 'body' to mean that the scope of the child distributed to the man is limited to "body," and does not include the "head "of the child.

We also find many instances in which *nanaw* is placed after a verb, indicating that among a wide variety of potential activities, the agent does merely a specific activity, as exemplified in (38).

(38) | *sia* | *mato:o* | *ila* | *tinal'oemaeh* | |
| --- | --- | --- | --- | --- |
| 3SG.NOM | thirty | PFV | year | |
| 3SG.NOM | sanshí | PFV | nían | |

'ana	*kano'*	*ma'*	*okik*	*ra:am*
no.matter	what	also	NEG	know
wúlùn	shémė	yě	NEG	AF.zhidào

kito'yaen	*h\<oem\>angi'*	**nanaw**
AF.hungry	cry\<AF\>	**NANAW**
AF.è	ku\<AF\>	**NANAW**

'He is 30 years old. He cannot do anything. He just cries when he gets hungry.'
'Ta sanshí sùe lė, shémė dou búhùi. Dùzi è lė zhĭhùi ku.'

---

² This *naw* might be a false start.

In example (38), *nanaw* modifies its preceding verb *hoemangi'* 'to cry,' indicating that the agent's capability is limited to "crying," and he can do nothing else.

In addition to noun and dynamic verb, *nanaw* is sometimes found to be placed after a stative verb. For example, in (39), *nanaw* modifies *'al'alak*, a stative verb meaning 'young.'

(39)   *'obay*   *'al'alak*   ***nanaw***
        PN   AF.young   **NANAW**
        PN   AF.nínqing   **NANAW**
        ''Obay is still young.'
        ''Obay hái hěn nínqing.'
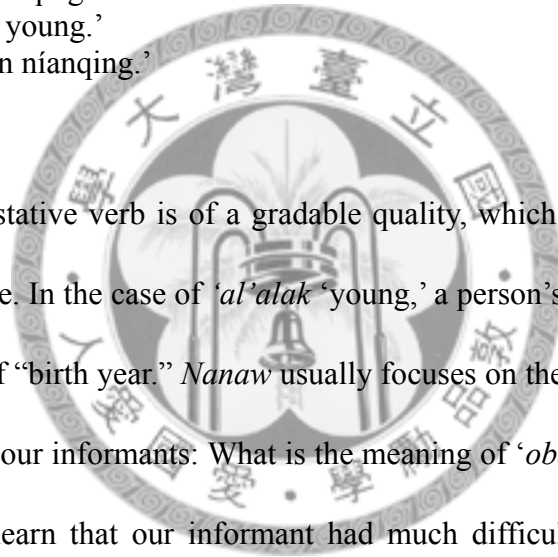
In this case, the stative verb is of a gradable quality, which can be transferred into some quantifiable scale. In the case of *'al'alak* 'young,' a person's state of being young is countable on a scale of "birth year." *Nanaw* usually focuses on the lesser end of the scale. We once asked one of our informants: What is the meaning of *'obay tatini' nanaw* ('Obay old *nanaw*), only to learn that our informant had much difficulty comprehending this utterance. Here we need to point out that in Saisiyat, no sharp grammatical distinction is made between quality and activity: They are all expressed as verbs: the former as stative verbs and the latter as dynamic verbs.[3] What are known as "adjectives" in English should be deemed stative verbs in Formosan languages (Zeitoun 2000; Zeitoun and Huang

---

[3] Nouns in Saisiyat functions as nominal predicate as in the following example:
    kizaw   'in-sia-a   minkoringan
    PN   POSS-3SG-POSS   wife
    'Kizaw is his wife.'
The division between word classes made in Indo-European language studies is blurry in Saisiyat.

2000).

In addition to nouns, active verbs, and stative verbs, *nanaw* is also collocated with quantifiers or numbers to indicate that a quantity is less than an expectation. Example (40) is an instance in which the speaker is talking about the number of the mushrooms she planted. The speaker uses *nanaw* to modify the quantifier *ititi'an* 'a bit,' meaning that the quantity is low.

(40) Life

142.　F:　*ititi'an*　　***nanao***　*ka*　　*p<in>amowa'*
　　　　　a.bit　　　**only**　NOM　plant<PFV>
　　　　　yidĭan　　**zhĭ**　　NOM　zhòng<PFV>
　　　'We plant very few (mushrooms).'
　　　'Wŏmėn zhòng dė xiangku hĕn shăo.'

And (41) is an instance in which *nanaw* modifies a numeral predicate. A dog typically has four legs, and the use of *nanaw* in (41) indicates that "three" is a number less than what we typically expect based on our understanding of dogs.

(41)　*hini*　　*'ahoe'*　*to:o*　***nanaw***　*ka*　　*tatakay*
　　　this　　dog　　three　**NANAW**　NOM　leg
　　　zhè　　gŏu　　san　　**NANAW**　NOM　jĭao
　　　'This dog has only three legs.'
　　　'Zhè zhi gŏu zhĭyŏu san zhi jĭao.'

The numerals can function as verbs in addition to nouns in Formosan languages (Li 2006). For example, a numeral predicate *roSa'* 'two' in IU 481 can have referential focus marking and aspect marking as shown in (42), and it can also undergo nominalization just like a verb, as in (43).

(42) Kathethel

481. ...(0.8)   *lasia*   *o:*   **si-\<in\>roSa'**   *ka==*
             3PL.NOM   INT   **RF-two\<PFV\>**   ACC
             3PL. NOM INT   **RF-lǐanggè\<PFV\>**   ACC

482. ...   *korkoring*
             child
             xǐaohái

483. ..   *ti-roSa'-en*   *nasia*
             cut-two-PF   3SG.GEN
             qie-èr-PF   3SG. GEN
    'They cut the child in two.'
    'Tamèn bǎ xǐaohái qie chéng lǐangbàn.'

(43) Life

12. M: ...(1.8)   *a==*
               FIL
               FIL

13. M: ...(0.9)   *koza'*   *ka-sinpan-an*   *kin*   *'akoy*
              how.much   KA-raise-NMZ   very   many
              duoshǎo   KA-yǎng-NMZ   hěn   duo
    'How many (animals) do you raise?'
    'Ah, na nimén yǎng hěn duo ma?'

14. B: ...(1.0)   *sinpan-an==*   **kina-roSa'-an**
              raise-NMZ   **NMZ-two-LF**
              yǎng-NMZ   **NMZ-èr-LF**

15. B: ...(1.0)   *e*   *...*   *mato:ol*
              FIL   thirty
              FIL   sanshí
    'They are kept in two places, thirty in all.'
    'Yǎng zài lǐanggè dìfang, zǒnggòng yǒu sanshí zhi.'

No matter modifying a verb, a noun, a stative verb, a quantifier, or a numeral predicate, *nanaw* always focuses on the lesser end of a scale, meaning that the kinds of activities, the types of entities, or the numbers, are less than an explicit or implicit standard.

Instances of the LIMITATION sense usually receive a Mandarin translation *zhǐ(yǒu)* 'only.' In addition, the informants sometimes use *cái* to translate instances that express a

limited quantity. The following example (44) is taken from our fieldnotes.

(44) | *sia* | *to:o* | ***nanaw*** | *tinal'oemaeh* | *'ana* | *kano'* | *ma'* | *haSa'* |
| 3SG.NOM | three | **NANAW** | year | whether | what | also | unable |
| 3SG.NOM | san | **NANAW** | nían | wúlùn | shémė | yě | búhùi |

'He is just three, and he cannot do anything.'
'Ta cái san sùi, wúlùn shémė dou búhùi.'

When an utterance involves the speaker's intention to counter a belief or a statement, the informants tend to use *cái* as the direct translation of *nanaw*. In the case of (44), there might be someone expecting a three-year-old to do a specific activity.

4.3.2. CONTINUATION

The CONTINUATION sense encompasses instances in which "**an activity or state extends or iterates over a period of time.**" It has two forces: [persistence] and [inherence].
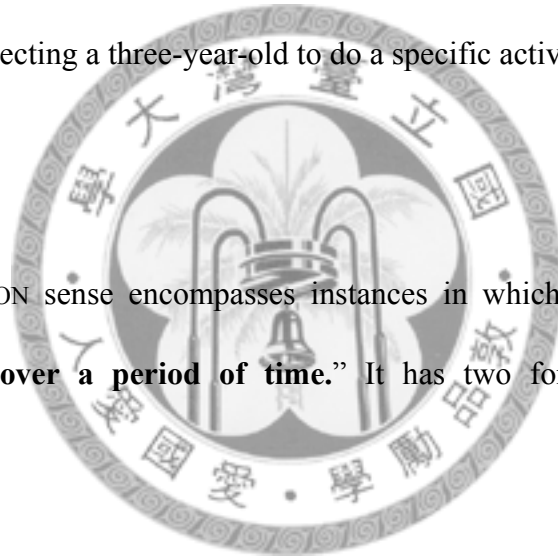
[persistence]

Example (45) is an instance of CONTINUATION. It is extracted from a text in which the speaker is narrating a story about flooding. There were heavy rains and the lands were flooded. And example (45) literally means that in a long period of time, 'the water and rain remained abundant.'

(45) flood
84.   ...(1.0)   *o:*
                FIL
                FIL
85.   ...   *'akoy*   ***nanaw***   *ka*   *ralom*   *ki*   *'a'oeral*
           AF.many   **still**   NOM   water   and   *C*a-rain
           AF.duo   **réngrán**   NOM   shǔi   hàn   *C*a-yǔ
      'It kept on raining.' '[4]
      'Yizhí xìa yǔ.'

In direct elicitation, we came across a number of instances of *nanaw* that collocates with future marker *'am* to denote a strong volition to carry out an action. In Saisiyat, the future marker *'am* is used to express a wide variety of irrealis events, including obligation, commission, and volition. Its collocation with *nanaw* denotes the subject's continuous volition, such as (46).

(46)   *yao*   *'am*   *rima'*   ***nanaw***
       1SG.NOM   FUT   AF.go   **NANAW**
       1SG.NOM   FUT   AF.qù   **NANAW**
       'I am going (there) in any case.'
       'Wǒ yídìng yào qù.'

Instances such as (46) are considered instances of [persistence]. It seems that the meaning of *'am* has a strong collocational effect on *nanaw,* and the meaning of willingness or determination is merely pragmatic elaboration.

One thing needs to be pointed out is that the grammatical environment of CONTINUATION partially overlaps with that of LIMITATION. Their partially sharing of the

---

[4] The free translation of (45) does not pair with the linguistic form. Because NTU-Formosan collects many long texts with complicated plots, translations with rich contextual elaboration can help corpus users to appreciate the plots in a more comprehensive manner.

same grammatical structure sometimes leads to ambiguity. For example, (39) that we have previously mentioned in fact has a second reading, as shown below.

(39') *'obay*   *'al'alak*   ***nanaw***
       PN   AF.young   **NANAW**
       PN   AF.níanqing   **NANAW**
       Reading 1: ''Obay is just a young lad.' (referring to his age)
               ''Obay hái hěn níanqing.'
       Reading 2: ''Obay remains young.' (referring to his energetic state)
               ''Obay yizhí dou hěn níanqing.'

These two interpretations apply to different "aspects" of youth. Reading 1 is classified into the LIMITATION sense, and it triggers "the low number of living years" as one aspect of being young. On the other hand, Reading 2 is classified in the CONTINUATION sense, and it triggers "the energetic state of human body" as another aspect of being young.
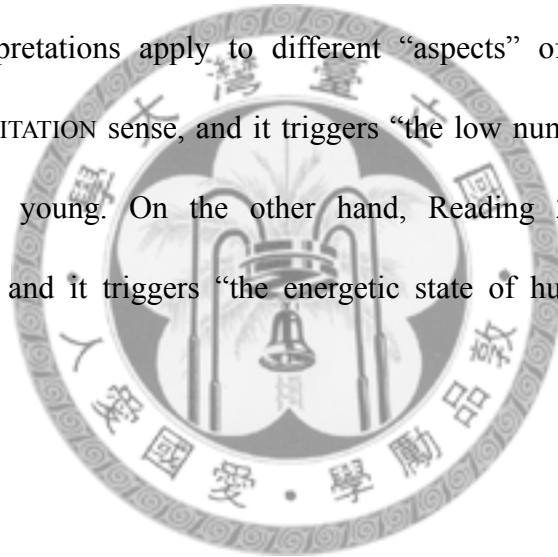
[Inherence]

The [inherence] force refers to the function of *nanaw* to be used as a marker to indicate the speaker's judgment on the factual status of a statement. For example, (47a) denotes an activity that happened, and by adding *nanaw*, (47b) means that according to the speaker's judgment, for this activity to happen is a norm.

(47)a.   'oem'oemaeh   *se'el-en*   *noka*   hinpoetoeh
        dry.land   grow-PF   GEN   weed
        hàndì   zhǎng-PF   GEN   zácǎo
        'The dry land grew weeds.'
        'Hàndì zhǎng le zácǎo.'

b.   'oem'oemaeh   *se'el-en*   ***nanaw***   *noka*   hinpoetoeh
        dry.land   grow-PF   **NANAW**   GEN   weed
        hàndì   zhǎng-PF   **NANAW**   GEN   zácǎo
        'Dry lands of course grow weeds.'
        'Hàndì jiùshì huì zhǎng zácǎo.'

Nanaw in its [inherence] sense is an evidential marker. It denotes the speaker's affirmative attitude to a statement. By using *nanaw*, the speaker conveys his strong belief that a proposition is true. We suggest that the difference between [persistence] and [inherence] lies in the scope of viewing. When the duration of perception time is specified or implied in the context, the activity is understood as occupying a specific temporal frame. When a wider scope of perception is taken, the interpretation is relevant to the "atemporal" generic nature of the subject. Experientially, there is a tangible link between the two interpretations: An activity that keeps happening in a temporal frame is likely to be a norm in the physiological world, and vice versa.

Nanaw in its [inherence] force also modifies expressions of negative voice. For example, in (48), the speaker is lamenting that there has not been any Saisiyat who has been elected representative ever since the Japanese colonial period. He uses *nanaw* in IU200 to emphasize the low number of Saisiyat people elected as representatives.

(48) election

198. M: ... *'ae-'aering-an* ... *noka rippon o: tabin haysani*
          RED-start-NMZ   GEN PN   DM until   now
          RED-kaishǐ-NMZ   GEN PN   DM zhídào xiànzài
   'From the Japanese Occupation until now,'
   'Cóng Rìjùshídài yizhí dào xiànzài…'

199. M: ...(2.3) *mayisaa*
          pass.that.way
          jìngguònàlǐ

200. M: ...(1.3) *'oka'* **nanaw** *koSa'en* .. *ina* ...
          NEG **still**   FIL   FS
          NEG **réngrán** FIL   FS
          ka SaiSiyat min'itol o: pa-sa:eng-en
          FIL PN   AF.register   DM CAU-sit-PF
          FIL PN   AF.dengjìcanxǔan DM CAU-zùo-PF
   'there has not been any SaiSiyat elected to office.'
   'hái méiyǒu yigė canxǔan dė Sàixià rén dangxǔan gùo.'

Instances of *nanaw* occurring in negation construction are numerous. Saisiyat uses different negators in different contexts, summarized in Table 4.1.

Table 4.1. Negators in Saisiyat (following Yeh 2000a: 121)

Negator	Function	Collocates
'oka'	negating existence	noun, verb, (answering yes/No question)
'okik	negating a property	nominal predicate, statics, dynamic verbs with aspect markers
'okay	negating existence of an activity	dynamic verbs
'izi'	prohibition	dynamic verbs
kayni	negating a volition	nouns and verbs

In (49), we show that *nanaw* can modify all these negators, emphasizing that an activity is never a fact or should never be done.

(49) a.
sia	*'<in>oka'*		*nanaw*	*'i-ra'oe*
3SG.NOM	**NEG<PFV>**		**NANAW**	NEG-drink
3SG.NOM	**NEG<PFV>**		**NANAW**	NEG-he

'He never drinks (wine).
'Ta cónglái méiyǒu he (jǐu).'

b.
kizaw	*'okik*	*'ina*	*sarara'*	*nanaw*	*hi*	*'obay*
PN	**NEG**	EXPER	like	**NANAW**	ACC	PN
PN	**NEG**	EXPER	xǐhuan	**NANAW**	ACC	PN

'Kizaw had never loved 'Obay.'
'Kizaw cónglái méiyǒu xǐhuan gùo 'Obay.'

c.
sia	*'<in>okay*	*kaat*	*nanaw*
3SG.NOM	**NEG<PFV>**	write	**NANAW**
3SG.NOM	**NEG<PFV>**	xǐe	**NANAW**

'He had never written a word.'
'Ta cónglái méiyǒu xǐe gùo zì.'

d.
'izi'	*ila*	*nanaw*	*i'-pa'erem*
NEG	PFV	**NANAW**	NEG-sleep
NEG	PFV	**NANAW**	NEG-shùi

'Don't fall asleep at any rate.'
'Júedùi bù kěyǐ shùi.'

e.
yao	*kayni*	*ila*	*nanaw*	*'iso'on*
1SG.NOM	**NEG**	PFV	**NANAW**	2SG.ACC
1SG.NOM	**NEG**	PFV	**NANAW**	2SG.ACC

'I don't want to be with you anymore.'
'Wǒ búyào gen nǐ zài yiqǐ lě.'

If we regard *nanaw* as an intensifier that marks the factual status of a statement at a high point on a positive scale, its occurrence with negative structures should then be understood as "the speaker's placement of the statement on the corresponding negative (unfavorable) scale," as explained by Horn (1989).

When modifying a nominalized activity, *nanaw* indicates that an entity is by nature predisposed to perform a specific action. For example, (50) is taken from a conversation.

(50) life

108.	F:	...	*koSa'en*	*kama-saseez*	***nanao***	*alikah*	*hayza'an*	*mita'*
			DM	KAMA-small	**originally**	fast	past	1IPL.GEN
			DM	KAMA-xĭaoxíng	**yúanběn**	kùai	gùoqù	1IPL.GEN
109.	F:	..	*s<om>pan*	*ka*	*tata:a'*	*isaa*	*kama-saseez*	***nanao***
			raise<AF>	ACC	chicken	that	KAMA-small	**originally**
			yăng<AF>	ACC	ji	nà	KAMA-xĭao	**yúanběn**
110.	F:	...	*'ana-tabo:-on*	*ma'*	*hinghae'*	*ma'ngel*	*sobae:oeh*	.\
			whatever-feed-PF	also	same	slow	grow	
			wúlùn-wèi-PF	yě	yiyìang	màn	zhăng	

'The chicken we raised before were small and it took long for them to grow, whatever you feed them.'

'Wŏmėn zhiqían yăng dė ji dou hěn jĭu cái zhăng dà, bùgŭan ni wèi tamėn shémė.'

The verb is nominalized, modified by *nanaw* to indicate that the subject has a natural disposition to perform a specific action. The use of *nanaw* in IU108 and IU110 denotes that the chickens are of a small size by nature. In direct elicitation, we also come cross several instances following the same grammatical pattern. For example, (51) indicates that snow is something with a natural tendency to melt. On the other hand, (52) is not acceptable to the native speakers, and their explanation is: *Patonay 'okay sahoewi' 'okik e:Sem* 'Iron, if not heated, does not melt.'

(51)	*hahoera'*	*kama-e:Sem*	***nanaw***
	snow;ice	NMZ-melt;dissolve	**NANAW**
	shŭe;bing	NMZ-róng;róngjĭe	**NANAW**

'Snow is the thing that melts (by nature).'
'Shŭe běnlái jìushì hùi rónghùa dė dongxi.'

(52) ?	*patonay*	*kama-e:Sem*	***nanaw***
	iron	NMZ-melt;dissolve	**NANAW**
	tĭe	NMZ-róng;róngjĭe	**NANAW**

Saisiyat lexical nominalization adopts verbal focus markings, and has four foci: genitive, patientive, referential, and locative focus.[5] [Inherence] modifies a nominalized activity by all of the four foci, as shown in (53).

(53)
Nominalization by agent focus
Pattern: *kama*-V   *nanaw*
Example: *babuy*   *kama-s<om>i'ael*   *'akoy*   *nanaw*
                pig   NMZ-eat<AF>   AF.many   **NANAW**
                zhu   NMZ-chi<AF>   AF.duo   **NANAW**
                'Pigs are those supposed to eat a lot.'
                'Zhu běnlái jìushì hěn hùi chi de.'

Nominalization by patient focus
Pattern: *ka(k)*-V-*en*   *nanaw*
Example: *hini*   *boway*   *ka-si'ael-en*   *nanaw*
                this   fruit   NMZ-eat-PF   **NANAW**
                zhè   shŭigǔo   NMZ-chi-PF   **NANAW**
                'Those fruits are the thing supposed to be eaten.'
                'Zhè shŭigǔo běnlái jìushì yào chi dè.'

Nominalization by referential focus
Pattern: *ka(k)*-V   *nanaw*
Example: *hini*   *ka-ko:as*   *nanaw*
                this   NMZ-comb   **NANAW**
                zhè   NMZ-shu   **NANAW**
                'This is the thing used to comb hair.'
                'Zhè jìushì yùnglái shutóu dè.'

Nominalization by locative focus
Pattern: *ka(k)*-V-*an*   *nanaw*
Example: *habaan*   *kak-bayoS-an*[6]   *nanaw*
                summer   NMZ-typhoon_blow-LF   **NANAW**
                xìatian   NMZ-gua.táifeng-LF   **NANAW**
                'Summer is the season that has typhoons.'
                'Xìatian běnlái jìushì táifeng jìjíe.'

---

[5] Examples in (53) are so-called "lexical nominalization" in Yeh (2000b). Although locative focus does not occur as verbal focus marking in modern Saisiyat, it is still used in lexical nominalization.
[6] Typhoon, as a noun, is *bayoSon*.

Nominalization in Saisiyat itself carries a reading that an entity shows expertise in doing an activity. The tendency for *nanaw* to co-occur with nominalization construction certainly has some cognitive implications. Compared with activities, entities are typically presupposed and self-contained; to nominalize an activity to some extent implies that the activity is a presupposed fact according to our socio-physiological knowledge.

4.3.3. Idiomatic sense: (NEG_EXTREME)

The other idiomatic sense, (NEG_EXTREME), denotes that "**the quantity of an implied entity is at the lowest extreme, usually in a negative situation.**" (NEG_EXTREME) is classified as an idiomatic sense because it constitutes a fixed constructional pattern with the negator *'oka'*, usually together with aspect marker *ila*.[7] Example (54) is extracted from a frog story.

(54) frog3
97. ... *sahae'*   *ray==*
         AF.fall   LOC
         AF.dìao   LOC
98. ... *ray kaehoey babaw sahae' ila== ray== rapoe hini'*
         LOC   tree   up   AF.fall PFV   LOC   ground here
         LOC   shù   shàng AF.dìao PFV   LOC   dì   zhèli
      '(The boy) fell on the ground from up the tree.'
      '(Ta) cóng shùshang dìao dào dì shàng.'
99. ... *kita'-en hini' 'oka' ila **nanaw** hini' 'ima sahae' ray==*
         see-PF   this   NEG   PFV **NANAW** here   PROG AF.fall LOC
         kàn-PF   zhège NEG   PFV **NANAW** zhèlǐ PROG AF.dìao LOC
      'It seems that there is nothing that can be done.'
      'Sìhu méiyǒu shémė bànfã.'

---

[7] According to Yeh (2000a, 2003), the marker *ila* is used to denote inceptives, and it also co-occurs with perfective aspect marker *<in>* and future marker *'am* to indicate that a new state is achieved or a new state is going to come about. In NTU-Formosan, it is analyzed as a perfective marker. Its collocation with *'oka' nanaw* has an effect: to call for the hearer's attention to the speaker's judgment.

The use of *nanaw* in IU 99 denotes a situation in despair. In direct elicitation, we came across a lot of instances of (NEG_EXTREME). For example, (55a) means that 'Obay is very sick, and according to the speaker's speculation, there might be no remedy. And (55b) indicates that the speaker is so hungry that he is about to be out of patience.

(55)a. *'obay   kak'ayeah   'aboe'   'atomalan*
      PN   AF.sick   AF.serious   very
      PN   AF.shengbìng   AF.yánzhòng   feicháng
      'am   'oka'   ila   nanaw   ra:am-en
      FUT   NEG   PFV   NANAW   know-PF
      FUT   NEG   PFV   NANAW   zhidào-PF
      ''Obay shengbìng hěn yánzhòng, kěnéng méi jíu lė.'
      ''Obay is very sick.There is no known remedy.'
   b. *yao   'oka'   ila   nanaw   'a-kito'yaen*
      1SG.NOM   NEG   PFV   NANAW   'a-hungry
      1SG.NOM   NEG   PFV   NANAW   'a-è
      'I am so hungry that I can't bear it anymore.'
      'Wŏ è dė shòubùlĭao lė.'

Instances of (NEG_EXTREME) usually depict a desperate situation: running out of remedy, money, hope, means, and so on. Note that *'oka'* typically negates the existence of an entity or an activity, but in instances of *'oka' nanaw*, the entity or activity being negated is linguistically unrealized. For example, in (55a), there is no explicit linguistic form for "remedy." In (55b), the implicit noun might be 'patience.'

4.4. The semantic network of *nanaw*

When classifying the instances of *nanaw* according to their functions, we find that its senses are closely related to each other, constituting an interconnected semantic network. This section presents how the senses of *nanaw* are chained to each other as a "family."

4.4.1. Establishing the prototype

Following the procedures proposed in Chapter 3, we look for the prototype of the linguistic family primarily based on reciting salience. If we obtain conflictive answers when querying the most salient meaning, we will employ the other two criteria, naturalness of predication and socio-phenomenological basicness, as subsidiary evidence.

To learn which sense is more salient, we ask the informants "What is the meaning of *nanaw*" in de-contextualized situations. The informants with little hesitation answer with Mandarin equivalent *zhǐyǒu*, the translation that is used for instances of LIMITATION. When they are asked to make up an utterance that contains *nanaw*, they also have a strong preference to give instances of LIMITATION. Specifically, the utterances they make up tend to contain numeral predicates or quantifiers, such as (40) and (41).

Although we assume a level of "sense" that should reflect conceptual discreteness, the utterances made up by the informants tend to converge on a specific type of structure. In the case of *nanaw*, a specific type of the LIMITATION sense collocating with quantity-related verbs receives cognitive saliency in the speakers' reciting practice. The

tendency for the prototype to be reflected at a low level is an issue which we will come

back to after we examine other cases of polysemy.

4.4.2. Networking the senses

The construction of semantic network is based on four types of construal changes

suggested in Croft and Cruse (2004): attention (salience), judgment (comparison),

situatedness (perspective), and constitution (gestalt). For detailed discussion on the four

types of construal changes, please see Section 3.4.2.2.

By linking the senses with these cognitive mechanisms, we find that the senses of

nanaw constitute a semantic network, as shown in Figure 4.1. In other words, the various

senses of *nanaw* should be polysemy rather than homonymy.
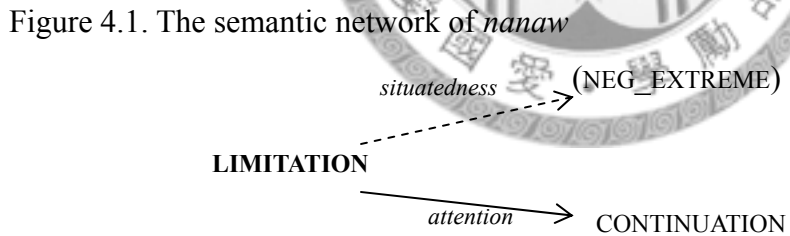
Figure 4.1. The semantic network of *nanaw*

In this section, we will explain the links in detail, attempting to build the semantic

network of *nanaw*, following the spirit of "family resemblance" (Wittgenstein 1963).

From LIMITATION to CONTINUATION

Although a specific type of the LIMITATION sense receives particular reciting saliency,

the extension of *nanaw* seems to be collaborative efforts of more than just the most

salient meaning, i.e. limitation of a quantity. The instances of LIMITATION that modifies a

verb — numeral, dynamic, and stative — might have jointly contributed to the extension

of *nanaw* to denote CONTINUATION. When modifying a verb, *nanaw* in its LIMITATION

sense indicates that an activity is the delimited scope as opposed to other potentials, and

CONTINUATION focuses on that delimited activity by narrowing down the viewing scope

and neglecting other potentials. Their respective viewing arrangements are sketched in

Figure 4.2

Figure 4.2. Perspective-taking and the relation between LIMITATION and CONTINUATION

(a) LIMITATION   (b) CONTINUATION
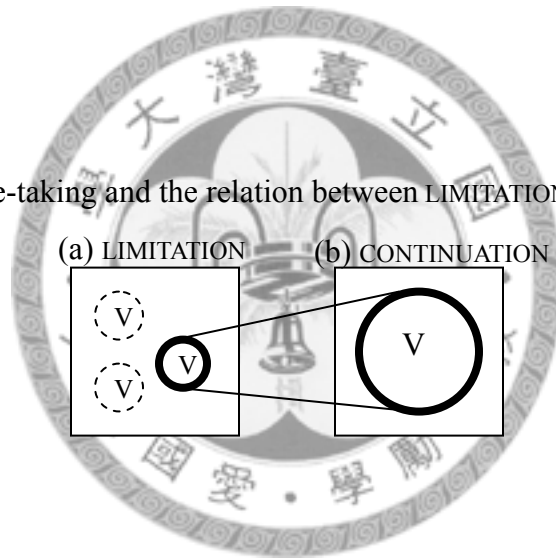
The LIMITATION sense means that a focused activity is a limited scope of predication

against other possibilities, as shown in Figure 4.2(a). And CONTINUATION makes a

different construal on the same scene — by focusing only on this delimited activity, as

shown in Figure 4.2(b). The only activity that receives our full attention is likely to be a

preoccupation or a habit of the subject, and in an atemporal context, it can be understood

as the nature or the fact of the world. The change of viewing scope and adjustment of

attentional focus exhibit the type of construal change known as **attention (salience)** in

Croft and Cruse (2004).

From LIMITATION to (NEG_EXTREME)

The relation between (NEG_EXTREME) and LIMITATION is evident. They both make judgments on a limited quantity, but (NEG_EXTREME) is different in three aspects: a) The quantity is at the lowest extreme, i.e. naught, b) The entity being quantified is implicit, and c) It always expresses a negative evaluative attitude. We find that the link between (NEG_EXTREME) and LIMITATION is motivated by **situatedness (perspective)**.

The LIMITATION sense indicates that a quantity is limited to a specific scope, and is usually less than expected as compared to other potentials. On a quantifiable scale, LIMITATION is prone to take the lesser end. A "low number" is experientially linked to "negative outcome." In instances of (NEG_EXTREME), this inference of "negative outcome" is strengthened and emphasized, when other semantic elements are deemphasized or neglected. Also, the speaker is becoming more and more engrossed in his evaluation on the delimited quantity, using it to make negative assessment of a situation, usually based upon his personal inferences, showing salient tendency of subjectification, motivated by **situatedness (perspective)**.

In Figure 4.1, we graphed (NEG_EXTREME) as linked to LIMITATION with a dashed arrow, meaning that their relations are relatively weaker compared to the links of other senses.

4.5. Semantics, pragmatics, and syntax at crossroads

In our investigation of *nanaw*, its meaning is always intertwined with the contexts it occurs in. Meanings may also be conventionalized with a specific grammatical pattern. The understanding of a linguistic form necessarily involves our understanding of its contextual constraints as well as its grammatical structures. As suggested by Nerlich and Clarke (2003), studies of polysemy give us an opportunity to view semantics, pragmatics, and syntax as a trinity. This section will demonstrate how the meanings of *nanaw* are integrated with the contextual and constructional environment.

4.5.1. Meaning in contexts

Wittgenstein (1963) rightly states in his *Philosophical Investigations* that "the meaning of a word is its use in a language"— Understanding the meaning of a linguistic form in terms of its contexts is far from a new proposal. There are typically two kinds of contexts: linguistic context and non-linguistic context (Lyons 1977). Following Lyons (1977) and supplemented by Croft and Cruse (2004), the contextual constraints that influence our understanding of a linguistic form may include the following:

(56) Types of contexts
    (I) Linguistic context
        a)   knowledge of the language
        b)   previous discourse
        c)   immediate linguistic environment

(II) Non-linguistic context

    a)   situational context: the time and place, topic, register, relation between participants

    b)   background knowledge: social convention, conversational maxims, common sense of the objective world, mutual-understanding between participants

We find that **knowledge of the language**, **immediate linguistic environment**, and **background knowledge** are particularly relevant to interpretations of *nanaw*. We will talk about knowledge of the language and background knowledge in this section. Immediate linguistic environment, i.e. the phrasal or sentential structure a linguistic form occurs in, has special impacts on the interpretation of *nanaw*, and we will thus save the discussion in Section 4.5.2.

For **knowledge of the language**, here we would like to point out that the position of *nanaw* in an utterance greatly determines its interpretation. *Nanaw* typically modifies its preceding phrase (an NP or VP). In example (57a), *nanaw* modifies the noun phrase that directly precedes it, and *nanaw*'s scope of modification never covers the elements at its left side. In (57b), *nanaw* modifies the preceding verb phrase, but the scope does not cover the subject NP. In (57c), *nanaw* is placed within a VP, and it modifies the entire VP. The scope is limited to the verb phrase, and does not cover the subject NP.

(57) a. *sia*   ***nanaw***   *malmalay*   *m-wa:i*
      3SG.NOM   **NANAW**   AF.slow   AF-walk
      3SG.NOM   **NANAW**   AF.màn   AF-zǒu
      'He was the only one that walked slowly.'
      'Zhǐyǒu ta mànmàn dė zǒu guòlái.'
      *'He kept walking slowly.'
      *'Ta yizhí mànmàn dė zǒu guòlái.'

    b. *sia*   *malmalay*   *m-wa:i*   ***nanaw***
      3SG.NOM   AF.slow   AF-walk   **NANAW**
      3SG.NOM   AF.màn   AF-zǒu   **NANAW**
      'He kept walking slowly.'
      'Ta yizhí mànmàn dė zǒu guòlái.'

    c. *'obay*   *k<om>ita'*   ***nanaw***   *hi*   *kizaw*
      PN   see<AF>   **NANAW**   ACC   PN
      PN   kàn<AF>   **NANAW**   ACC   PN
      Modifying V: ''Obay kept looking at Kizaw.'
                 ''Obay yizhí kàn Kizaw.'
      Modifying V: ''Obay merely looked at Kizaw (and he didn't do anything else).'
                 ''Obay zhǐyǒu kàn Kizaw.'
      Modifying N: ''Obay merely looked at Kizaw (and at no one else).'
                 ''Obay zhǐyǒu kàn Kizaw.'
      * 'Only 'Obay (and nobody else) saw Kizaw.'
      * 'Zhǐyǒu 'Obay kàndào.'

Nevertheless, the position of *nanaw* is more indicative of the meaning of the utterance only in agent-focus clauses. In PF and RF clauses, the scope of modification is relatively flexible. Although there is a tendency for *nanaw* to modify the elements that precede it, the tendency can be violated. For example, in (32) and (33) presented in Section 4.2.1, the meaning is relatively unpredictable.

For **background knowledge**, the interpretation of *nanaw* in an utterance is dependent upon our understanding of the real world. Example (58) is an instance in which our common sense of the objective world rejects any reading.

(58)? *'obay*   *tatini'*   **nanaw**
     PN   old   **NANAW**
     PN   lǎo   **NANAW**

The counterpart of this utterance *'a'alak nanaw* (young *nanaw*) can be interpreted in terms of LIMITATION as well as CONTINUATION sense, as we have shown in (39)'. On the other hand, *tatini' nanaw* (old *nanaw*) cannot receive a LIMITATION reading 'He is merely an old man.' Being old entails a high number of living years, which is conflictive with the prototypical LIMITATION meaning of *nanaw*; the LIMITATION reading of *tatini' nanaw* (old only) is thus ruled out. Interestingly, the informants also tend not to interpret (58) in CONTINUATION sense. One informant responded that it is weird for someone to "remain old" or to be "inherently old," which implies that the person was old since he was born. This interpretation is deemed unacceptable because it is conflictive to our real world experience. The CONTINUATION sense is only acceptable when more contextual information is added. For example, giving a temporal framing by telling the informants: "Someone saw 'Obay two years ago, and now he saw 'Obay again, finding 'Obay has not grown old since the last time they met." Without this additional contextual information, the informants often have difficulty comprehending this utterance.

4.5.2. Constructional patterns and polysemy

For linguistic contexts mentioned in 4.5.1, we did not talk about the "immediate linguistic environment." This is in fact a determining factor for interpretation of an utterance, which we considered very important, and is thus left to this section.

In diachronic studies of semantic change, Lehmann (1985) states that the shift of meaning is not a lexical issue, but involves the change of its co-occurring environment as a whole. This often involves a process known as "fixation" by which an item comes to occupy a fixed slot. Although the present study is not a diachronic one, this view has a different significance when applied in synchronic linguistic studies: regarding disambiguation, changes of meaning are manifested via changes of form, and constructional patterns often help us pinpoint an exact meaning among many possibilities. The effect is a very strong one.

A best example to illustrate this point is the (NEG_EXTREME) construction of *nanaw*. The (NEG_EXTREME) is substantive in the sense that the modified element is restricted to the negator *'oka'*.[8] Also, the (NEG_EXTREME) construction has a particularized pragmatic function that is not entirely predictable from its components. The speaker, via the use of this construction, expresses his evaluation or assumption on a specific topic, with a strong negative connotation.

Using a specific constructional pattern to express specific information has an advantage: more effective processing of an utterance. Goldberg (2006), an eminent specialist in construction-related phenomena, raises a crucial question in her latest book: **WHY constructional generalizations are learned?** By experiments, she found that constructional generalizations are learned because they are useful in communication,

---

[8] "Substantive" means that in a construction, most slots are fixed with specific lexical items or types of lexical items. Another relevant term "schematic" means just the opposite, i.e. with more open lexical slots (cf. Langacker 1987a; Croft and Cruse 2004).

especially for predication of the meaning of "an entire utterance." This view is attested in our study. Among the many possible meanings of *nanaw*, some of them only occur in specific structural patterns. As suggested by construction grammars, meanings reside partially in the form, and the basic unit of language should be pairs of form-meaning composites.[9] This also leads to the implication that when a construction becomes more idiomatic, its meaning will become more dependent on the structure, making it less predictable on the basis of the prototype. The (NEG_EXTREME) construction has a salient member *'oka' (ila) nanaw* that has the least open slots, and the interpretation has the least dependence on the semantic structure of the prototype, i.e. LIMITATION. In Figure 4.1, we draw a dashed line between LIMITATION and (NEG_EXTREME) to show their weak associations.
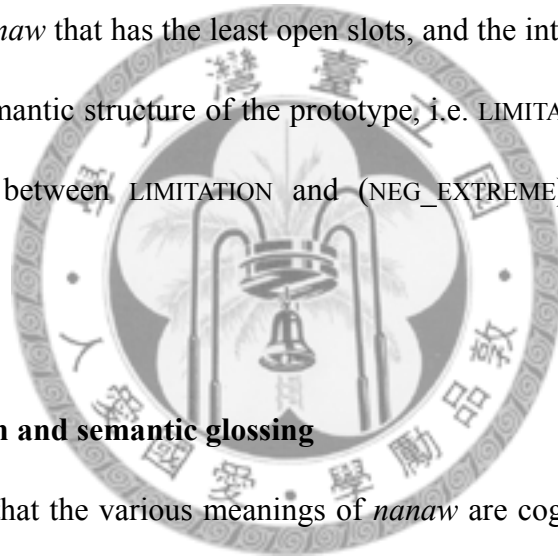
4.6. Direct translation and semantic glossing

We have shown that the various meanings of *nanaw* are cognitively related to each other when some of them have formed fixed structures with the collocating linguistic elements. Speakers rely greatly upon contexts to make precise interpretations, which naturally results into inconsistent direct translations of *nanaw* in different contexts. In this section, we further point out that *nanaw*'s semantic network is very unique, which results into a number of seemingly irrelevant Mandarin translations.

---

[9] We use "construction grammars" in a broad sense following a camp of theorists such as Charles Fillmore, George Lakoff, Ronald Langacker, Adele Goldberg, as some of the most representative ones. The capitalized term "Construction Grammar" is saved specifically to refer to Goldberg's framework.

4.6.1. Direct translations of a polyseme

In the course of linguistic investigation, we find that the informants use a wide variety of Mandarin translations to explain the meaning of *nanaw* in different contexts. Readers might have found some of these translations in the examples cited from NTU-Formosan. In Table 4.2, we briefly summarize the Mandarin translations in use. When we look into Mandarin dictionaries for the meaning of those translations, we find that they correspond to different senses/forces of *nanaw*.[10]

Table 4.2. Senses of *nanaw* and the dictionary meanings of its Mandarin translations

Translations Senses of *nanaw*		*zhǐ(yǒu)* 'only'	*réngrán* 'still'	*hái* 'still'	*yúanběn* 'originally'
LIMITATION		Yes			
CONTINUATION	[persistence]		Yes	Yes	
	[inherence]				Yes
(NEG_EXTREME)					

We can see from Table 4.2 that there is no Mandarin linguistic form that is semantically as well as syntactically comparable to the (NEG_EXTREME) sense of *nanaw*. In other words, (NEG_EXTREME) does not have a Mandarin proper "semanto-syntactic equivalent" (in the sense of Krzeszowski 1984). In this regard, the informants have much difficulty translating the instances of the (NEG_EXTREME) sense into Chinese. The informants often solve this problem by employing another Mandarin translation, e.g. *réngrán* 'still,' which is used as a translation also for [persistence]. The other solution is

---

[10] Please see Chapter 3 (§ 3.4.3) for the three Mandarin Chinese dictionaries chosen. Please see Appendix E for the detailed dictionary meanings of each of these Chinese translations.

also very common: Giving only pragmatic translation, and avoid translating the single lexical item *nanaw*. The second solution in fact has posed unwanted problems to our data analysis. In earlier stages of our investigation, because the informants often tell us *'oka' ila nanaw* means 'no method,' we used to misinterpret *nanaw* as meaning 'method.' Only at later stages when more data are collected and comparison is made possible can we be certain that it is in fact a fixed construction to denote hopelessness.

4.6.2. Idiosyncrasy of semantic extension

Table 4.2 also reveals another fact: In Mandarin dictionaries, these Mandarin glosses have little semantic overlap. Except *réngrán* 'still' and *hái* 'still' that are used interchangeably to express continuation of an activity, other translations have little functional overlap in Chinese. We argue that this is an effect of a low "degree of predictability of a cross-linguistic semantic hierarchy."[11]

In a cross-linguistic perspective, the semantic network of *nanaw* is a typologically-unimportant one manifested only in few languages. It is not easy to find any of the two senses of *nanaw* to share the same linguistic form in languages other than Saisiyat. Because the idiosyncratic semantic development is rarely found in other languages, direct translation often yields seemingly irrelevant meta-language translations. As we have shown, the Mandarin equivalents we obtained via the direct translation strategy are formally unrelated, and have clear functional division in Mandarin. The

---

[11] This term is borrowed from Sweetser (1986).

instances of *nanaw*, glossed by these seemingly irrelevant Mandarin translations (and English translations as well), are likely to be treated as homonymy.

4.7. Summary

In this chapter, we present the analysis of a Saisiyat polyseme *nanaw* whose various meanings are related to each other, and semanticized in specific contexts or grammatical structures. The semantic network of *nanaw* exhibits low degree of cross-linguistic predictability of a semantic hierarchy. In Mandarin (and in English as well), we do not find a single lexical item that exhibits similar semantic development as *nanaw*. Direct translation thus renders a number of seemingly irrelevant Mandarin translations. In this case, *nanaw* could easily be misconstrued as a homonymy because its multiple Mandarin translations have little functional overlap in Mandarin, and the relation between its meanings cannot be easily figured out. It is valuable to point out idiosyncratic semantic development of this kind, for it reveals a unique way to construe the relations in the physio-logical world.

Chapter 5   The Case Study of Saisiyat *Nahaen*

5.1. Preliminaries

In this chapter, we present a case study of Saisiyat *nahaen* which we often come across in language fieldwork. When the informants are asked to interpret its meaning, different Mandarin direct translations are employed according to the contexts it occurs in. The situation seems to suggest that the various meanings of *nahaen* are not related: their sharing the same linguistic form might be a sheer accident because what we have is a case of homophone. Nevertheless, some of its Mandarin translations exhibit functional overlaps, and we are thus interested to learn whether those meanings are related instantiations of a prototype interconnected by cognitive-pragmatic associations.

This chapter is organized as follows. In Section 5.2, we present a brief syntactic sketch of *nahaen*. In Section 5.3, we classify all instances of *nahaen* into senses following the tests of autonomy suggested in Chapter 3 (§ 3.4.1.). In Section 5.4, by identifying the relations between its senses, we show that the senses of *nahaen* are mutually related to each other as members of the same family. Some of the meanings are pragmatic forces that are specified in the immediate contexts. It is difficult to extract the semantic elements of a linguistic form and totally exclude pragmatic considerations. Relevant discussions will be presented in Section 5.5. Finally, in Section 5.6, we turn back to look at the Chinese glosses of *nahaen* in our language fieldwork. Special focus will be placed on cross-linguistic mismatches of semantic partitioning and its impact on

the direct translation approach to language investigation.

5.2. A sketch of Saisiyat *nahaen*

In NTU-Formosan, the lexical item *nahaen* in fact has three transcriptions: *nahan*, *nahaen*, and *naehaen*. The transcription of a linguistic form may not be consistent in NTU-Formosan, because phonetic variations are common in spoken data due to modulation of nearby phonological environments. When carefully pronounced, it is /nahaen/. In our discussion, this lexical item is referred to as *nahaen*, but if the examples are extracted from NTU-Formosan, its various spellings are retained.

Nahaen typically occurs right after a verb phrase, as shown in (59a). It does not occur in any position preceding a modified verb, as shown in (59b-d).

(59)a.
yao	*hayza'*	*ila*	*min-osa'*	*raroemoe'an*
1SG.NOM	EXIST	PFV	AF-go	Xiangtian.lake
1SG. NOM	EXIST	PFV	AF-qù	Xiangtianhú

ko'hael	*'am*	*rima'*	***nahaen***
next.year	FUT	AF-go	**NAHAEN**
míngnían	FUT	AF-qù	**NAHAEN**

'I have been to Xiangtian Lake before. I am going there again next year.'
'Wǒ yǐqían qù-gùo Xìangtianhú, míngnían hái yào qù.'

b.*
nahaen	*ko'hael*	*'am*	*rima'*
NAHAEN	next.year	FUT	AF-go
NAHAEN	míngnían	FUT	AF-qù

c.*
ko'hael	***nahaen***	*'am*	*rima*
next.year	**NAHAEN**	FUT	AF-go
míngnían	**NAHAEN**	FUT	AF-qù

d.*
ko'hael	*'am*	***nahaen***	*rima'*
next.year	FUT	**NAHAEN**	AF-go
míngnían	FUT	**NAHAEN**	AF-qù

Saisiyat verbs have three kinds of focus marking, and n*ahaen* can modify all three of them: agent-focus (AF), patient-focus (PF), and referential focus (RF). Examples are presented (60a-c). As shown in these examples, *nahaen* strictly stays in the postverbal position regardless of the focus types.

(60)a *sia*   *rima'*   *'<oem>alep* **nahaen**
      3SG.NOM   AF-go   hunt<AF>   **NAHAEN**
      3SG. NOM   AF-qù   dǎlìe<AF>   **NAHAEN**
      'He went hunting again.'
      'Ta yòu qù dǎlìe lė.'

   b. *'ahoe'*   *Sebet-en* **nahaen** *noka*   *'obay*
      dog   hit-PF   **NAHAEN** GEN   PN
      gǒu   dǎ- PF   **NAHAEN** GEN   PN
      ''Obay hit the dog again.'
      ''Obay yòu dǎ gǒu lė.'

   c. *korkoring* *si-saehae'*   *nahaen*
      child   RF-fall   **NAHAEN**
      xǐaohái   RF-diaolùo **NAHAEN**
      'The child was seen falling down again (and the speaker is affected).'
      'Yòu kànjìan xǐaohái dìao xìaqù.'

If a verb phrase contains more than one element, *nahaen* can occur in the midst of the verb phrase. In (61), it precedes a tense aspect marker. In (62), it occurs between the verb and the direct object. Also, in a serial verb structure, *nahaen* can be inserted between two verbal elements, as shown in (63).

(61)a. *'obay   ma'erem   ila   **nahaen***
    PN   AF.sleep   PFV   **NAHAEN**
    PN   AF.shùi   PFV   **NAHAEN**
    ''Obay slept again.'
    ''Obay yòu shùi lè.'

  b. *'obay   ma'erem   **nahaen**   ila*
    PN   AF.sleep   **NAHAEN**   PFV
    PN   AF.shùi   **NAHAEN**   PFV
    ''Obay slept again.'
    ''Obay yòu shùi lè.'

(62)a. *yao   s\<om\>i'ael   ka   'aelaw   **nahaen***
    1SG.NOM   eat\<AF\>   ACC   fish   **NAHAEN**
    1SG. NOM   chi\<AF\>   ACC   yú   **NAHAEN**
    Reading 1: 'I am still eating the fish.'
           'Wǒ hái zài chi yú.'
    Reading 2: I ate fish again.
           'Wǒ yòu chi yú lè.'

  b. *yao   s\<om\>i'ael   **nahaen**   ka   'aelaw*
    1SG.NOM   eat\<AF\>   **NAHAEN**   ACC   fish
    1SG. NOM   chi\<AF\>   **NAHAEN**   ACC   yú
    Reading 1: 'I am still eating the fish.'
           'Wǒ hái zài chi yú.'
    Reading 2: I ate fish again.
           'Wǒ yòu chi yú lè.'

(63)a. *'obay   rima'   ila   '\<oem\>alep   **nahaen***
    PN   AF-go   PFV   hunt\<AF\>   **NAHAEN**
    PN   AF-qù   PFV   dǎlìe\<AF\>   **NAHAEN**
    ''Obay went hunting again.'
    ''Obay yòu qù dǎlìe lè.'

  b. *'obay   rima'   ila   **nahaen**   '\<oem\>alep*
    PN   AF-go   PFV   **NAHAEN**   hunt\<AF\>
    PN   AF-qù   PFV   **NAHAEN**   dǎlìe\<AF\>
    ''Obay went hunting again.'
    ''Obay yòu qù dǎlìe lè.'

As we have mentioned in Section 4.2.2, many adverbials in Formosan languages exhibit properties of verbal predicates. These verb-like adverbials can have focus markers, and they can also stand alone as a clause. Saisiyat *nahaen*, nevertheless, does not have verbal properties. *Nahaen* cannot be marked by focus markers (AF, PF, and RF), and it cannot

stand alone in a clause without another verb. To make a contrast, we cited an example of typical verbal predicate from Chapter 4, reprinted in (64). In the answer of speaker B, the verb-like adverbial *maybalbalay* can make a clause. On the contrary, *nahaen* alone cannot make a clause, as shown in (65). The answer of speaker B in (65) is not acceptable because *nahaen* has to occur together with the modified verb to constitute a clause.

(64) *A:* *sia*   **maybalbalay**   *ay*   *manraan?*
          3SG.NOM   **AF.slow**   Q   AF.walk
          3SG. NOM   **AF.màn**   Q   AF.zǒu
      B: *'ihi'.*   **maybalblay.**
          BC   **AF.slow**
          BC   **AF.màn**
      A: Did he walk slowly?
      B: Yes. Slowly.
      A: Ta mànmàn zǒu gùolái ma?
      B: Dùi. Mànmàn de.

(65) *A:* *Obay*   *S<om>bet*   *ila*   **nahaen**   *ay?*
          PN   hit<AF>   PFV   **NAHAEN**   Q
          PN   dǎ<AF>   PFV   **NAHAEN**   Q
    B: *'ihi'.*   **nahaen.**
          BC   **NAHAEN**
          BC   **NAHAEN**
      A: Did he hit (someone/something) again?
    *B: Yes. Again.
      A: Ta yòu dǎ le ma?
    *B: Dùi. Yòu dǎ le.

Saisiyat has many verb-like adverbials that have focus marking and they can make a clause alone, but *nahaen* appears not to be one of them. It is a typical adverbial that has to be used together with a verb that it modifies.

5.3. Senses of *nahaen*

This section investigates all the functions of *nahaen* that we come across in NTU-Formosan and in our direct elicitation. According to the sense judgment criteria proposed in Chapter 3 (§ 3.4.1), the meanings of *nahaen* can be classified into the followings:

REPETITION
    [recurrence]
    [continuation]
    [addition]
SUCCESSION
PRECEDENCE
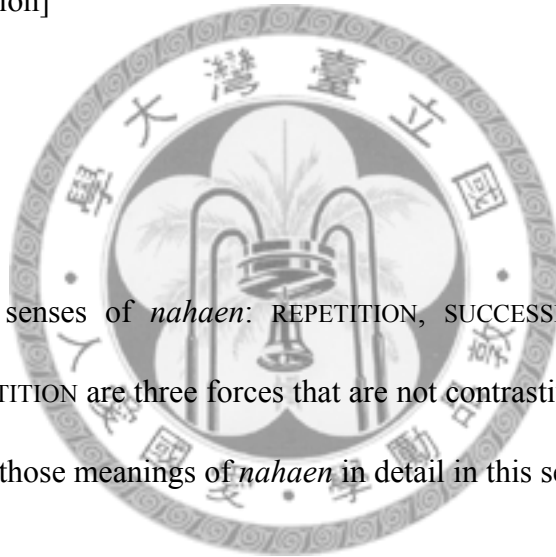
There are three senses of *nahaen*: REPETITION, SUCCESSION, and PRECEDENCE. Subsumed under REPETITION are three forces that are not contrastive or antagonist to each other. We will explain those meanings of *nahaen* in detail in this section.

5.3.1. REPETITION

The REPETITION sense encompasses instances that denote this meaning: "**An activity previously occurring takes place for the second time.**" It can have three forces as subtle contextual modulations: [recurrence], [continuation], and [addition]. They are explicated below.

The [recurrence] force denotes an activity that happens for the second time. A recurring event involves an action (A) and its reduplicate (A') in chronological order, and *nahaen* signals an attentional focus on the recurring A'. Example (66) is taken from a Saisiyat legend. Saisiyat people believe their ancestors, when very old, would molt and become young again. The word *nahaen* in IU 13 receives the [recurrence] interpretation because there is a felt discontinuous phase between the men's twice of being youth – one before and one after molting.

(66) molaw

24.	…	*hiza*	*tatini'*		*ila*	
		FIL	grow.old		PFV	
		nàgè	lǎo		PFV	
25.	..	*m-olaw*		*ka*	*bangeS*	
		AF-molt		ACC	skin	
		AF-tùepí		ACC	pífu	
26	..	*minayhael*				
		next.generation				
		zǐsun				
27.	…	*'am=*				
		FUT				
		FUT				
28.	…(1.0)	*'am*	*'al'alak*	**naehaen**	.\	
		FUT	young	**again**		
		FUT	níanqing	**yòu**		
29.	…	*hoe=pay*				
		AF.be.tired				
		AF.lèi				
30.	…	*o:*	*biSbiS*	*'atomalan*		
		INT	AF.painful	very		
		INT	AF.tòngkǔ	feicháng		
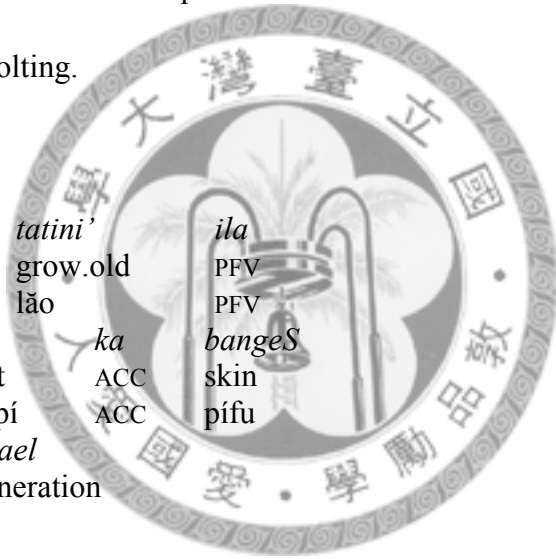
31. ...(0.8) *hinghae'   ki   siba:i'   hara   haysani   siba:i'*
                same   and   snake   like   now   snake
                xiangtóng yǔ   shé   xiàng xiànzài   shé
     'They would grow old and their skin would molt; (it was like that) for generations.
     It was very tiring and painful, just like a snake molting.'
     'Zhèyàng yidài yòu yidài. Tamėn lao lė tùipí yòu biàn níanqing lė. Feicháng dė lèi
      yòu feicháng dė tòngkŭ. Jìu gen shé yiyàng. Xìang xìanzài dė shé yiyàng.

Example (67) is also an instance of [recurrence]. The use of *nahaen* denotes the state of an entity being reverted to its previous state. IU 9-14 describes how the bamboo shoots are dehydrated for preservation. In IU16, *nahaen* is employed to indicate that the dried bamboo shoots are recovered to its previous "moist" state. Similar event construal has been called "restitutive" in studies of Mandarin repetitive adverbs *zài* and *yòu* (Han 2004).
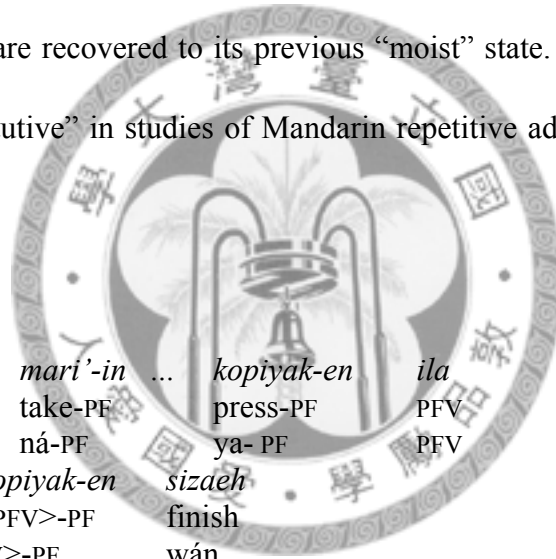
(67) anhi2
9. ...(1.3) *na==   mari'-in   ...   kopiyak-en   ila*
              FIL   take-PF   press-PF   PFV
              FIL   ná-PF   ya- PF   PFV
10. ...(1.4) *k<in>opiyak-en   sizaeh*
              press<PFV>-PF   finish
              ya<PFV>-PF   wán
11. ...(1.1) *in-timo'-en*
              PFV-salt-PF
              PFV -yùngyényen- PF
12. ...(1.3) *isa==*
              DM
              DM
13. ...(1.5) *mari'-in*
              take-PF
              ná- PF
14. ...(0.9) *tabe-en   ila   ray==   ...   'a   taboway*
              put-PF   PFV   LOC   FIL   jar
              fàng- PF   AF   LOC   FIL   wèng
     "The bamboo is pressed, saltened, and put in a jar."

15.	...(1.2)	*So:*	*'a-s<m>i'ael*	*ila*					
		COND	FUT-eat<AF>	PFV					
		COND	FUT-chi<AF>	PFV					
16.	...(1.2)	*senge-en*	**naehaen**	*ray*	*ralom*				
		soak-PF	**still**	LOC	water				
		pào-PF	**réngrán**	LOC	shǔi				
17.	...	*a==*	*talek-en*	*ma'...*	*nak*	*isaa*	*ma*	*kin*	*kayzaeh*	*si'ael-en*
		FIL	heat-PF	DM	like	that	DM	very	good	eat-PF
		FIL	jiarè-PF	DM	xìang	nà	DM	hěn	hǎo	chi-PF

'If (one) feels like eating (it), it would taste better if heated.'

'Rúgǔo xǐang chi, xian zài pào gùo shǔi, zhǔ gùo dė wèidào háishì nàmė dė xinxian hǎochi.'

The use of *nahaen* can be generalized to situations wherein A' is not a "strict reduplicate" of A. Example (68) illustrates such a case. In this example, the speaker states that a man carrying a goat passes by a tree, and soon after that, another boy riding a bicycle also passes by the same tree.

(68) pear 4

27.	...(0.8)	*o:*	*rima'*	*ila*	*hiza*
		DM	AF-go	PFV	there
		DM	AF-qù	PFV	nàlĭ
28.		*kita'-en*	*m-wa:i'*	*ila*	**naehaen**
		see-PF	AF-come	PFV	**still**
		kàn-PF	AF-lái	PFV	**hái**
29.		*'aehae'*	*ka==*		
		one	NOM		
		yi	NOM		
30.	...	*kamo'alay*			
		young.man			
		níanqingrén			
31.		*kama-manra:an*			
		man			
		nánrén			
32.	...	*'ima*	*papama'*	*ray==*	
		PROG	AF.ride	LOC	
		PROG	AF.qí	LOC	

33. ...(1.0) *kapapama'an==*
             vehicle
             jǐaotàche
     '(Off) they went. (Then I) see another boy coming; he was riding a bike.'
     'Tamėn zǒu lė. Yòu kàndào lìngyigė nánháizi lái lė. Ta qízhė jǐaotàche.'

The use of *nahaen* in IU 28 indicates the same activity "passing by the tree" happens for the second time. In this instance, A' is not a strict reduplicate of A, since the actors are different (a man versus a child), and their manners are different (walking versus riding).

The [recurrence] force of *nahaen* in fact encompasses events with different identity relation between A and A'. The informants agree that *nahaen* typically introduces recurrence that is exactly identical to its preceding activity, i.e. an A-A' structure. However, in examples such as (67) and (68), the following activity is not strictly identical to the previous activity, constituting a loose A-B structure.
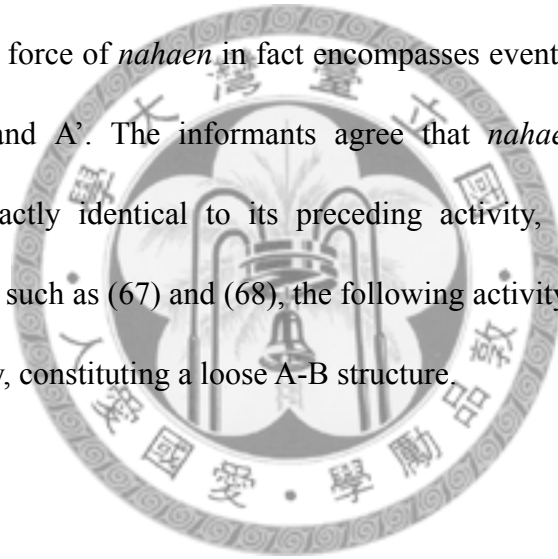
[Continuation]

The [continuation] force indicates that an activity lasts for a period of time without an interruption. Like [recurrence], it involves two stimuli A and A', but [continuation] overlooks the temporal gap between A and A', rendering them a continuous activity. Example (69) is extracted from a fragment of Pear story. In this story, an old man picked several baskets of fruits, and one of them was stolen by a boy. During the time the fruits were stolen, the old man remained on top of the pear tree. In IU 44, the use of *nahaen*

109

indicates that the action of "picking fruits on top of the tree" lasts for a period of time as conceived by the speaker.

(69) pear 3
44.   ...(1.2)   *isaza   tatini'   rima'   r<om>okrok*   **naehaen**
     FIL   old.man   AF.go   pick<AF>   **still**
     FIL   lǎorén   AF.zǔo   zhai<AF>   **hái**
     babaw   *ka*   *boway*
     above   ACC   fruit
     shàngmìan   ACC   shǔigǔo
  'The old man was still up in the tree picking fruits.'
  'Lǎorén hái zài shùshàng zhai shǔigǔo.'

The [continuation] force is also found in disjunctive utterances, and it denotes that an activity or a state persists regardless of a non-favorable condition, as shown in (70) below. Based on (70), we claim that the [continuation] reading of *nahaen* pertains to not only temporal domain but also to atemporal domain which is used for abstract disjunctive reasoning.

(70) *'ima   hasa'   h<om>ayap   kabkabahae'   koSa'-en   kabkabahae'*   **nahaen**
  'IMA   unable   fly<AF>   bird   say-PF   bird   **NAHAEN**
  'IMA   búhùi   fei<AF>   nǐao   shuo-PF   nǐao   **NAHAEN**
  'Búhùi fei dė nǐao háishì jìaozùo nǐao.'
  'Birds that cannot fly are still called birds.'

The relation between recurrence and continuation of an activity is widely discussed in previous literature (Dowty 1979; Yeh 1998; Han 2004). For example, Chinese *hai*, now typically meaning 'still,' is historically derived from the sense of 'return' or 'restitute.' In English, an utterance like *He ran, and ran, and ran* denotes repetition of an activity as

well as continuation. The different between repetition and continuation lies in the speaker's subjective interpretation on "what happens in the midst of A and A'?" Koffka (1935, cited in Ungerer and Schimid 1996) postulates a so-called "gestalt principle of continuation" by which human beings tend to perceive elements as a whole if they only have few interruptions. Observe the following figure, and we can find that a smeared alphabet can be read "C" or "O." Since we did not see what is under the smear, we may neglect the part that we did not see, and make it a "C." Or else we mentally visualize a continuous line and make it an "O."

Figure 5.1. A smeared alphabet
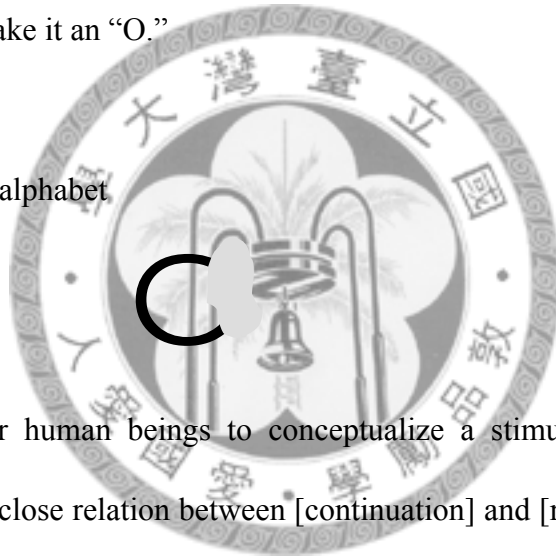
The tendency for human beings to conceptualize a stimulus as continuous and complete explains the close relation between [continuation] and [recurrence]. Most of the time it is the contexts that coerces a specific reading. At the perception stage, the speaker perceives A and A' at two different times on the temporal scale, with a gap between them. An objective conception is that A' is the reduplicate of A, and a recurrence reading thus overrides. Nevertheless, in some conditions, the construal of A and A' fulfils what has been called the "Gestalt principle of good continuation," as the following:

  a. A' is exactly the same as A.
  b. A has a property that can persists for a period of time.

Under the conditions, the speaker can easily fill in the gap with imaginary persistence of the activity. The combination of two segments A-A' is construed as one meaningful unit A. This also involves the speaker's adjustment of his scope of viewing. As has been pointed out in Dowty (1979), continuous reading pertains to a wider event-internal view, and gaps are likely to be deemphasized in order to highlight the integrity of the event segments. On the other hand, when a narrower event-external view is taken, the repetition meaning will override. Most of the time, it is the contexts that help us to determine which perspective to be taken.

[Addition]

The [addition] force is coerced by *nahaen*'s collocation with numeral predicates. Instances of [addition] are found in contexts when the speaker is making a list, and something occurs to him as an additional member to this list, usually as an afterthought. Example (71) is extracted from an episode in which the speaker talks about the breeds of deer he keeps. After reciting a list of breeds, he recalls another kind of similar deer but he forgets its breed name. In IU24, a numeric verb *'aehae'* 'one' precedes *nahaen*, which coerces the additive reading of *nahaen* and the informants intuitively translate it as *lìngwài* 'another.'

112

(71) election
18.	M:	...	*piza'*	*sinraehoe'*	*hiza*	*koSa'en*	*ka*	*wa'ae'*
			how.many	species	there	FIL	NOM	deer
			duōshǎo	zhǒnglèi	nàlĭ	FIL	NOM	lù

'How many species are there?'

'Nà lù zǒnggòng yǒu jǐ zhǒng ah?'

19.	B:	*...(1.9)*	*to:o'*	*sinraehoe'*
			three	species
			san	zhǒnglèi

20.	M:	...	*to:o'*	*sinraehoe'*
			three	species
			san	zhǒnglèi

21.	M:	*...(1.1)*	*'aehae'*
			one
			yi

22.	M:	...	*koSa'en*	..	*kasakiray*	*bangol*	*ka==*	*wa'ae'*
			FIL		wild	wild	NOM	deer
			FIL		yě	yuánshĭsenlín	NOM	lù

'Three. One is wild.'

'Oh, san gè. San gè zhǒnglèi. Yìzhǒng shì yělù.'

23.	B:	...	*e*
			BC
			BC

24.	M:	*...(0.8)*	*'aehae'*	**naehaen**
			one	**another**
			yi	**lìngwài**

25.	M:	*...(0.9)*	*sinkano'on*
			named.what
			jìaoshémè

'There is one more wachamaccalit.'

'Lìngyìzhǒng shì jìaozùo shémè lái dè.'

Semantically, to add a new member to a category is construable as repeating the act of enumerating. And grammatically, as we have previously shown in Section 4.3.1, Saisiyat numerals exhibit verbal properties: they can have focus markings and tense/aspect markers. We thus subcategorize the [addition] force under the REPETITION entry to show that the addition reading is a contextual elaboration of *nahaen*.

113

5.3.2. SUCCESSION

The SUCCESSION sense encompasses instances which depicts that "**in an event that contains more than one activity, those activities follow a temporal sequence.**" It is classified as a distinct sense because it contrasts with REPETITION in terms of the truth-condition autonomy. For example, (72) is acceptable even when the subject has not planted peanuts before.

(72) | *k<em>e:b* | *ka* | *hinpetoel* | *o* |
cut\<AF\>	ACC	weed	LIG
ge\<AF\>	ACC	cǎo	LIG
mamoa'	*ila*	***nahaen***	*ka*	*tawtaw*
AF.plant	PFV	**NAHAEN**	ACC	peanut
AF.zhòng	PFV	**NAHAEN**	ACC	huasheng

'(Someone) weeded the field, and then planted peanuts.'
'Xian gecǎo, zài zhòng huasheng.'

In narration, SUCCESSION is sometimes used as a strategy to manage the plots, arranging the storyline into coherent chronological order. Example (73) illustrates an instance of *nahaen* used for construction of action succession. It is extracted from a fragment of frog story in which a boy and his dog keep looking for a frog. The sequential activities, "to go aside", "to rest on a piece of wood" and "to catch the frog" are linked by *nahaen* to show that they are temporally relevant to the episode "to search for a specific object." Three instances of *nahaen* in IU 259, 262, and 267, are translated into Mandarin as *xian* 'first.' It is very likely that the informants have equated it with Mandarin *xian…zai…* ('first…then…') construction.

114

(73) frog5

258. ...(1.4) *m-wa:i'*
        AF-come
        AF-lái

259. ... *kabih* **naehaen** *langi==*
        side **first** side
        pángbian **xian** pángbian
    'They came first to the side.'
    'Tamėn xian dào zhèbian.'

260. ...(0.8) *'akoy* *'atomalan* *ila* *kita-en* *ka* .. *a* *takem*
        many very PFV see-PF NOM FIL frog
        duo hěn PFV kàn- PF NOM FIL qingwa
    'They saw many frogs.'
    'Tamėn kàndào xŭduo qingwa.'

261. ... *lasia* *isaa*
        3PL.NOM there
        3PL.NOM nàlĭ

262. ...(0.8) *masa:eng* *ri'saon* **naehaen** *ray*
        AF.sit there **first** LOC
        AF.zùo nàlĭ **xian** LOC

263. ...(0.8) *'ataba:i* *babaw*
        wood top
        mùtóu shàngmìan
    'They were sitting on top of the dead piece of wood.'
    'Tamėn zùozài nà kuài mùtóu shàngmìan.'

264. ...(3.2) *isahini* *ka==*
        now NOM
        xìanzài NOM

265. ... *a* *korkoring*
        FIL child
        FIL xĭaohái

266. ...(1.3) *ma==*
        DM
        DM

267. ... *a* *r<om>akep* ... *tasihoeroe*[1] *ila* **naehaen**
        FIL catch<AF> ?? PFV **first**
        FIL zhuo<AF> ?? PFV **xian**

268. ... *kayzaeh* *ka==* ... *a* *takem*
        good NOM FIL frog
        hăo NOM FIL qingwa
    'Now the child caught... , found his frog.'
    'Xìanzài xĭaohái zhuo ... zhăodào tadė qingwa.'

---

[1] It might be a variant of *tihoeroe* 'find.'

There is a loose schematic construction "VERB *nahaen* VERB *nahaen…*" that connects activities A, B, C, and so forth. In direct elicitation, we can see this structure more clearly, as shown in (74).

(74) *yao*   *minSala'*   [*baiw*   *ka*   *taumo'*   **(nahaen)**]₁

      1SG.NOM   AF.first   AF.buy   ACC   banana   **NAHAEN**

      1SG. NOM AF.first   AF.mǎi   ACC   xiangjiao **NAHAEN**

                          [*baiw*   *ila*   *ka*   *lapuwar*   ***nahaen***   ]₂

                          AF.buy   PFV ACC   guava   **NAHAEN**

                          AF.mǎi   PFV ACC   balè   **NAHAEN**

                          [*baiw*   *ila*   *ka*   *zozo'*   ***nahaen***   ]₃

                          AF.buy   PFV ACC   plum   **NAHAEN**

                          AF.mǎi   PFV ACC   lǐzi   **NAHAEN**

'I first bought bananas, and then bought guavas, and then bought plums.'

This schematic structure has an additional interactional function: In narration, the lexical item *nahaen* in this construction prepares the hearer to expect a series of activities temporally ordered, and the sum of them makes a holistic "episode."[2]

5.3.3. PRECEDENCE

The PRECEDENCE sense depicts that **"an activity is carried out on a temporal basis, and there might be subsequent activities yet to do."** It is classified as an autonomous sense because it shows truth-condition autonomy. For example, (75) can have two meanings in different contexts. The two readings are conflictive in some contexts. The speaker producing this utterance in its PRECEDENCE sense might not have rested before.

---

[2] The term "episode" is defined as a discourse unit that can be justified by episode closure and thematic unity, with "paragraph" as its written counterpart (See Su 1998).

(75)   *yao*   ***iya***   *s<in>angay*   ***nahaen***
       1SG.NOM   want   rest<IN>[3]   **NAHAEN**
       1SG.NOM   yào   xiuxí<IN>   **NAHAEN**
       Reading 1: 'I want to rest a bit.' (PRECEDENCE)
       Reading 2: 'I want to rest again.' (REPETITION)

And PRECEDENCE is also in conflict with SUCCESSION in terms of exclusiveness (cf. § 3.4.1). Precedence indicates that someone does an activity on a temporal basis to prepare for another coming event. SUCCESSION on the other hand depicts that someone does an activity that is subsequent to a previously-mentioned one. PRECEDENCE projects subsequent activities in the future whereas SUCCESSION projects preceding activities in the past, and the two readings cannot co-exist in one situation.

Example (76) is extracted from a conversation between two elder informants talking about Saisiyat tomb sweeping ceremony. The young people packed up the foods (sacrificial offerings) before the elders had a chance to enjoy them, and they were told not to do that as yet. In IU 225, the use of *nahaen* indicates that the young men should "halt their act for a while" and wait for the elder men.

---

[3] In utterances containing *iya* 'want,' the verb in the complement clause carries an infix *<in>*. This is very likely to a polysemy of the perfective aspect marker, but to be conservative, we currently do not gloss it as PFV.

(76) holiday

214. C: (0) *uh=*
          BC
          BC

215. K: (0) *isaa*   *a-s<m>i'ael*   *ila*   *o:*
        there   FUT-eat<AF>   PFV   DM
        nàlǐ   FUT-chi<AF>   PFV   DM

216. K: ... *kita-en 'oka' ila*
        see-PF   NEG   PFV
        kàn-PF   NEG   PFV

    '(After the meeting was over and) they were ready to eat, all the food was gone.'
    '(Děngdào jùhùi yǐjing jíeshù, lǎorén) tamèn zhǔnbèi yào chi, sǔoyǒu chi dè dongxi dou méi lè.'

 … (5 IUs omitted)

222. K: ... *isahini*   *kaysa'an*   *'am*   *manabih*   *lasia*
        now   today   FUT   AF.say   3PL.NOM
        nà   jinnían   FUT   AF.shuo   3PL.NOM
    'They were already told (...) this year.'
    'Nà jinnían tamèn yǐjing bèi shuo lè.'

223. K: ... *'am*   *kayzaeh*   *ila*   *<L2kongkoL2>*   *k<om>oSa:*
        FUT   AF.good   PFV   [Hak]announce   say<AF>
        FUT   AF.hǎo   PFV   [Hak]xuanbù   shuo<AF>
    'The situation has improved.'
    'Qíngkùang yǐjing gǎishàn lè.'

224. C: (0) *uh*
          BC
          BC

225. K: .. *'izi'*   *['izi' potoy*   **naehaen]**
        NEG   AF.pack   **first**
        NEG   AF.dǎbao   **shǒuxian**
    'Don't pack (before everybody has eaten).'
    '(Dàjia háiméi chi zhiqían) xian bù dǎbao.'

The activity marked by *nahaen* is a "prelude" to something more important yet to come. In this case, the subsequent activity is "the elder people enjoying the food." Saisiyat has another adverbial *minsa'la'* 'first' that denotes an activity that precedes other

activities, as in (77a), or someone gets ahead of others, as in (77b).

(77)a. *yao*   **minsa'la'**   *rima'*   *raroemoe'an*   *lobih*   *may*   *walo'*   *ila*

1SG.NOM **first**   AF.go   PN   AF.return pass PN   PFV

1SG.NOM **xian**   AF.qù PN   AF.húi   jingguò PN   PFV

'I went to Raroemoe'an first, and pass Walo' when I returned.'

'Wǒ xian qù Xìangtianhú. Húilái shí jingguò Donghé.'

b. *sia*   **minsa'la'**   *ma'erem*   *ila*

3SG.NOM **first**   AF.sleep   PFV

3SG.NOM **xian**   AF.shùi   PFV

'He went to sleep ealier than others did.'

'Ta xian qù shúi lè.

According to the informants, *nahaen* of its PRECEDENCE sense differs from (77a) in that it is frequently used to mitigate speaker's suggestion or invitation, for the purpose of politeness. Sometimes, there is in fact no subsequent activity after the prelude. The speaker simply exploits the structure of PRECEDENCE to presuppose that the addressee might have something important to do, and thus to moderate a face-threatening act (in the sense of Brown and Levinson 1987). As we have seen in (76), *nahaen* in its PRECEDENCE sense is used to make a suggestion, and the following example (78) is taken from our fieldnotes. It shows how *nahaen* is used to express the speaker's invitation in a polite manner.

(78) *si'ael*   **nahaen**

eat   **NAHAEN**

chi   **NAHAEN**

'Come have a bite (before you leave) !'

'Chi gè dongxi (zài zǒu) ba!'

This utterance is used when the speaker invites or suggests the hearer to do

something. To show politeness, the speaker will assume that the hearer might be busy doing something, and his invitation or suggestion is downgraded as a prelude.

5.4. The semantic network of *nahaen*

After identifying the senses of *nahaen*, we attempt to learn whether they are accidental sharing of the same form or "principled" extension motivated by cognitive or pragmatic mechanisms. This section explicates the relations between the senses of *nahaen*, by identifying the prototype and finding the cognitive/pragmatic associations between the senses. The result shows that those senses might be members of the same "family," and *nahaen* might not be as polysemous as hinted by its multiple meta-language translations.
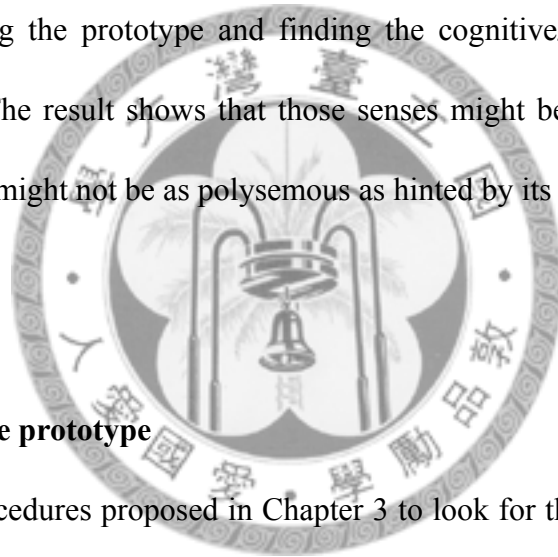
5.4.1. Establishing the prototype

We follow the procedures proposed in Chapter 3 to look for the prototypical sense of *nahaen* by **salience to native speakers**, and supplemented by two other optional criteria: **naturalness of grammatical predication** and **social-phenomenological basicness**. According to our experience, the saliency test alone usually successfully yields a consistent answer.

We first asked the informants "What is the meaning of *nahaen*" in de-contextualized situations. They intuitively give the counterparts *hái* 'still,' *réngrán* 'still,' *zài* 'again,' and *yòu* 'again' almost with equal salience. We further asked the informants to make an

utterance containing *nahaen*, and they are prone to make instances of [recurrence] (of the REPETITION sense).[4]

We use two strategies to find the sense that is salient to native speakers, but in the case of *nahaen*, we find that the strategy of asking the informants to make utterances containing a specific lexical item is more objective and reliable. It involves less interference of meta-language. On the contrary, by asking the informants "What is the meaning of X," we often obtain meta-language translations which are themselves ambiguous. For example, Mandarin *hái* 'still,' although typically denoting continuation of an activity, is sometimes used to express repetition/iteration/concession of an activity. For example, *hái* can denote a recurrence that is countering an expectation, as shown in (79).

(79) Mandarin *hai* used to express recurrence of an activity

jìao	*ta*	*búyào*	*lái,*	*ta*	**_hái_**	*lái*
call	3SG	NEG	come	3SG	**still**	come

'(I) have told him not to come, yet he came again.'

Mandarin *hái* being used to denote multiple different meanings can be found in Appendix E. In fact, *zài* 'again' and *yòu* 'again' also have many meanings in addition to 'recurrence of an activity.' For example, both of them can denote 'succession of activity.' It is arguable what the informants actually intend to tell us by using these meta-language

---

[4] At earlier stages of investigation, we suspect [continuation] as the prototypical meaning of *nahaen*, because a large amount of tokens receive Mandarin translations *hái* 'still' or *réngrán* 'still' which are used primarily to denote continuation of an activity. However, the [recurrence] force as the prototypical meaning of *nahaen* is attested via the utterance-making-up task, and the predominance of [recurrence] is also manifested in the fact that other senses of *nahaen* denote sequencing of two or more activities which are not identical (i.e. A-A' sequence instead of A-A continuity).

lexemes.

Although the network of *nahaen* encompasses many different situations of A-A' sequence: strict repetition, loose repetition, continuous, disjunctive, chronological, etc. One specific type receives more cognitive saliency than other subtypes, i.e. the one with the reading of strict repetition, such as (66). A tendency of "low-level prototype" is again attested here (cf. also § 4.4.1). The most salient meaning of *nahaen* to the informants is "smaller" than "sense," usually at the level where a specific subtype of word classes can be determined and semantic features can be stated.

5.4.2. Networking the senses

When classifying the instances of *nahaen* into senses, we learn that many of its meanings are cognitively contrastive, but they also have some relations that can explain their sharing of the same form. In this section, we will explicate how they are chained to each other according to the types of construal change (Croft and Cruse 2004) that we have reviewed in Chapter 3. We construct the semantic network of *nahaen* as shown in Figure 5.2. Senses of *nahaen* are related directly or indirectly to the prototype, constituting a network that exhibits "chaining" (Austin 1961) or "family resemblance" (Wittgenstein 1963) of meanings.

Figure 5.2. The semantic network of *nahaen*

REPETITION {
[continuation]
[recurrence] —*attention*→ SUCCESSION —*attention*→ PRECEDENCE
[addition]
}
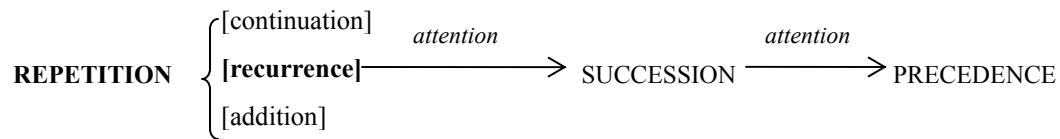
The interconnection of the senses suggests that *nahaen* is a polysemy, and the sharing of the same form is not an accident. In the following, we will explicate the cognitive and pragmatic mechanisms that motivated the links between the prototype and the other senses.

From REPETITION to SUCCESSION

The SUCCESSION sense is an extension from the [recurrence] force. As we have mentioned, there are some instances of [recurrence], such as (68), in which the subsequent activity A' is not a strict reduplication of the preceding A. The loss of identity relation between A and A' can be a perfect ground for SUCCESSION reading to take place. What makes A and A' (or A'' and so on) a "cluster" that can be marked by *nahaen* is the fact that their combination makes an "episode" in which activities are causally or chronologically related. In this respect, the speaker is neglecting the "identity relation between A an A', and the meaning is shifting from specific resembling relation to a looser temporally-ordered relation. The shift involves the speaker's extension of viewing scope, from an activity-centered narrow scope to an episode-centered wide scope. In this respect, **attention (salience)** plays a crucial role. According to Croft and Cruse (2004), change of

attention can have the following situations: 1) select some relevant elements and ignore the others, 2) broaden or narrow the scope of attention in appropriate contexts, 3) adjust the granularity to specific information or to general ones, and 4) to move one's attention across the time to create stative-dynamic contrasts. The second type of attentional shifts accounts for the relation between [recurrence] and SUCCESSION.

From SUCCESSION to PRECEDENCE

When the polysemy network of a lexical item is established, it would become the solution to the future communicative problems, and serve as the foundation of further semantic extensions (Campbell 1998). As the link between REPETITION and SUCCESSION is established, SUCESSION can instantiate further extensions through constant usage, and we find PRECEDENCE has strong functional connections with SUCCESSION, motivated by **attention (salience).**

PRECEDENCE is similar to SUCCESSION in many aspects. First, they both depict a series of activities, A, A', A" and so on. Also, A' and A" are not reduplicates of A. In addition, the series of activities are ordered only by temporal sequence. Nevertheless, PRECEDENCE is different from SUCCESSION in that its focus is placed on A, which is the sole activity that receives linguistic coding. Also, unlike SUCCESSION wherein all activity segments are typically of equal importance, in PRECEDENCE, the activity marked by *nahaen* simply serves as the prelude of the upcoming activities, driven by politeness considerations to eliminate the treats of direct request or suggestion. The shift of focus to the first activity

of an event conforms to change of **attention (salience)**.

5.5. Context-sensitivity of lexical semantics

Meanings are constantly triggered by context, linguistic or extra-linguistic. In section 5.3.1, we employ Figure 5.1 to show that a smeared "C" is potentially construable as "O," due to a natural human tendency to conceive a stimulus as continuous and closed. Now compare Figure 5.3 with Figure 5.1. It is self-evident that when contexts are considered, some potential interpretations are ruled out, and some are preferred.

Figure 5.3. A smeared alphabet in its context
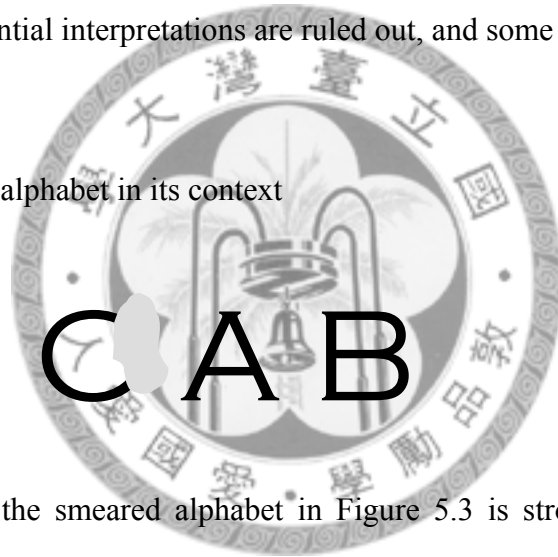
The interpretation of the smeared alphabet in Figure 5.3 is strongly restricted to "C" because *cab* makes a sense whereas *oab* does not. However, this is not the end of the story. Collocating alphabets are not the sole contextual clue that helps us to disambiguate. *Cab* makes sense in English, but not necessarily in other languages. In that sense, "context" should really be deemed an extended notion, including a broader background situation, settings, and social-cultural environment. As we have shown in Chapter 4 (§ 4.5), "contexts" in the present dissertation may include the followings:

(80) Types of contexts

    (I) Linguistic context

        a)   knowledge of the language

        b)   previous discourse

        c)   immediate linguistic environment

    (II) Non-linguistic context

        a)   situational context: the time and place, topic, register, relation between participants

        b)   background knowledge: social convention, conversational maxims, common sense of the objective world, mutual-understanding between participants

In this section, we will explicate the contextual factors that drive the semantic extensions of *nahaen*, and some specific structural patterns that have become or are becoming fixed with its collocated linguistic elements.
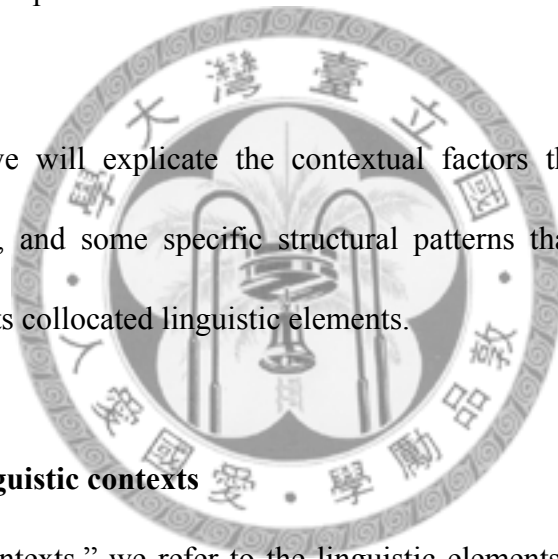
5.5.1. Meaning in linguistic contexts

By "linguistic contexts," we refer to the linguistic elements that co-occur with the lexical item *nahaen*. They include 1) the properties of those elements, and 2) the structure by which they are collocated. Linguistic contexts can be global, including the previous discourse and the register.

Boundedness of the activity

We have illustrated that [recurrence] and [continuation] involve change of

perspectives. One factor behind the choice of viewpoint is the boundedness of the activity: the repetitive sense applies to "accomplishments," "achievements" and "activities," whereas the continuous sense is prone to be activated by "states" and "activities." This is the case mostly because "accomplishments" and "achievement" hint at event termination whereas "states" and "activities" do not. Most Saisiyat verbs do not have pairs of bounded/unbounded contrast. For example, *ma'az'azem* can be a bounded activity 'think up' or an unbounded activity 'miss someone/something.' However, occasionally there are pairs of contrast, for example, *tihoeroe'* 'find' and *komi:im* 'look for.' They prompt different interpretations as shown in (81), but when *nahaen* modifies an unbounded activity, there are still two possible readings.

(81)a. *lasia   tihoeroe'   **nahaen**   ka   rayhil*
       3PL.NOM   find   **NAHAEN**   ACC   money
       3PL.NOM   zhǎodào **NAHAEN**   ACC   qían
       'He found (the) money again.'
       'Ta yòu zhǎodào qían lè.'
    b. *lasia   komi:im   **nahaen**   ka   rayhil*
       3PL.NOM   look.for   **NAHAEN**   ACC   money
       3PL.NOM   xúnzhǎo   **NAHAEN**   ACC   qían
       Reading 1: He is still looking for the money.
             'Ta hái zài zhǎo qían.'
       Reading 2: He is looking for money again.
             'Ta yòu zài zhǎo qían lè.'

Still, boundedness of an activity does not always predict whether the intended meaning is repetition of the activity or continuation of it. Note that (8) depicts a state of "being young," yet it receives [recurrence] reading because of the special contexts of the story that tells the re-juvenescence of human beings in a folklore. In the text, there is an

explicit mention of aging preceding *'al'alak nahaen* 'young NAHAEN,' and thus forcing a reading of [recurrence].

Clausal aspect

Saisiyat *ila* is analyzed in NTU-Formosan as a perfective marker that frequently denotes change of state.[5] When co-occurring with *ila*, *nahaen* always means [recurrence], even for activities that are intrinsically unbounded. For example, (81b) can be coerced to mean recurrence of an activity when the activity is marked by *ila*, as shown in (82).

(82)   *lasia*   *k<om>iim*   *ila*   **nahaen**   *ka*   *rayhil*
       3PL.NOM   look.for<AF>   PFV   **NAHAEN**   ACC   money
       3PL.NOM   xúnzhǎo<AF>   PFV   **NAHAEN**   ACC   qían
       'He is looking for the money again.'
       'Ta yòu zài zhǎo qían le.'

The contextual effect forced by *ila* is strong. This is due to the fact that "change of state" and continuation is incompatible. Change of state implies that there is a non-A' state so that we can talk about a change to A.' It thus forces a gap between A and A' and rules out the continuation reading.

Constructional pattern

Our data show that there is a structural pattern of *nahaen* that is probably under the

---

[5] It is analyzed as a marker of inception in Yeh (2000a, 2003).

process of "fixation", i.e. the process by which an item comes to occupy a fixed slot (Lehmann 1985). The pattern we are talking about is the SUCCESSION *nahaen*. It loosely follows this structure:

V$_1$ (*nahaen*), V$_2$ *nahaen*, V$_3$ *nahaen* …

This structure also strongly predicts our interpretation of *nahaen* as connecting a series of successive activities. However, this structural pattern is not as "local" as the kinds of constructions discussed in traditional studies of constructions (Fillmore et al. 1988; Langacker 1987a; Goldberg 1995). As shown in (73) in Section 5.3.4, this structure is constructed over 10 intonation units, and is clearly not at the sentential level. Called "constructional pattern" in this section, SUCCESSION is an interactional pattern in a broad sense of "construction" (Ford et al. 2003). Ford et al. (2003) suggest that some activities in discourse will accelerate the formation of specific interactional patterns. For example, adjectives are used in attributive structure "Adj. + N" as well as predicative structure "N + *be* + Adj.", and they are the "crystallization" (in terms of Ford et al.) of two kinds of activities we often do in discourse: introducing news referent and assessing already known referents, respectively. Similarly, what the speaker is doing by using SUCCESSION *nahaen* is to manage a series of activity to form a smooth storyline. The sequencing of incidents is of great importance in story-telling, and we believe this is the motivation that necessitates the formation of SUCCESSION structure.

There is a fuzzy boundary between interactional construction and what is traditionally called "grammar." As claimed by Langacker (1999), if a linguistic pattern is more

schematic and systematic, it is easily taken as "grammar" of a language. But "schematicity" is a matter of degree. The SUCCESSION structure may not be so schematic as to be called syntax or even a "construction" (in traditional definition), but it loosely follows a structure that to some extent correlates specific lexical elements to make a unit, and hearers can rely on this structural pattern to attain the intended meaning.

5.5.2. Meaning in extra-linguistic contexts

Extra-linguistic contexts, in our definition, include situational, social-cultural, and common knowledge. As we have mentioned, the distinction between [recurrence] and [continuation] (of the REPETITION sense) lies in the scope of perspective. Sometimes, whether the speaker is taking a narrow or wider perspective is as shared tactic between the speaker and the hearer. Observe the following example:

(83) Life
151. M: ...(1.4) *rangi'* *hiziyo'* *rima'* **naehaen** *ray* ... *bangol* ... *talboeyoe'*
               VOC   PN   AF.go **again** LOC   forest   AF.hunt
               VOC   PN   AF.qù **zài**   LOC   senlín   AF.lìe
      'Does Hiziyo still go hunting?'
      'Hiziyo háishì yǒu dǎlìe ma?'
152. F: ...(1.7) *'oka'* *ila*
               NEG   PFV
               NEG   PFV
      'No.'
      'Méiyǒu.'
153. M: ... *'oka'* *ila*
               NEG   PFV
               NEG   PFV
      'No.'
      ''Méiyǒu oh.'

154. F: *m==*
      FIL
      FIL

155. F: ...(0.9) *sia*   *mamowa'*   *ila*   *ka*   *ka'niw*

3SG.NOM	AF.plant	PFV	ACC	mushroom
3SG.NOM	AF. zhòng	PFV	ACC	xianggu

'He plants mushroom.'
'Ta zhòng xianggu.'

Speaker M and F are talking about their daily chores, and M suddenly asks whether Hiziyo' hunts. This statement in IU 151 in fact has many readings; according to its structure, [recurrence], [continuation], or PRECEDENCE are all possible readings. Knowing that Hiziyo' has hunted habitually for a period of time, M's question immediately invites F to construct a wider event-internal perspective, yielding a continuation reading (particularly habitual) in accordance to the background information of the situation.

The language users' experience in the speech community also constrains the possible reading of *nahaen*. Example (84) is an instance that manifests interesting social-cultural considerations.
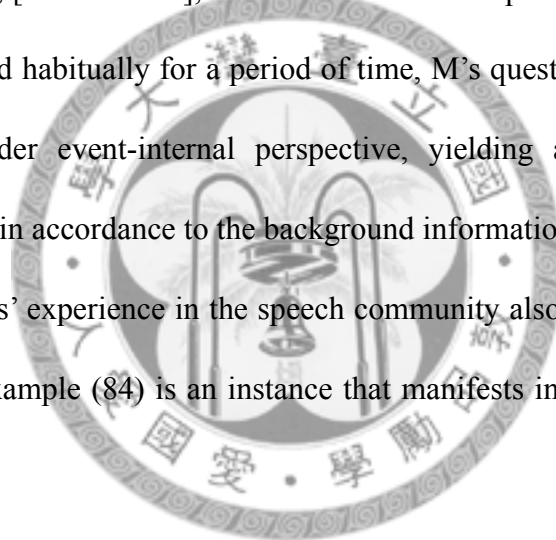
(84) *lasia*   *roSa'*   *rasiwazay*   ***nahaen***

3PL.NOM	two	separate	**NAHAEN**
3PL.NOM	èr	fenkai	**NAHAEN**

They broke for a while. (divorce)
'Tamén xian fenkai yizhènzi.'

There are two possible interpretations of this utterance: REPETITION and PRECEDENCE. However, the informants have a salient tendency to interpret this utterance in a PRECEDENCE sense. This is abnormal, because REPETITION is more salient in the *nahaen*

family. When asked for the reasons of this interpretation, the informants explained that Saisiyat society strongly follows the monogamy system, and there are very few cases of divorce, and remarriage is only found when one of the couple passes away. The REPETITION interpretation will involve the following scripts.

(85) They got married.
    > They got divorced.
    > They got married for the second time.
    > They got divorced for the second time.

The situation described in (85) is novel, and the informants reported that they have never seen such cases before. In this respect, how a marriage should be going on in the Saisiyat society has entrenched in their knowledge system and influenced their interpretation of a linguistic form in an unconscious manner.

5.6. Direct translation and semantic glossing

In the course of language investigation, we obtain different Mandarin translations of *nahaen* when it is used in different contexts. In this section, we show that the semantic network of *nahaen* is of some degree of cross-linguistic predictability. The Mandarin glosses of *nahaen* have a lot of functional overlap in Mandarin, and each Mandarin gloss does not correspond to a single sense of *nahaen*.

132

5.6.1. Direct translation of a polyseme

We begin by examining the relation between the senses of *nahaen* and the meaning of its Mandarin glosses. There are multiple direct translations used by the informants to interpret the meaning of *nahaen* in different contexts. Some of them can also be found in the examples extracted from NTU-Formosan. If we neglect some of the morphological variations, those direct translations can be grouped into seven, as shown in Table 5.1. In this table, we also summarize the dictionary meanings of those Mandarin translations as compared with the senses of *nahaen*.[6]

Table 5.1. Senses of *nahaen* and the dictionary meanings of its Mandarin translations

Translations / Senses of *nahaen*	*réngrán* 'still'	*hái* 'still'	*yòu* 'again'	*zài(dù)* 'again'	*lìngwài (lìng yigė)* 'another'	*yi-hŭir* 'a while'	*(shŏu)xian* 'first'
REPETITION							
[recurrence]		yes	yes	yes			
[continuation]	yes	yes		yes			
[addition]		yes	yes	yes	yes		
SUCCESSION			yes	yes			
PRECEDENCE						??[7]	yes

What has been revealed in Table 5.1 is that there is likely a division between PRECEDENCE and the other two senses in Mandarin. The Chinese gloss *xian* 'first' and *yi-hŭir* 'a while' has little functional overlap with other glosses according to their

---

[6] To know the Mandarin dictionaries employed in this study, please refer to Chapter 3 (cf. § 3.4.3). To see the dictionary meanings enlisted in the dictionaries for each Mandarin translation, please refer to Appendix E.

[7] When making suggestion or invitation, PRECEDENCE has a 'trivialize' function like Mandarin "*yi*-X" structures such as *yi-xìa* 'a while,' *yi-dĭan* 'a little bit,' or *yi-hŭir* 'a while,' and so on.

dictionary meanings. REPETITION and SUCCESSION are so closely intertwined that their Mandarin translations have functional extensions in a cross-boundary manner. In other words, many of them can be used to denote repetition of an activity as well as SUCCESSION of several activities.

5.6.2. Cross-linguistic predictability of semantic hierarchy

The complicated functional overlap and the inconsistent word translation might have been the effect of cross-linguistic predictability of semantic hierarchy. Most meanings of *nahaen* are found in other languages to be expressed by similar or the same linguistic forms. Among the meanings of *nahaen,* the three forces of REPETITION have a high degree of cross-linguistic predictability. The relation between [recurrence] and [continuation] has been attested in many languages, as has been discussed in Section 5.3.1. The other force, [addition], also has a strong association with [recurrence] and [continuation]. For example, in Japanese, the difference between [recurrence] and [addition] lies in whether we are talking about one more "activity" or one more "entity," and they can be expressed with the same lexical item, as shown in (86).

(86) Japanese
    a. ***mou***   *ichido*   *itte*   *kudasai*
         MOU   once   say   HONORIFIC
         'Please repeat the words again.'
    b. ***mou***   *hitsutsu*   *boutsu*   *o*   *katta*
         MOU   one   boot   OBJ   buy.PAST
         'I bought another pair of boots.'

It is thus not surprising when we found that Deutsch *noch*, used primarily to denote continuation, can also be used to encode repetition of an activity or addition of an entity, as shown in (31). It shows that the differences between [continuation], [recurrence], and [addition] are not conceived as so different in some languages, as evidenced by their sharing of the same marker.

(87) Deutsch
  a. *Ich   bin   **noch**   beschäftigt.*
      1SG.NOM   COP-1SG   **still**   busy
      'I'm still busy.' ([continuation])
  b. *Mach   das   bloß   nicht   **noch**   einmal.*
      make   DET-NEU   just   NEG   **still**   one.time
      'Never do that again!' ([recurrence])
  c. ***Noch**   etwas,   bitte?*
      still   something   please
      'Anything else?' ([addition])

SUCCESSION is comparatively less commonly associated with REPETITION. However, occasionally such an association can be found in crosslinguistic data. For example, as shown in (88), Mandarin Chinese encodes REPETITION and SUCCESSION both with the lexical item *zài*.

(88) Mandarin Chinese
  a. *sanshí   nían   hòu   **zài**   húidào   zhè   pìan   tŭdì*
      thirty   year   after   **again**   return.to   this   CLF   land
      ta   găndòng   dé   lùo   lèi
      3SG   feel.moved   COMP   fall   tear
      'After thirty years, he returned to this place. He was moved and cried.'
  b. *chui   gan   tóufã   hòu   **zài**   shu-tóu*
      blow   dry   hair   after   **again**   comb-hair
      jìu   búhùi   shanghài   fã-zhí
      CONJ   NEG   hurt   hair-quality
      '(You should) dry your hair first, and then comb them, so that (you) won't hurt your hair. '

It needs to be pointed out that some extensions of *nahaen* are not very predictable in a cross-linguistic aspect, such as SUCCESSION and PRECEDENCE; the sharing of the same form of REPETITION and SUCCESSION is also not very common. Nevertheless, its degree of cross-linguistic predictability is higher than the case of *nanaw*, and this also means that we are more likely to find similar semantic extensions in many other genetically unrelated languages. When the Mandarin translations are used directly as the gloss of *nahaen* in the corpus, the corpus users might discover the relation between the tokens of *nahaen* because its Mandarin glosses are functionally related. However, it is after all not economic if the corpus users have to take extra efforts to find out the polysemy network of *nahaen*.

In addition, multiple Mandarin glosses cannot reflect semantic partitioning of *nahaen*. Concepts that are conceived as relevant and expressed by the same lexical item *nahaen* in Saisiyat have been partitioned into several groups in Mandarin coded by different linguistic forms. Using the multiple Mandarin translations as glosses, we are likely to exaggerate the conceptual differences of the instances of *nahaen*, thus failing to show the way of conceptual categorization specific to Saisiyat.

5.7. Summary

In the course of our investigation, the lexical item *nahaen* is translated by the informants into multiple formally different Mandarin lexical items. After we look closely into the meanings of *nahaen*, we find that the meanings of *nahaen* are related to each

other, constituting a semantic network that is cognitively and pragmatically motivated. The functional difference between its tokens is not as drastic as has been hinted by its multiple Mandarin glosses.

The semantic network of *nahaen* has some degree of cross-linguistic predictability, and some of the associations can be found in Mandarin. Many of its Mandarin translations have salient functional overlaps in Mandarin. With careful observation, we are likely to find the relation between the instances of *nahaen*, and less likely to misconstrue the instances as homonymous.

Chapter 6   The Case Study of Saisiyat *Ma'*

6.1. Preliminaries

Semantic partitioning is to a large extent language-specific. For concepts that are conceived as relevant in one language, their relations may not be highlighted in another language and are thus expressed with formally irrelevant linguistic forms. For example, 'to click (a computer icon)' and 'to point at (something with finger)' are expressed with the same lexical item *dǐan* in Mandarin, but in English, the relation between these two actions are not highlighted, and are expressed by different linguistic forms.[1] Direct translation often fails to tell us the way of semantic partitioning specific to the target language, because the translation has to conform to the system of the meta-language. In fieldwork, direct translation is, however, a strategy that is frequently used, and the use of a meta-language in fieldwork seems to be inevitable in many cases (Matthewson 2004). In fact, there have been attempts to replace the use of meta-language with non-verbal experiments in fieldworks, but they can elicit limited types of linguistic data, and are also constrained by the design of experiments (Matthewson 2004).

There have been a lot of studies of comparative semantic that report the mismatches of semantic partitioning in different languages, such as Wierzbicka (1996), Bowerman and Choi (2003), Botne (2005), etc. In the previous two chapters, we present two studies

---

[1] This example is drawn from the speech of Lily I-wen Su "Metaphor as a way of categorization: Evidence from metalanguage glosses" delivered in Academia Sinica, May 16, 2005.

that exhibit cross-linguistic semantic mismatches. But are there cases that show crosslinguistic similarity? According to Sweetser (1986), when a semantic network is of a high degree of abstractness, it is more likely to have cross-linguistic semantic predictability, and similar semantic extensions may be found in other languages. "Coordination" and "negation" are two cases that are suggested by Sweetser to have high abstraction.

In this chapter, we present a case study of Saisiyat *ma'*, a marker that is very close to English 'also.' This lexical item has multiple functions, but it also has a high degree of cross-linguistic predictability. The functions that are denoted by *ma'* in Saisiyat are found to be predominantly denoted by a single lexical item in Japanese and also in Mandarin. The high degree of crosslinguistic semantic predictability of this lexical item is found to be accompanied by a high degree of abstraction, which is manifested by its heavy dependence upon collocating structures.

This chapter is organized as follows. Section 6.2 presents a short syntactic description of Saisiyat *ma'*. Our analyses are presented in Section 6.3, explicating the senses of *ma'* by examples taken from NTU-Formosan and our direct elicitation. In section 6.4, we will identify their cognitive or pragmatic links by which the meanings are chained together as a "family" (following Wittgenstein's notion of "family resemblance"). We found that the meaning of *ma'* is closely integrated to its linguistic contexts, and several constructional patterns related to *ma'* are highly predicative of its interpretation. We will talk about semanticization and its relation to contexualization as well as construction formation (or

structural patterning). Finally, in Section 6.6, we will turn back to examine the direct translations of *ma'* and discuss some of its implications of semantic glossing.

6.2. A syntactic sketch of Saisiyat *ma'*

Saisiyat *ma'* is an adverb that indicates that an expression is similar to another expression that is previously mentioned.[2] It typically occurs between the subject and the predicate. Interestingly, it can have its scope of modification to the right or to the left. For example, (89) can have two readings, depending on whether *ma'* modifies its preceding or following elements.

(89) *kizaw* ***ma'***   *S<om>bet*   *ka*   *'ahoe'*
      PN   **MA'**   beat<AF>   ACC   dog
      PN   **MA'**   dă<AF>   ACC   gŏu
      Reading 1: Someone beat the dog, and Kizaw also beat the dog.
      Reading 2: Kizaw beat other thing, and she also beat the dog.

Saisiyat has three focus constructions: AF (agent focus), PF (patient focus), and RF (referential focus). Verbs are affixed with focus markers to signal the semantic role of the sentential subject (the nominative case). The lexical item *ma'* can occur in all three constructions, as shown (90).

---

[2] We exclude cases wherein *ma* is the omission of *'ima* (progressive aspect marker). Cases of *ma=* (with a lengthening) or *ma:* (with a lengthened vowel) are pause fillers in Saisiyat. They are also excluded in our current discussion.

(90)a.   *'obay*   *S<om>bet*   *ka*   *'ahoe'*
           PN   beat<AF>   ACC   dog
           PN   dă<AF>   ACC   gŏu
           iban   **ma'**   **S<om>bet**   *ka*   *'ahoe'*
           PN   **MA'**   **beat<AF>**   ACC   dog
           PN   **MA'**   **dă<AF>**   ACC   gŏu
           ''Obay beat the dog, and Iban also beat the dog.'
           ''Obay dă gŏu, Iban yĕ dă gŏu.

   b.   *iban*   *Sebet-en*   *ni*   *'obay,*   *kizaw*   **ma'**   **Sebet-en**   *ni*   *'obay*
           PN   beat-PF   GEN   PN   PN   **MA'**   **beat-PF**   GEN   PN
           PN   dă-PF   GEN   PN   PN   **MA'**   **dă-PF**   GEN   PN
           ''Obay beat Iban, and he also beat Kizaw.'
           ''Obay dă Iban, yĕ dă Kizaw.

   c.   *oya'*   *si-Sebet*   *ka*   *'ahoe'*
           mother   RF-beat   ACC   dog
           mama   RF-dă   ACC   gŏu
           korkoring   **ma'**   **si-Sebet**   *ka*   *'ahoe'*
           child   **MA'**   **RF-beat**   ACC   dog
           xiăohái   **MA'**   **RF-dă**   ACC   gŏu
           'Mother saw the dog being beaten, and the child saw that, too.'
           'Mama kàn dào gŏu bèi dă, xiăohái yĕ kàn dào gŏu bèi dă.'

In conversation, *ma'* is sometimes found to occur in utterance-initial position. For example, in (91), the subject of IU 120 is the entire utterance uttered by speaker F in IU 118-119. In speech interaction, *ma'* is often used to voice an evaluation on a topic of conversation, or to manage change of turns between the speaker and hearer, often followed by a short pause.

(91) Life

118.	F:	...(0.8)	*ma'an*	*p<in>amowa'-an*	*ititi'an*	*nanao* .\	
			1SG.GEN	plant<PFV>-NMZ	a.bit	only	
			1SG. GEN	zhòng<PFV>-NMZ	yidiǎn	zhǐ	

119.	F:	..	*okik*	*lal'oz*	*noka*	*mae'iyaeh*	*mari'-in*	*ba:iw-in* .\
			NEG	enough	GEN	person	take-PF	buy-PF
			NEG	zúgòu	GEN	rén	ná- PF	mǎi- PF

'Just a bit. Not enough to sell.'
'Zhǐyǒu yidiǎn. Búgòu mài.'

120.	M:	...(0.8)	***ma'***	..	*nakisaza*	*kayzaeh*
			FIL		like.that	AF.good
			FIL		xiàngnàyàng	AF.hǎo

'That is good.'
'Nà hǎo ah.'

This structure is only found in spontaneous speech interaction. In direct elicitation, it is not acceptable to begin an utterance with *ma'*.
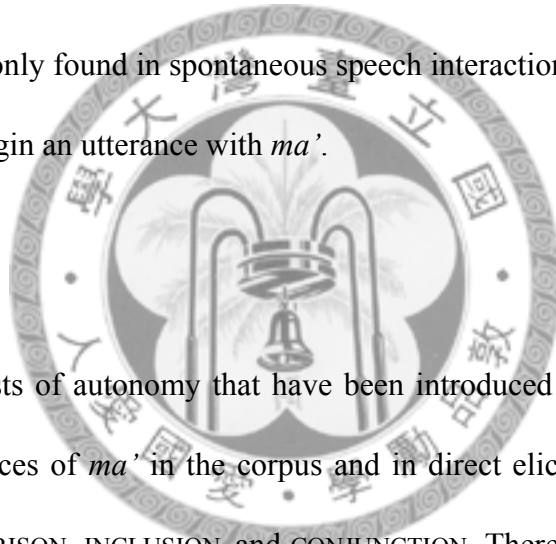
6.3. Senses of *ma'*

Following the tests of autonomy that have been introduced in Chapter 3 (§ 3.4.1), we examine all instances of *ma'* in the corpus and in direct elicitation. There are three senses of *ma'*: COMPARISON, INCLUSION, and CONJUNCTION. There is one idiomatic sense, (CONNECTION) that occurs together with *isaa* 'that' as a discourse cohesive device.

Instances of one sense may have different emphases to denote subtle changes of configuration, or to fulfill specific conversational demands. They can be divided into at least seven forces. Those forces are not contrastive in terms of autonomy, including perspectives, active zones, contextual modulations, etc. Please see Appendix D for their definitions in Croft and Cruse (2004). In sum, the senses and forces of *ma'* can be listed below.

COMPARISON

    [similarity]

    [listing]

    [counter-expect]

    [periphery]

INCLUSION
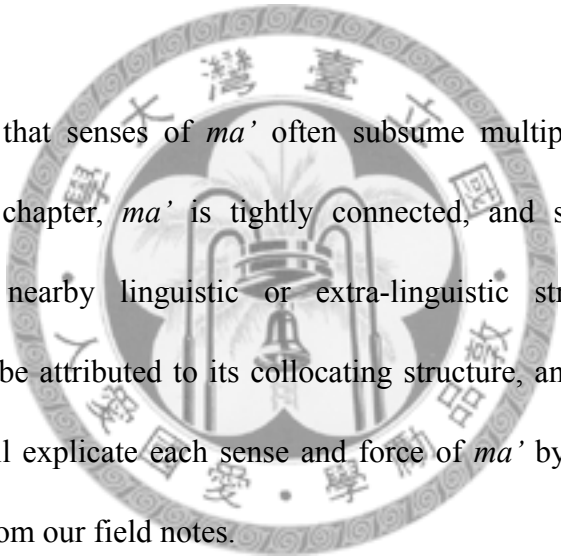
    [indefiniteness]

    [extremity]

    [collectiveness]

CONJUNCTION

(CONNECTION)

It is noteworthy that senses of *ma'* often subsume multiple forces. As we will explain later in this chapter, *ma'* is tightly connected, and sometimes obligatorily collocated, with its nearby linguistic or extra-linguistic structure; many of its interpretations should be attributed to its collocating structure, and are not autonomous. In this section, we will explicate each sense and force of *ma'* by citing examples from NTU-Formosan and from our field notes.

6.3.1. COMPARISON

The COMPARISON sense emphasizes that **"two or more expressions share at least one identical element."** Four forces are subsumed under COMPARISON, each with slightly different emphases: [similarity], [listing], [counter-expect, and [periphery].

[Similarity]

The [similarity] force refers to cases wherein one expression shares some comparable commonalities with a previously-mentioned expression. An expression that is marked with *ma'* is similar to a previously-mentioned expression if they depict the same action, share the same actor, impose an action on the same patient, take place in the same location or time, and so forth. Example (92) is extracted from a Saisiyat folklore, in which a mysterious lady from water came to the Saisiyat tribe and taught the Saisiyat people to weave. This excerpt describes how the lady met a young Saisiyat hunter. They were shy when they met each other at first, but no longer so after the lady heard the farting of a dog. She laughed, so did the young man.

(92) Kathethel 2

156.	...(1.9)	*in'alay*	*ri'sa*	*bazae'*	*ka*	*intot*	*hiza*
		be.from	there	AF.hear	ACC	fart	that
		cóng	nàlǐ	AF.ting	ACC	fàngpì	nà

157.	...	*kabinao*
		lady
		xiǎojiě

158.	...(1.0)	*isa*
		FIL
		FIL

159.	...(0.8)	*baza'*	*na'ora==*
		AF.hear	suddenly
		AF.ting	túrán

160.	...	*sawa'*	*ila*	*hiza*	*ka==*	*kamo'alay*	***ma'***	*sawa'*	*ila*
		AF.laugh	PFV	that	NOM	young.man	**also**	AF.laugh	PFV
		AF.xìao	PFV	nà	NOM	níancingrěn	**yě**	AF.xìao	PFV

'Upon hearing the dog fart, the lady laughed. The young man also laughed.'
'Tingdào fàngpì sheng, nà nǔrén xìao lė, níanqingrén yě xìao lė.'

When the two joined activities share the same subject, the use of *ma'* indicates that "the subject is doing other activities in addition to the first-mentioned." In (93), the speaker is talking to another Saisiyat regarding the work he undertakes to raise his deer.

(93) Election

8.	B:	...(0.8)	*o:*	*matawaw*	*maelahaeng*	*ka==*	*s<in>pan*
			DM	AF.work	AF.care	ACC	raise<PFV>
			DM	AF.gongzùo	AF.zhàogù	ACC	yǎng<PFV>
			ka	*wa'ae'*			
			ACC	deer			
			ACC	lù			

9.	B:	...	*t<om>abo*	*s<om>apeh*
			feed<AF>	sweep<AF>
			wèi<AF>	dǎsǎo<AF>

10.	B:	...(0.8)	*o:*
			DM
			DM

11.	B:	...(1.1)	***ma'***	*...*	*'ima*	*...(1.0)*	*kikoso'*
			too		PROG		weed.grass
			yě		PROG		chúcǎo

'I feed the deer, sweep the place, and do the weeding too.'
'Oh, wǒ zhàogù wǒ yǎng de lù ah, wèi tāmen chī dōngxi ah, děsǎo ah, oh háiyǒu, chúcǎo ah.'

In (93), the speaker randomly provides a list of actions. He then employs *ma'* in IU11, upon recalling one thing that can be added to the list. In a sense, *ma'* functions as an additive marker. However, when the speaker wishes to emphasize the chronological order of joined activities, he will rather choose to use *nahaen* instead (which we have introduced in Chapter 5).

The two activities do not have to share exactly identical elements; in many cases, an expression is comparable to another because they share a cognitive unifying "ground":

in the same frame, for the same purpose, or conceptually contiguous via any possible link.

For example:

(94) *'obay* *'alehel,* *iban* **ma'** *minhaetis*
    PN   cough   PN **MA'** sneeze
    PN   késòu PN **MA'** dǎpentì
    "Obay coughed, and Iban also sneezed.'
    "Obay késòu, Iban yě dǎpentì.'

In (94), the two activities do not share the same actor, and the activities themselves are not identical. However, the use of *ma'* is acceptable in this case because the two activities "coughing" and "sneezing" are in the same frame, i.e. the frame of being sick. In fact, the compared expressions sometimes look contradictory, as shown in (95).

(95) *'ima* *Sobaoeh* *Sobaoeh*   *'atomalan*
    'IMA AF.big   AF.big   very
    'IMA AF.dà   AF.dà   feicháng
    'ima *'ol'ola'an* **ma'** *'ol'ola'an* *'atomalan*
    'IMA AF.small **MA'** AF.small very
    'IMA AF.xǐao   **MA'** AF.xǐao feicháng
    'The big ones are too big, and the small ones are too small.'
    'Dà dė tài dà, xǐao dė yòu tài xǐao.'
    (→ I didin't find a satisfactory one)'

What the speaker intends to convey by (95) is that he does not find the size he wants: Neither the big ones nor the small ones are satisfactory. When the speaker's intention is taken into consideration, these two compared expressions then become not contradictory at all; they are similar in that the speaker does not like both sizes.

The [listing] force also denotes similarity of two or more activities. However, unlike [similarity] which compares two expressions in an "asymmetric" relation, i.e. the later-mentioned activity resembling a first-mentioned, the [listing] force views every expression as symmetric, which is attested by the fact that every expression being compared carries the *ma'* marker. Example (96) is extracted from a story of flood. When a flood came, the protagonist put all the necessities on boat, and made some arrangements.

(96) Flood

44.	...(1.0)	*ka==*		
		FIL		
		FIL		
45.	...	*pazay*	***ma'***	*nonak*
		rice	**too**	self
		mǐ	**yě**	gèzì
46.	..	*ralom*	***ma'***	*nonak*
		water	**too**	self
		shǔi	**yě**	gèzì

47.	..	*'ahoe'*	***ma'***	*nonak*	*mae'iyaeh*	***ma'***	*nonak*
		dog	**too**	self	man	**too**	self
		gǒu	**yě**	gèzì	rén	**yě**	gèzì

'There was room for putting the rice and water, as well as for the animals and the family members.'
'Yǒu dìfang fàng mǐ, yǒu dìfang fàng shǔi, yǒu dìfang fàng gǒu, měiġė rén dou yǒu gèzì dė dìfang.'

Unlike [similarity] wherein only the latter-mentioned activity carries *ma'* marker, in (96), every activity carries the marker *ma'*, iconically showing the equal salience of each activity, constituting a group wherein all members are identical in some respect.

[Counter-expect]

For [counter-expect], an activity resembles a previously-mentioned activity, but in a disjunctive mood. Instances of [counter-expect] is usually preceded by a non-favorable situation for an activity to take place. The use of *ma'* indicates that although in a non-favorable situation, the activity still takes place, countering an implicit or explicit expectation. In example (97), the informant made up a story in which Kizaw went fishing but was caught in a rain. When it is raining, one typically finds a shelter to wait for the rain to stop. When one still insists going home in spite of bad weather, it is countering our normal expectation.

(97)   *'a'oeral*   *'ima*   *'<oem>oeral*   *kizaw*   **ma'**   *'am*   *lobih*   *ila*
       rain   PROG   rain<AF>   PN   **MA'**   want   AF.return   PFV
       xiàyǔ   PROG   yǔ<AF>   PN   **MA'**   yào   AF.húi   PFV
       'Kizaw insisted going home regardless of the rain.'
       'Yǔ zài xià, Kizaw háishì yào húiqù.'

In Chapter 5, we introduced an example of disjunctives that is marked by *nahaen*, as in (98 a). In fact, the same scene can be expressed by using *ma'*, as shown in (98b).

(98) a.   *'ima*   *haSa'*   *h<om>ayap*   *kabkabahae'*   *koSa'-en*   *kabkabahae'*   **nahaen**
          'IMA   NEG   fly<AF>   bird   say-PF   bird   **NAHAEN**
          'IMA   NEG   fei<AF>   niǎo   shuo-PF   niǎo   **NAHAEN**
          'Búhùi fei dè niǎo háishì jìaozùo niǎo.'
          'Birds that cannot fly are still called birds.'
     b.   *'ima*   *haSa'*   *h<om>ayap*   *kabkabahae'*   **ma'**   *koSa'-en*   *kabkabahae'*
          'IMA   NEG   fly<AF>   bird   **MA'**   say-PF   bird
          'IMA   NEG   fei<AF>   niǎo   **MA'**   shuo-PF   niǎo
          'Búhùi fei dè niǎo háishì jìaozùo niǎo.'
          'Birds that cannot fly are still called birds.'

According to the informants, although *nahaen* is acceptable in disjunctive utterances, a more natural way of expression disjunction is via the use of *ma'*. In other words, although both (98a) and (98b) are acceptable, (98b) is considered more appropriate than (98a).

[Periphery]

In many cases, an expression is marked by *ma'* to convey the speaker's opinion in a modest way. For example, (99) is extracted from a conversation. Two speakers are talking about the ducks raised by M. Speaker F assumes that M does not raise ducks by automated incubator because the ducks will become tasteless. To deny F's assumption, M marks his negative answer with *ma'* to mitigate possible threat.

(99) 86. F: ... *'aewhaey*   *ka-si'ael-en*   *./*
                   AF.bad   KA-eat-PF
                   AF.bùhǎo   KA-chi-PF
    'Taste bad?'
    'Wèidào bùhǎo?'
    87. M: ... **ma'**   *'okay* *koSa*   *'aewhaey*   *ka-si'ael-en* *.*
                        also   NEG   AF.say   AF.bad   KA-eat-PF
                        yě   NEG   AF.shuo   AF.bùhǎo   KA-chi-PF
    'It's not that they taste bad.'
    'Búshì shuo wèidào bùhǎo.'

Biq (1989) in her study of Mandarin *yě* 'aslo', the counterpart of Saisiyat *ma'*, reports that *yě* is used to express an evaluation or judgment in a roundabout way, which is very much like the function of Saisiyat *ma'* in example (99). We call this force [periphery]

because we consider the speaker is making a kind of comparison, associating an expression to a presupposed or implicit standard, but deliberately placing the compared expression in the peripheral position that passably meets the standard. To downgrade the degree of typicality has a pragmatic function to mitigate the imposition of a statement.

The following example (100) can better illustrate the effect of downgrade. Example (100) extracted from a Saisiyat folktale (a sequel to (92)). After the nymph from the water married the Saisiyat youth, she taught the Saisiyat women how to weave. Some Saisiyat women thus became arrogant for the weaving skills they possess, and offered a challenge to the nymph. The following expresses such a sentiment of a Saisiyat woman.
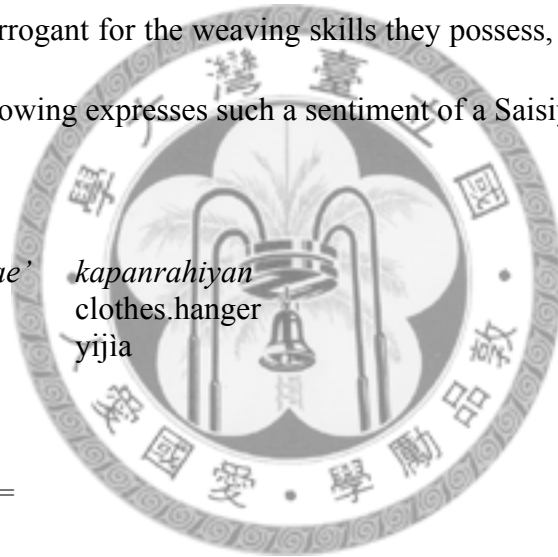
(100) kathethel 2
408.	...(0.9)	*'aehae'*	*kapanrahiyan*		
		one	clothes.hanger		
		yi	yijìa		
409.	...	*yo*			
		also			
		yò			
410.	...(1.2)	*ma==*			
		FIL			
		FIL			
411.	...(2.1)	*to:o'*	*may'amet*	*ila*	*zi'saza==*
		three	AF.filled.to.the.brim	PFV	there
		san	AF.saimǎn	PFV	nalǐ
412.	...	*hao*			
		DM			
		DM			
413.	...(1.4)	*nisia*	*ma'* ..	*'ima*	*k<om>oSa:*
		3SG.GEN	DM	PROG	say<AF>
		3SG. GEN	DM	PROG	shuo<AF>
		ma'an	*hini*	*'aehae'*	*kapanrahiyan*
		1SG.GEN	here	one	clothes.hanger
		1SG. GEN	zhè	yi	yijìa

414.	...(0.8)	***ma'***	*ak*	*'akoy*	*'atomalan*	*ila*
		also	FS	AF.many	very	PFV
		yĕ	FS	AF.duo	hĕn	PFV

'The other person said, 'I have hung one hanger. There are too many clothes.''
'Yigė yijìa yòu san jìan. Kùa dė mănmăn dė. Ta shuo, wŏ yŏu yigè yijìa dė yifú,yĕ yĭjing hĕn duo lė.'

The clothes woven by the arrogant woman entirely occupy one clothes-hanger, and there are three other clothes in addition to those hung on the hanger. The use of *ma'* indicates that the amount may not be high, but is good enough. This is used as a politeness device to be humble. In fact, the arrogant Saisiyat woman considers herself a very good weaver, and she is actually very proud of the amount of the clothes she has woven in a short time. This is evident from (101) which is a sequel to (100).

(101) kathethel 2

415.	...	*kayzaeh*						
		AF.good						
		AF.hăo						
416.	...	*ma'an*	*komoSa:*	*la'oz*	*ila*	*hini*		
		1SG.GEN	DM	AF.enough	PFV	this		
		1SG. GEN	DM	AF.gòu	PFV	Zhè		
417.	...	*la'oz*	*ila*	*si'an*		*hini*	*kin*	*'akoy*
		AF.enough	PFV	AF.exceedingly.many		this	very	AF.many
		AF.gòu	PFV	AF.feicháng		zhè	hĕn	AF.duo

'Very good. My clothes are enough. Many clothes.'
'Hĕn hăo. Wŏdė yĭjing gòu duo lė. Feicháng dė duo.'

In (101), the speaker does not necessarily mean that the delimitated entity is in fact of a low quantity. According to Brown and Levinson (1987), "downgrading" is one strategy to show the speaker's politeness considerations to not being too assertive or pompous. Employing *ma'* to downgrade an evaluation can make the speaker's viewpoint

more acceptable to the hearer.

[Counter-expect] and [periphery] has a fine distinction. We suggest that they are highlighting different aspects of the following script of "atypical resemblance:"

(102) a. An expression (E') is not similar to a previously-mentioned expression (E),

b. and for E' to resemble E is countering our expectation,

c. and even when E' resembles E, it is not a typically situation.

[Counter-expect] highlights (102a) and (102b), whereas [periphery] highlights (102a) and (102c).

6.3.2. INCLUSION

The second sense of *ma'* is INCLUSION. Unlike COMPARISON which views the sharing of a property as a "chaining" relation, i.e. each entity resembling the other in a sequential order.[3] INCLUSION on the other hand takes a "collective" view, conceptualizing the sharing of a property by viewing the members as collectively involved in a situation. Reflected in its structure, it is used clausal-internally, and is more structurally dependent.

INCLUSION includes three forces: [indefiniteness], [extremity], and [collectiveness]. We will explicate each force in this section.

---

[3] We borrow the terms "chaining" and "collective" from Lichtenberk (1985). According to Lichtenberk (1985), "chaining" denotes that participant A stands in a certain relation to B, B stands in the same relation to C, C to D, etc. "Collective" denotes that "two or more participants are together involved in a situation, in the same Initiator-type role."

[Indefiniteness]

Speakers often use indefinite markers *'ana* together with *ma'* to form a construction denoting shared commonalities of a group of entities. The indefinite morpheme *'ana* 'whether' is collocated with interrogatives, such as *hiya'* 'who,' *kano'* 'what,' and *hayno'* 'where' to indicate that any indefinite member in a specific group has a specific property shared by all members. Example (103) is extracted from a two-party conversation on the topic of election. They lament that Saisiyat candidates never win out in any election, because the Saisiyat people are not united to support their own candidate to compete with Hakka opponents. In this excerpt, the speaker uses *ma'* in IU 109 to express that Saisiyats should support ANY Saisiyat candidate regardless of who is to stand out as the representative.
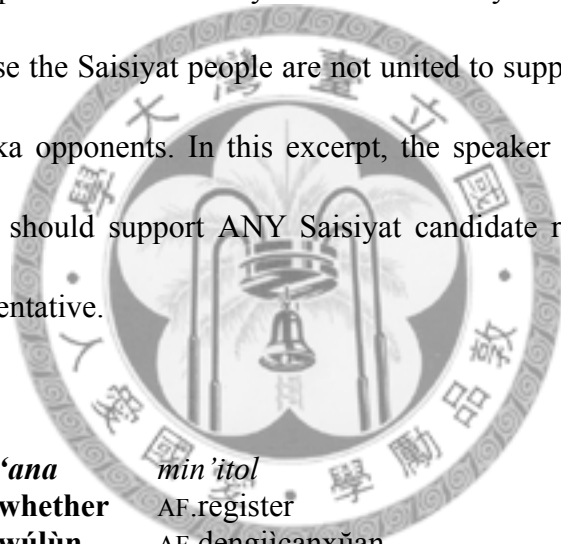
(103) Election
107.   M: ...(0.9)   **'ana**   *min'itol*
                      whether   AF.register
                      wúlùn   AF.dengjìcanxǔan
108.   M: ...(1.5)   *hia'*   *k<in>ita'*
                      who   see<PFV>
                      shéi   kàn<PFV>
109.   M: ...(1.6)   *makakreng*   *no*   *matawaw*   ...   *min'itol*
                      AF.hardworking   DAT   AF.work   AF.register
                      AF.yònggong   DAT   AF.gongzùo   AF.dengjìcanxǔan
                      mita'   **ma'**   *kaela'haeng-en*
                      1IPL.GEN   **also**   care-PF
                      1IPL.GEN   **yě**   zhàogù- PF
'Whoever is elected, as long as they work diligently, we will also support him.'
'Wúlùn shéi canxǔan, zhǐyào rènzhen zùoshìi, womėn dou hùi zhichí.'

The collocation of *ma'* with *'ana* is very common in our database. The marker *'ana* can precede an indefinite pronoun (104a) or the names of the members in the group as in (104b).[4]

(104)a. **'ana**   *kano'*   **ma'**   *kayzaeh*
      whether what   **MA'**   AF.good
      wúlún   shémė   **MA'**   AF.hǎo
      'Anything will do.'
      'Shémė dou hǎo.'

    b. **'ana**   *iban*   *ki*   *'obay*   *ki*   *'oemaw*   **ma'**
      whether PN   COM   PN   COM   PN   **MA'**
      wúlún   PN   COM   PN   COM   PN   **MA'**
      ma'an   *sarara-en*
      1SG.GEN like-PF
      1SG.GEN xǐhuan-PF
      'I like Iban, 'Obay, and 'Oemaw, whoever he is.'
      'Bùguǎn Iban, 'Obay, háishì 'Oemaw wǒ dou xǐhuan'

According to Haspelmath (1997), indefiniteness is often expressed with interrogative markers in a cross-linguistic aspect. By using indefinite markers (and interrogatives), the speaker means that any random member of a group (in this case, the Saisiyat people) will own a shared property (in this case, to have their people's support).

[Extremity]

Collocation with *'ana* involves not only [indefiniteness]; it also involves the employment of non-typical members of a category to denote the sharing of a property in

---

[4] The lexical item *'ana* has multiple readings. In different contexts, direct translation renders multiple interpretations roughly equivalent to English 'regardless,' 'whether,' 'at least,' 'whatever,' 'although,' and 'ever.' A proper way of glossing this lexical item demands thorough investigation into its semantic extension. It is nevertheless beyond the scope of the present dissertation, and will be saved for future study.

the category. The speaker often mentions the most peripheral case to emphasize that other

typical cases of course possess the property, as shown in (105), two utterances elicited

from our fieldwork.

(105) a. By mentioning the lowest number (always in negative construction)

'ana	*'aehae'*	***ma'***	*oka'*	*'i*	*mari'-i*	*ka*	*'aelaw*
whether	one	**MA'**	NEG	NEG	take-PF	NOM	fish
wúlùn	yi	**MA'**	NEG	NEG	ná-PF	NOM	yú

'(He got) no fish at all' (Literally: 'not even one fish')
'Lían yi tíao yú dou méi dìao dào.'

b. By mentioning the most unexpected member

'ana	*mayakay*	***ma'***	*hasa'*
whether	AF.talk	**MA'**	unable
wúlùn	AF.shuōhùa	**MA'**	búhùi

'(He) can't even speak.'
'Lían shuōhùa dou búhùi.'

In example (105a), by mentioning the lowest number, i.e. "one," the speaker means

that the subject caught not even one fish. If "one," an extremely low number, is not

included, other numbers will be excluded as well. (105b) employs the most unexpected

activity to denote someone's incapability. Speaking is an ability that everybody has. If

one cannot speak, it implies that he can do nothing else either. Like the case of Mandarin

yě investigated in Biq (1989), *ma'* of the [extremity] force manifests abstract reasoning of

"scalar" inference — Members of extremity, atypicality, and unpredictability are

employed to show the absolute inclusion of central, typical, and predictable members.

[Collectiveness]

Shared commonality can also be expressed by *ma'* via its collocating with quantifiers such as *sa'sa'ih* 'each' and *saboeh* 'all.' Example (106) is extracted from a conversation in which two Saisiyats are talking about the tomb-sweeping ritual.

(106) Holiday

60.	C:	...(1.5)	***sa'sa'ih***	*ka*	*mae'iyaeh*	***ma=='***
			every	LNK	person	**DM**
			měi	LNK	rén	**DM**
61.	C:	..	*'iska*	*nonak*	*baiw*	
			each	self	AF.buy	
			měi	zìjǐ	AF.mǎi	

'Everybody buys his own things to offer.'
'Měigè rén mǎi zìjǐ dè dōngxi.'

The Saisiyats have a "family system." All Saisiyats can be divided into seven "manayahae" (kin). Each "manayahae" has two to four "sinrahoe'" (clan), distinguished by their family names. In this text, the speaker is of the Titiyon family which is kin to the Bobotol family. Although kin members participate in the tomb-sweeping ceremony together, in this example, the speaker uses *ma'* with distributiveness marker *sa'sa'ih* in IU 60 to emphasize the fact that each family buys their own offerings.

The morpheme *saboeh* 'all' can also collocate with *ma'* to denote sharing of a commonality in a group. The instances are not found in NTU-Formosan, but we came across several instances in direct elicitation. The following example (107) is taken from our field notes. It means that two people BOTH/ALL do the same action, i.e. hitting Iban.

(107)   *'obay*   *S\<om\>bet*   *hi*   *iban*   *kizaw*   *S\<om\>bet* *hi*   *iban*

PN	hit\<AF\>	ACC	PN	PN	hit\<AF\>	ACC	PN
PN	dǎ\<AF\>	ACC	PN	PN	dǎ\<AF\>	ACC	PN

      rosa'   *may'iaeh* **ma'** *S\<om\>bet* ***saboeh*** *hi*   *iban*

two	people	**MA'**	hit\<AF\>	**all**	ACC	PN
èr	rén	**MA'**	dǎ\<AF\>	**quánbù**	ACC	PN

    ''Obay hit Iban, and Kizaw also hit Iban. They both hit Iban.'
    ''Obay dǎ Iban, Kizaw yě dǎ Iban. Liǎng gè rén dou dǎ Iban.'

However, the collocation of *ma'* with *saboeh* to denote collectiveness is not very common, and the meaning in fact comes largely from the context. The preceding context "*'obay s\<om\>bet hi iban kizaw s\<om\>bet hi iban*" cannot be omitted. Simply saying "*rosa' may'iaeh ma' s\<om\>bet saboeh hi iban*" has a reading that some people hit Iban, and now there are two other people who also hit Iban.
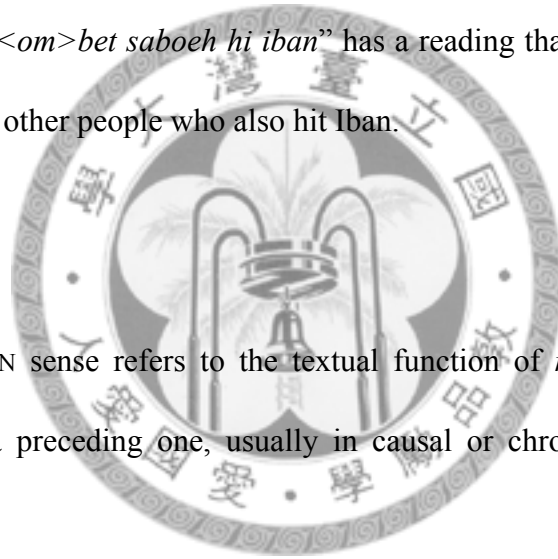
6.3.3. CONJUNCTION

The CONJUNCTION sense refers to the textual function of *ma'* when it is used to connect a clause to a preceding one, usually in causal or chronological relation. For example:

(108) Life

221. F: ..   *so:*   *la'oz*   *ila*   **ma'**

COND	enough	PFV	**DM**
COND	zúgòu	PFV	**DM**

222. F: ...(1.2) *ka-hi-hiwa'-en*   *ila*   *ka-si'ael-en*

KA-RED-kill-PF	PFV	KA -eat-PF
KA-RED-sha-PF	PFV	KA -chi- PF

                ka-powa'-en   *so:*   *tatini'-in*

KA -why-PF	COND	old-PF
KA -wèishémė- PF	COND	lǎo- PF

    'The ducks can be killed as soon as they grow old. Why do you raise them until old age?'
    'Yazi dà lė jìu kěyǐ sha, wèihé yào yǎng dào lǎo?'

The speaker is talking about the ducks she keeps. The use of *ma'* indicates that when a condition stands, a result naturally follows. Cognitively and pragmatically, CONJUNCTION has a close relation with the COMPARISON sense, but they are different in that COMPARISON compares two clauses that are similar in some way, whereas CONJUNCTION conjoins two clauses that are not similar but related in some way. We will explain this point in detail in Section 6.4.

6.3.4. Idiomatic sense: (CONNECTION)

The (CONNECTION) idiom refers to two fixed expressions *isaa ma'* and *ma' isaa* that are frequently used in spontaneous speech to manage discourse cohesion. We classify instances of these two fixed expressions in the (CONNECTION) sense, and use parenthesis to indicate that it is an idiomatic sense that has a strong tendency to collocate with specific lexical items. In the following, we will show that in the instances of (CONNECTION), *ma'* is tightly collocated with *isaa*, and the collocation has specific discourse functions. Examples (109) and (110) contain the use of *ma' isaa* and *isaa ma'* respectively.

(109) Life
147. F: *m==*
            FIL
            FIL

148. F: ...   *'okik*   *wa'isan*   *ka-ba:iw-an*   **ma'**   *isaa*   ...
            NEG   expensve   KA-buy-NMZ   **also**   **there**
            NEG   gùi   KA-mǎi- NMZ   **yě**   **nàlǐ**
            'am   *nak'ino'*
            FUT   so.what
            FUT   nàyòurúhé
        'The market price is not good. What can we do?'
        'Shìchǎng shàng dė jìagé bùhǎo, wǒmėn zěmė bàn?'

(110) holiday
21. K: ...   *patabir*   *ila*
           AF.do.ritual   PFV
           AF.jìsì   PFV

22. K: ...(1.0)   *patabir*   *so:*
           AF.do.ritual   COND
           AF. jìsì   COND
    'We do the rituals.'
    'Ránhòu jìu jìzǔ.'

23. K: ...   *sizaeh*   *ila*   ***isaa***   ***ma***
           AF.finish   PFV   **then**   **??**
           AF.jíeshù   PFV   **ránhòu**   **??**

24. K: ..   *'a-'ayna:a'*   *yami*
           RED-AF.wait   1EPL.NOM
           RED- AF.tĕng   1EPL.NOM

25. K: ...   *to:o'*   *sinraehoe'*
           three   clan
           san   shìzú
    'After we finish (weeding), we wait for the others. There are three clans altogether.'
    'Jìbài wán lė, ránhòu wǒmėn (měiyigė rén) jìu děng san gė xìng dė ren tongtong dào cí.'[5]
    <u>Note: *isaa ma'* is a fixed expression.</u>
    <u>Note: *isaa ma'* shì gùdìng yùngyǔ.</u>

      Sometimes, the informant feels that this *ma'* is related to the COMPARISON sense,

and hence gives a Mandarin translation *yě*, meaning 'also.' But in many cases, the

---

[5] The speaker is of the Kaybaybaw family, which is of the same kin with the other two clans: Minrakes and Sayna'ase.

informant cannot explain the meaning of *ma'* and he leaves it unexplained, as in (110). For example (110), the informant reported that *isaa ma'* should be treated as a fixed expression, and his comment has been marked down in our note in NTU-Formosan.

There seems to be some differences between *isaa ma'* and *ma' isaa*. Unfortunately, due to the limit of corpus size, only two tokens of *isaa ma'* are available, and in direct elicitation, the informants often fail to provide instances for this highly discourse-oriented function. We are thus not capable of identifying the differences between *isaa ma'* and *ma' isaa*.[6]

In fact, *isaa* functions as a connective marker even without collocating with *ma'*.[7] Example (111) describes the chronological order of two activities: people return home and then they eat the offerings brought back from the tomb. The use of *isaa* in IU 98 is merely to coordinate two temporally-connected activities.

---
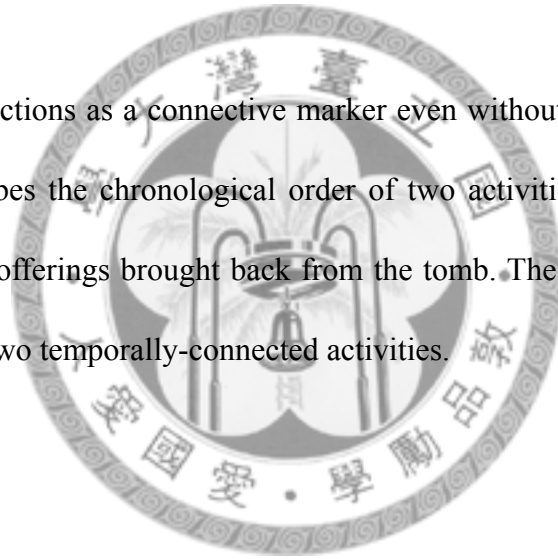
[6] It appears that *ma' isaa* connects an activity with a preceding prerequisite. The two activities have tighter causal relations, and are more likely conceived the elements of the same event. On the other hand, the *isaa ma'* weakly connects two events only with temporal sequential relation, and they are more likely to be conceived as two separate events that are linked for discoursal connective reasons. We nevertheless need more data to verify this observation.

[7] The lexical item *isaa* typically refers to a place far away comparable to English 'there,' or a specific situation or consequence, meaning 'that way.' It forms a contrastive set with *isahini* 'here; this way.' When meaning 'this way,' *isahini* signals the speaker's cognitive approximation of a situation or consequence, as opposed to *isaa* which carries an implication of the speaker's cognitive detachment from the situation or consequence. The connective function of *isaa* is likely to be a derived meaning or the result of grammaticalization.

(111) Holiday

97. K: .. *aras-en*   *si-patabir*
            take-PF   RF-offer
            dàiqù- PF   RF jìsì
     'Take them to be offered.'
     'Dongxi dai guoqu baibai,'
98. K: ... *lobih*   **isaa**
            AF.return   **then**
            AF.húilái   **ránhòu**
     'Then come home.'
     'Ranhou huilai jiali,'
99. C: .. *sia*
            3SG.NOM
            3SG. NOM
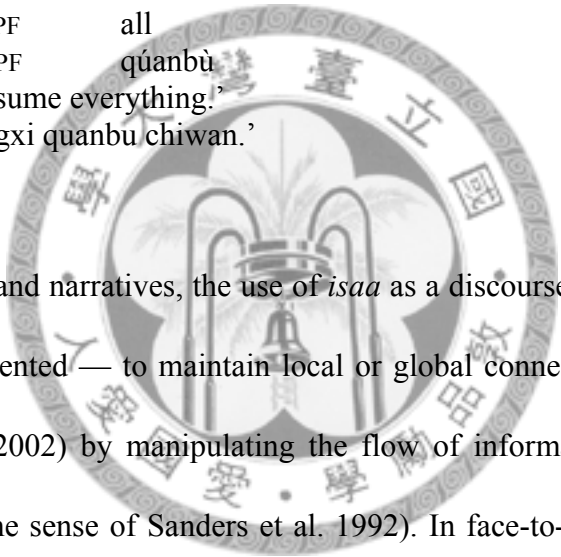
100. K: (0) *si'ael-en*   *saboeh*
            eat-PF   all
            chi-PF   qúanbù
     'Then they consume everything.'
     'Tamen ba dongxi quanbu chiwan.'

In conversation and narratives, the use of *isaa* as a discourse marker is common: It is highly discourse-oriented — to maintain local or global connectivity (in the sense of Traugott and Dasher 2002) by manipulating the flow of information in a smooth and coherent manner (in the sense of Sanders et al. 1992). In face-to-face interaction, it can have another function: to hold the speaker's floor. In (111), the speaker's use of *isaa* signals that there is more information to be said, and the speaker has not yet finished his turn of talking, this explains the latching in IU 100, because by the use of *isaa* to hold his own turn, the speaker K does not expect the interruption in IU 99.

As *isaa* and *ma'* show similar functions for discourse cohesion, they might have undergone the process of "routinization" or "idiomatization" (Hopper and Traugott's 1993). By these two terms Hopper and Traugott emphasize the process of meaning

negotiation in communication, and underscores conventionalization of meaning-structure pairing for the purpose of minimizing processing efforts. Recall (110); the informant reports that *isaa ma'* is a fixed expression. In some cases, *ma'* and *isaa* are very highly agglutinated that the combination of *ma' isaa* is interpreted by the informants as one lexical item. Example (112) is taken from NTU-Formosan.

(112) Life

190.	F:	...	*so==*	*rikrika-en*	*ila*	**ma'isaa**			
			COND	hot-PF	PFV	**like.that**			
			COND	rè-PF	PFV	**xiàngnàyàng**			

'If the weather is hot,'
'Rùo tianqì rè,'

191.	F:	*'ana*	*p<in>amowa'-an*	*tatimae'*	*ma*	*ayayo'*	*ila*	*saboeh*
		regardless	plant<PFV>-NMZ	vegetable	FUT	wilt	PFV	all
		wúlùn	zhòng<PFV>-NMZ	shucài	FUT	kuwěi	PFV	qúanbù

'the plants will all wilt.'
'zhíwù dou hùi kuwěi.'

In this respect, (CONNECTION) has become so fixed in terms of their syntagmatic selection and functions that we list it independently as one single entry of idiomatic sense.

6.4. The semantic network of *ma'*

By classifying the instances of *ma'* into senses, we found that the meanings of *ma'* are closely related to each other, and some of them are very difficult to be distinguished from one another. This section presents our analysis of how the senses of *ma'* are related to each other as a "family" (in the sense of family resemblance as suggested in Wittgenstein 1963).

162

6.4.1. Establishing the prototype

To construct a semantic network, the first step is to look for the prototypical meaning of *ma'* to which other senses are related. To look for the prototypical meaning of *ma'*, we follow the procedure proposed in Chapter 3. The prototypical meaning must have **Reciting Salience**. If we get uncertain answers by the test of reciting salience, we will then rely on **Naturalness of Predication** and **Socio-phenomenological Basicness** to help our judgment.

When the informants are asked "what is the meaning of *ma'*?" in de-contextualized situations, they intuitively give the counterparts *yě*. When they are asked to make an utterance containing *ma'*, they are prone to make instances of the [similarity] force, citing examples with different subjects doing the same action, such as (92).

Biq (1989) suggests that Chinese *yòu* and *yě* differ mainly in that the former compares different activities done by the same actor, whereas the latter compares different actors doing the same activity. Interestingly, Saisiyat although has the tendency to express the former by *nahaen* (which we have studied in Chapter 5) and the latter by *ma'*, the division is not an obligatory one. The lexical item can express the same activity by different actors, as in (92), as well as the same actor doing different activities as in (93). Nevertheless, when the informants are asked to make an utterance containing *ma'*, their answers has a salient tendency to contain similar expressions of different actors such as (92). The informants' unanimous preference might have shown that the prototypical meaning of a lexical item tends to be distinguished from the meanings of other lexical

items, e.g. *nahaen*. It is typically acknowledged that prototype should "have a higher degree of family resemblance (measured by sharing of features) to other category members and a lower degree of resemblance to members of other categories (Croft and Cruse (2004: 78)." Our observation supports this claim.

It is worth noting that although we assume the significance of "sense," which is a basic unit in almost all studies of lexical semantics, "sense" in fact has little cognitive salience to speakers. The most salient meaning of a lexical item that comes to the speakers' mind is typically at the level where meanings can be structurally and functionally distinguished, i.e. a "force" or even lower level of instantiation.[8]

6.4.2. Networking the senses

According to the types of construal change (Croft and Curse 2004) that we have reviewed in Section 3.4.2.2, we construct the semantic network of *ma'* as illustrated in Figure 6.1.[9]

---
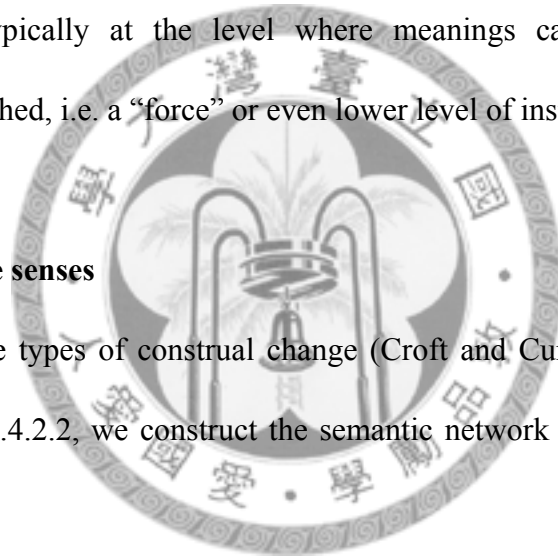
[8] This "low-level prototype" tendency is also attested in Chapter 4 and Chapter 5. The prototypical meaning of *nanaw* is a specific type of LIMITATION, that is, limitations that are specific to numbers and quantifiers. The prototypical meaning of *nahaen* is a specific type of REPETITION, i.e. repetition of an activity that has a pause in between.
[9] The extension is not necessarily from the most salient meaning, but has to be on the basis of the sense where the most salient meaning resides. This might have shown that extension usually happens when some degree of abstractness is generalized.

Figure 6.1. The semantic network of *ma'*
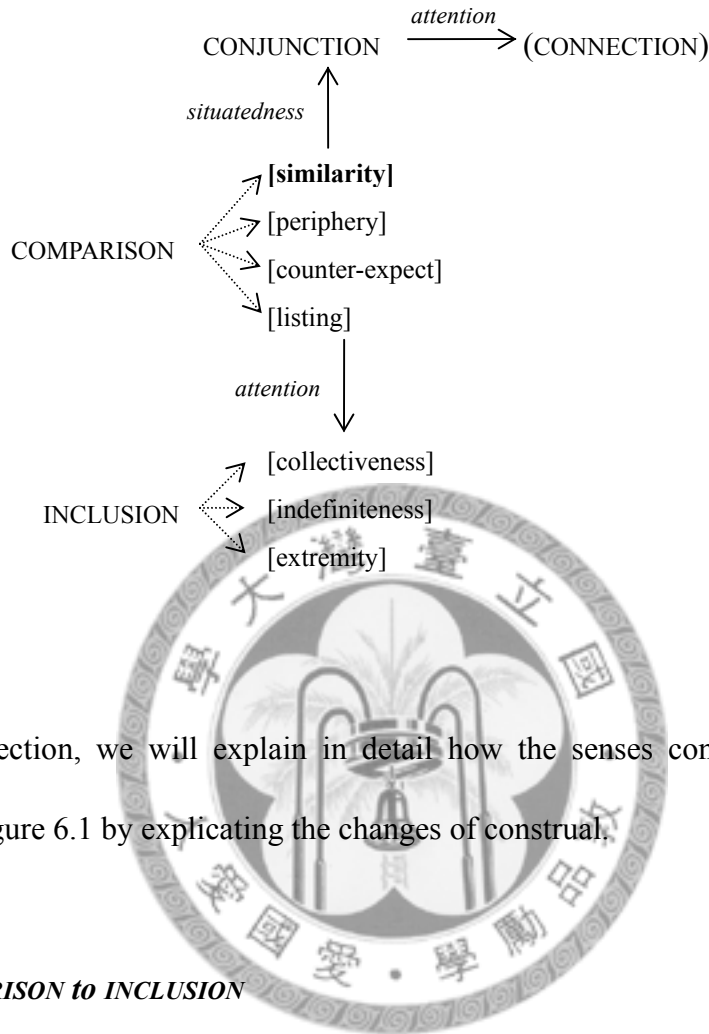
In this section, we will explain in detail how the senses construct the semantic network in Figure 6.1 by explicating the changes of construal.

From COMPARISON to INCLUSION

The relation between COMPARISON and INCLUSION can be explained by the type of construal known as **attention (salience)**. One force of COMPARISON, [listing], is directly responsible for the instantiation of INCLUSION.

The typical structure COMPARISON involves one expression that is compared with a preceding expression. The two expressions have apparent asymmetry in terms of information flow, and the preceding expression functions a "ground" which the following expression is compared to. The [listing] force has a significant change of this information

structure: Although it also involves comparison of expressions, every activity carries a *ma'* marker, iconically showing that the expressions being compared are nearly equal in terms of their grounding status. The discrepancy between the expressions is no longer highlighted, which serves as a springboard for speakers to construe an "entirety" view that is taken by the INCLUSION sense.

This shift involves a change of viewing scope. When one takes a narrow scope such as that of COMPARISON, he is likely to find out the differences between the individuals, and to view the individuals as different entities. INCLUSION on the other hand takes a wider, holistic viewing scope, shifting the focus to the entirety rather than the componential individuals. The shift manifests the kind of construal change that is called **attention (salience)** in Croft and Cruse (2004). The viewing scope has been adjusted from narrow to wide, and the focus has been shifted from individuals to the entirely.

From COMPARISON to CONJUNCTION

We claim that COMPARISON and CONJUNCTION are closely related. Before we explain their functional relation, it needs to be pointed out that the grammatical status of *ma'* in COMPARISON is different from that in CONJUNCTION. In its COMPARISON sense, *ma'* behaves as an adverbial, whereas in its CONJUNCTION sense, *ma'* behaves as a conjunction. There is in fact a fine distinction between adverbial and conjunction, as we will explain in this section. In studies of Mandarin, there have also been similar debates on the grammatical status of its counterpart *yě* (cf. Tsai 2006).

As shown in (113) and (114), the COMPARISON and CONJUNCTION senses of *ma'* both

involve two expressions A and B.

(113) COMPARISON

    [*palono'*   *toliyaep*]$_A$ [*kahoey*   ***ma'***   *toliyaep*]$_B$
      ship   AF.float   wood   **MA'**   AF.float
      chúan   AF.piao   shùmù   **MA'**   AF.piao
    'The ship floated (on water), and the wood also floated (on water).'
    'Chúan piao (zài shǔi shàng), shùmù yě' piao (zài shǔi shàng).'

(114) CONJUNCTION

    [*So:*   *'ino'an*]$_A$ ***ma'***   [*yao*   *'am*   *rima'*   *ila*   ]$_B$
      COND   when   **MA'**   1SG.NOM   FUT   AF.go   PFV
      COND   héshí   **MA'**   1SG.NOM   FUT   AF.qù   PFV
    'Next time, I will go.'
    'Xìacì wǒ yào qù.'
    (Literally: If there is a chance next time, I will go.)

We argue that the act of "comparison" is the functioning of "grounding." The

COMPARISON sense of *ma'* aligns a novel experience to an established, activated one. As

sketched in Figure 6.2, the prior expression (A) is the "ground" which the "figure" (later

mentioned B) is compared with.

Figure 6.2. The grounding configuration of COMPARISON and CONJUNCTION
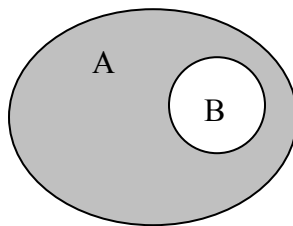
"Grounding" is deemed cognitively relevant to the formation of complex sentences (Talmy 1978; Croft 2001). The event in the subordinate clause is a cause or precondition for the event in the main clause, and is thus serves as a basis or "ground" by which a figure is contrasted to. In conditional sentences, the protasis (if-clause) can be considered the background on which the apodosis (then-clause) is presented (cf. Fauconnier 1985), exhibiting a ground-figure relation in the sense that the condition sets up a frame for the consequence. In this aspect, the same sketch Figure 6.2 also illustrates the construal of CONJUNCTION.

Compared with COMPARISON, the scope of CONJUNCTION goes beyond local elements, and extends to larger units such as the entire clauses. In addition, CONJUNCTION is more textually-oriented, which confirms to the type of construal change known as **situatedness (perspective)**. Due to the speaker's awareness of his role in a discourse, he intends to manage the structure of his discourse and to mind the sequence of presenting information.

From CONJUNCTION to (CONNECTION)

In Saisiyat, there is in fact no sharp distinction between coordination and subordination. For example, (115) can easily yield two readings that are not contrastive.

(115) *patonay rirang-en awka' ila*
      iron   rust-PF   disappear PFV
      Subordination reading: Iron will be eaten away if it gathers rust.
      Coordination reading: Iron gathers rust and is eaten away.

Given the blurred distinction between subordination and coordination, the relation between CONJUNCTION and (CONNECTION) is evident. Croft (2001) argues that the relation between subordination and conjunction is an issue of configuration change: from "figure-ground" configuration to "complex figure" arrangement. This can be sketched as in Figure 6.3.

Figure 6.3. The grounding configuration of CONJUNCTION and (CONNECTION)
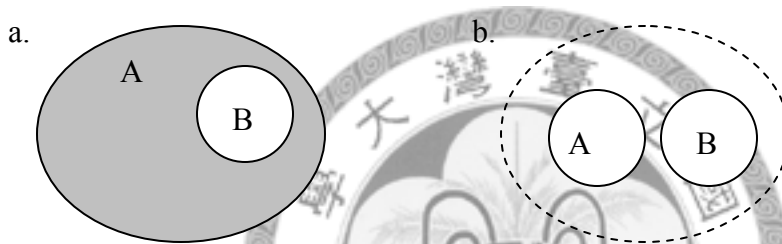
For CONJUNCTION, as we have explained, the first expression serves as the "ground" to which the later-mentioned expression, i.e., the figure, is related, as in Figure 6.3 (a). On the other hand, the clauses or passages conjoined by (CONNECTION) are of a 'sequential' relation that is looser than a temporal or causal one. Although the coordinated passages appear to be quite free, not any two passages can be naturally conjoined; they should involve the speaker "conceptualizing the paired elements as a whole unit having something in common" (Croft 2001: 336). We can find a common denominator between A and B: they are sequential expressions in the same episode. The presentation of two conjoined elements with any conceivable cognitive unity, as sketched in Figure 6.3 (b), is

called "complex figure" in Croft (2001).[10] The change of viewing arrangement and figure-ground alignment manifests the kind of construal change that is related to our **attention (salience)** as defined in Croft and Cruse (2004).

6.5. Semantic meaning and constructional patterns

In the course of investigation, we find that the meanings of *ma'* is tightly integrated with the structure it occurs in, constituting fixed structural patterns with it syntagmatic co-occurring elements. In addition to "fixation," a high degree of obligatorification (in the sense of Lehmann 1985) is attested: the choice of lexical items and structures are systematically constrained, and the use of *ma'* with some specific lexical items to denote an intended meaning is obligatory. The fact that *ma'* is structurally bounded in fixed forms might have shown that the semantic contents of *ma'* is more attenuated and abstract; many clues for meaning specification have been allotted to its structural patterns. In this section, we will show the structural patterning of *ma'*, with an emphasis on the communicative motivation for conventionalizing specific linguistic patterns.

6.5.1. Meaning of *ma'* in constructions

In our analysis, we find that *ma'* is highly dependent upon its collocating structure. Table 6.1 summarizes the relation between the meanings and structural patterns of *ma'* based on all the tokens available to us.

---

[10] The (CONNECTION) idiom *ma' isaa* or *isaa ma'* nevertheless connects passages or events (e.g. He woke up and saw his wife.) but not nominal phrases (e.g. husband and wife).

Table 6.1. Meanings and constructional patterns of *ma'*

Meanings	Structures	
COMPARISON		
[similarity]	NP$_1$ VP$_1$, NP$_2$ *ma'* VP$_1$	
[listing]	NP$_1$ *ma'* VP$_1$, NP$_2$ *ma'* VP$_1$	
[periphery]	(NP$_1$ VP$_1$), NP$_2$ *ma'* VP$_1$	(same as [similarity])
[counter-expect]	(NP$_1$ VP$_1$), NP$_1$ *ma'* VP$_1$	
INCLUSION		
[indefiniteness]	*ana'* INDEFINITE_PRO *ma'* VP	
[extremity]	*ana'* NP/VP *ma'* VP	
[collectiveness]	*sasa'ih/saboeh* ma' VP	
CONJUNCTION	CL$_1$ *ma'* CL$_2$	
(CONNECTION)	CL$_1$ *ma'isaa* CL$_2$	
	CL$_1$ *isaa ma'* CL$_2$	

The division of the structural patterns is not clear-cut. For example, strictly speaking, [periphery] shares the same structural pattern with [similarity]. However, from a discourse-interactional perspective, the "grounds" of comparison of [periphery] tends to be an implicit, unexpressed standard shared by the community or a presupposed expectation. In view of interactional grammar (e.g. Ford et al. 2003), [periphery] can be said to be structurally distinguished from [similarity] — its social-discourse function of "making evaluation" might have accelerated dropping (or de-emphasis) of the grounds of comparison, making it structurally distinct from [similarity]. Communicative goals, such as "making judgment," often motivate the emergence of a new meaning, and possibly a new structure. Some instances of [periphery] such as *ma' kayzaeh* 'also good' constantly activates a 'not very good' interpretation in our elicitation.

6.5.2. Semanticization, contextualization, and structural patterning

Our treatment of [periphery] is a reflection of our attitude toward the distinction between semantics and pragmatics: We deem the distinction between semantics and pragmatics as a necessary division when it comes to applications, yet we believe they are closely intertwined phylogeneticallly. Semantics is concerned with meanings that are relatively stable, independent from the context. Pragmatics, on the other hand, is concerned with subjective beliefs and inferences made by participants for communicative purposes. However, according to Hopper and Traugott (1993), at the early stages of grammaticalization, conversational implicatures often become part of the semantic polysemy of a form, a process known as "semanticization."

The consolidation of meaning is often accompanied by condensation of structures. To facilitate comprehension of a linguistic form, "conventionalization" is the feature that the speaker must follow in order to "ritualize" the conversation ceremony (Goffman 1981). Innovative languages are in use only when the current linguistic forms fail to convey the speaker's information or intention. Even when innovations are necessary, the speaker has to take into consideration the inferencing ability of his audience, "inviting" the hearer to make appropriate interpretation in the situated contexts (Hopper and Traugott 1993; Traugott and Dasher 2002). In fact, resent studies of construction (pairing of form and meaning) begin to pay attention to the pragmatic aspects. For example, Blank (2003) suggests that the formation of a construction involves three stages: 1) idiosyncratic innovation of the speaker, 2) usualization of this idiosyncratic innovation,

172

and 3) lexicalization in the language. Fritz (1998, cited from Gyori 2002) also includes social-cultural elements in semantic theorization. The process of semanticization is suggested to have involved routinization, standardization, and conventionalization.

In this view, "semanticization" is to a large extent the product of contextualization, and formally, the semanticization of a linguistic form is often accompanied by its integral with its linguistic contexts, i.e. structural obligatorification. Contexts are the resources of collocational patterns: implicatures embedded in particular linguistic/situational contexts often become stored meanings, and the co-occurrence patterns become frozen expressions to which the meanings are attached, a process that we shall call "contextualization" in the present study. And in a broad sense, compounding, lexicalization, idioms, selective restriction, and schematic structural patterning should be regarded types of construction as suggested in Croft and Cruse (2004).

The formation of ritualized structural patterns, including strictly-defined "constructions" (in the sense of Fillmore et al. 1988 and Goldberg 1995, 2006) and broadly-defined "patterning of verbal interaction" (in the sense of Ford et al. 2003), is to a large extent motivated by communicative demands, i.e. to decrease the hearer's processing efforts (Hopper and Traugott 1993).

6.6. Direct translation and semantic glossing

Given the different meanings of the lexical item *ma'*, the interpretation of this lexical item is likely to be rich. In this section, after identifying the senses of *ma'* and

constructing the semantic network, we come back to look at the use of its direct translations.

6.6.1. Direct translation of a polyseme

When translating the utterances containing *ma'*, the informants use *yě* widely cross-textually as the direct translation of the lexical item *ma'*. The subtle functional divisions of *ma'* does not yield many different direct translations as we have expected. For (CONNECTION), translations such as *ránhòu* 'then', *jiù* 'then', *nàyàng* 'that way', and *nà* 'that' are used in addition to *yě*. When translating the instances of (CONNECTION), the informants try to express it with the best Mandarin equivalent. However, in Mandarin, there is no appropriate comparable structure to (CONNECTION), which explains the diverse Mandarin translations for (CONNECTION).

It is worth noted that DM (discourse marker) and FIL (pause filler) are frequently used as the gloss of *ma'*. The meaning of *ma'* is highly abstract and discourse-oriented, and the informants feel difficult to find a translation for it. Sometimes the informants report that *ma'* is a pause for information flow, and we give it a DM gloss. Other times the informants report that *ma'* is to fill a pause, and we give it a FIL gloss. Occasionally, the informants report that *ma'* has no meaning at all, and we give it a ?? gloss.

If the textual function (CONNECTION) is neglected, the translation of *ma'* provided by the informant is consistently *yě*. In the next section, we will show that some constructions of Saisiyat *ma'* should rather be translated into *yòu, hái,* and *dou* rather than *yě*. However,

Saisiyat informants do not use them as a translation of the single lexical item *ma'*. This might have been the effect of a higher degree of cross-linguistic predictability.

6.6.2. The semantic partitioning of *ma'* in Mandarin

For the functions of Saisiyat *ma'*, there are at least three other Mandarin lexical items, *yòu*, *hái*, and *dou* that compete with *yě*. Based on previous studies in Mandarin (Li and Thompson 1981, Biq 1989, Yeh 1998, Liu 2001), their way of partitioning the functions of Saisiyat *ma'* can be summarized in Table 6.2.

Table 6.2. Meanings of *ma'* partitioned in Mandarin

Meanings	*Mandarin pragmatic translations*
COMPARISON	
[similarity]	*yě > yòu*
[listing]	*yě*
[counter-expect]	*hái*
[periphery]	*yě*
INCLUSION	
[indefiniteness]	affirmative : *dou* negative : *yě/dou*
[extremity]	*dou/ yě*
[collectiveness]	*dou*
(CONNECTION)	no consistent equivalent

Among the Mandarin equivalents, *yòu* differs from *yě* in that the former compares different activities carried out by the same actor, whereas the latter compares different actors doing the same activity (Biq 1989). In fact, when two activities share the same subject, usually both *yòu* and *yě* are acceptable, but *yòu* focuses on chronological order

175

whereas *yě* focuses on addition of activities. Their difference can be manifested by the

contrast between (116) and (117).

(116) Activities sharing the same predicate
    ta   *hùi ,*   *wǒ*   **yě**   *hùi*
    3SG   able   1SG   **also**   able
    'He can do it, and I can do it, too.'

(117) Activities sharing the same subject
    a. *ta*   *gǎnmào*   **yòu**   *fashao*
        3SG   have.flu   **again**   have.fever
        'He had flu, and then had fever.'
    b. *ta*   *gǎnmào*   **yě**   *fashao*
        3SG   have.flu   **also**   have.fever
        'He had flu, and also had fever.'

    The [counter-expect] force of *ma'* indicates that an activity persist regardless of an

unfavorable situation. In Mandarin, one should use *hái* to express this scene with roughly

the same semantic and pragmatic essence. This can be shown in (118).

(118) Counter-expect in Mandarin
    zài   *xìayǔ*   *ta*   **hái**   *yào*   *húi*   *jia*
    PROG   rain   3SG   **still**   want   return   home
    'It was raining but he insisted going home.'

    Interestingly, most forces of the INCLUSION sense is covered by Mandarin *dou*, but

the informants never use *dou* as the direct translation of the lexical item *ma'*. For

[collectiveness], *dou* is the only choice in Mandarin, as shown in (119).   For [extremity],

the Mandarin speakers use either *yě* or *dou*, as shown in (120).

(119) Collective in Mandarin
    a. ***měi***   *gè*   *rén*   ***dou/*yě***   *zhidào*   *jiankang*   *dè*   *kěgùi*

176

every	CLS	person	**all/also**	know	health	NMZ	value

'Everyone knows the value of health.'

b. ***sǔoyǒu***   *rén*   ***dou/*yě***   *zhidào*   *jìankang*   *dė*   *kěgùi*

   all   person   **all/also**   know   health   NMZ   value

   'All people knows the value of health.'

(120) Extremity in Mandarin

   (lían)   *cíangdào*   ***yě/dou***   *hùi*   *fa*   *shàn*   *xin*

   even   robber   **also/all**   able   show   kind   heart

   'Even robbers may act kindly (sometimes).'

For [indefiniteness], the affirmative structure in Mandarin has to be expressed with *dou*, and *yě* is not acceptable. In negative structures, however, either *dou* or *yě* is acceptable, as shown in (121).

(121) a. Indefiniteness (affirmative) in Mandarin

   ta   *shémė*   ***dou/?yě***   *hùi*

   3SG   what   **all/also**   able

   'He can do everything.'

   b. Indefiniteness (negative) in Mandarin

   ta   *shémė*   ***yě/dou***   *bú*   *hùi*

   3SG   what   **also/all**   NEG   able

   'He can't do anything.'

The affirmative-negative division is not found in Saisiyat. Both affirmative and negative constructions of INCLUSION are expressed with *ma'* in Saisiyat.

6.6.3. Cross-linguistic semantic predictability and direct translation

The relation between the meanings of *ma'* is very tight and has a high cross-linguistic predictability. We have shown that Mandarin *yě* has many functional extensions that are similar to that of Saisiyat *ma'* except for some forces of the INCLUSION

sense. We find that in another language, Japanese, which has little genetic relation with

Saisiyat, there is a counterpart *mo* that covers nearly all the functions of Saisiyat *ma'*.[11]

Examples in (119) shows that Japanese *mo* covers nearly all the functions of the

COMPARISON sense.

(122) Japanese (counterpart examples of the COMPARISON sense of Saisiyat *ma'*)
    a. similarity
        watashi   *wa*   *gakusei*   *desu*   *anata*   ***mo***   *gakusei*   *desu*
        1SG   SUB   student   DECL   2SG   **also**   student   DECL
        'I am a student, and so are you.'
    b. listing
        mikan   ***mo***   *linggo*   ***mo***   *banana*   ***mo***   *daisuki*   *desu*
        orange   **also**   apple   **also**   banana   **also**   like   DECL
        '(I) like oranges, apples, and bananas.'
    c. periphery
        saru   ***mo***   ki   kara   ochiru
        monkey   **also**   wood   LOC   fall
        'Even monkeys fall from trees occasionally.' ('Even an expert makes mistakes.')
    d. counter-expect
        kasoku   ga   iru   no   de   shigoto o   yametaku   **mo**   yamerarenai
        family   TOP   exist   NMZ   CONJ   job   OBJ   stop.want   **also**   stop.NEG
        'As I have a family to keep, I must continue this job even if I don't want to.'

Examples in (123) further show that Japanese *mo* also covers the functions of

INCLUSION. Although Japanese *mo* does not have similar structures like (CONNECTION), all

other functions of Saisiyat are highly predictable in Japanese.

(123) Japanese (counterpart examples of INCLUSION)
    a. indefiniteness

---

[11] Although some forces are expressed with *dou* and *hái* in Mandarin Chinese, we find that in Southern-Min dialect of Chinese, all the forces of COMPARISON and INCLUSION are expressed with the same linguistic item, *mā* 'also.' In other words, the cross-linguistic predictability of the semantic network of Saisiyat *ma'* is also attested in Southern-Min.

dare	**mo**	ga	kare	no	sholi	o	sinziteita
who	**also**	TOP	3SG	NMZ	victory	OBJ	believe. PST

'Everybody believed he would win.'

b. extremity

kono	mura	no	mono	wa	**hitori**	*mo*	shiranai
this	village	NMZ	person	SUB	**one.person**	**also**	know.NEG

'I don't know a single one of the villagers.'

c. collectiveness

sannin	**mo**	no	hito	ga	onaji	machigai	o	shita
three.person	**also**	NMZ	person	TOP	same	error	OBJ	do.PST

'All ten people made the same mistake.'

In the case of *ma'*, direct translations offered by the informants are highly adherent to the semantic core of the lexical item, less influenced by contextual considerations. The leap between the semantic core and the pragmatic translation can be easily bridged, and the informants feel less uncomfortable using the semantic core as the direct translations.
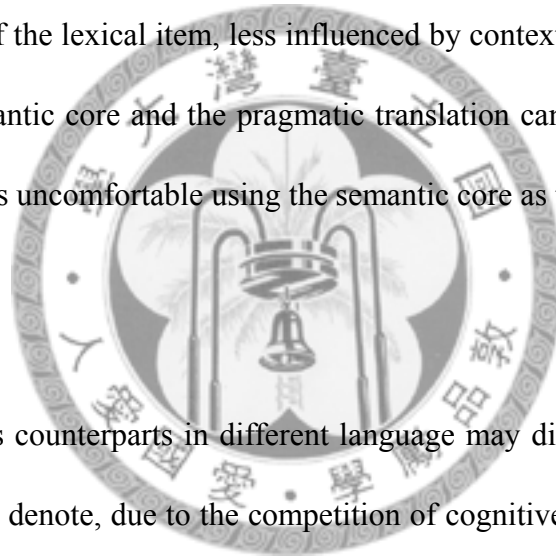
6.7. Summary

Forms deemed as counterparts in different language may differ slightly in terms of the meanings they can denote, due to the competition of cognitive-pragmatic motivations of semantic change in different languages (Croft 2001). There are, occasionally, cases that are found to have higher cross-linguistic predictability. In this chapter, a lexical item *ma'* may fall into such a category. It has many functions, but most of its functional extensions are found in at least two languages. The semantic network of *ma'* exhibits a high degree of cross-linguistic predictability. In addition, it also shows high degree of abstractness. Saisiyat *ma'* is highly dependent upon its collocating linguistic elements, forming constructional patterns that show the feature of fixation and obligatorification (in

the sense of Lehmann 1985). Due to its high degree of abstraction and cross-linguistic predictability, Chinese speakers can easily find similar development in their own language, and the relations between the multiple functions of *ma'* is more easily conceivable. Direct translations from Saisiyat to Mandarin offered by the Saisiyat informants are relatively consistent.

Chapter 7   Towards an Explanation: Meaning and Perspectivization

7.1. Preliminaries

Many studies of polysemy center on the meaning extension of lexical items in one single language. The rise of a cross-linguistic approach to polysemy is potential to provide another angle to view the problem of one-to-many mappings between form and meaning.

One aim of the present study is to examine some theoretical assumptions of polysemy via cross-linguistic comparisons. In the progress of our investigation, we find that the meaning of the three delimited lexical items is enriched or impoverished with the speaker's change of construal. Tendencies of subjectification (Langacker 1987a) and intersubjectification (Traugott 2003) are attested. In this chapter, we will discuss the findings of our case analyses with special emphasis on their manifestation of perspectivization, and. take a cross-linguistic view to examine the notion of perspectivization

The organization of this chapter is as follows. In Section 7.2, we briefly sketch the theories of perspectivization and the relation between perspective-taking and meaning extension. Some of the ideas have also been reviewed in Chapter 2. In Section 7.3, we take an intra-language point of view to examine the changes of perspective of our three case studies. In Section 7.4, we then take a cross-linguistic point of view, comparing the semantic partitioning of the three Saisiyat lexical items with their Mandarin counterparts.

Section 7.5 is devoted to discussions on language universality/relativity and the relation between language and thought. A comparative approach to languages inevitably invites thoughts about language universality versus relativity, which in turn brings about some implications to the relation between language and thought.

7.2. Perspectivization and semantic change

Theories of perspectivization use a wide variety of terms with different emphases to refer to the phenomena of speaker's viewpoint adjustment. In this study, following Graumann and Kallmeyer (2002), we use the term "perspectivization" in a broad definition to refer to the grounds taken by the speaker to view a particular scene: In the process of verbalization, the speaker could perceive a stimulus from different angles, and different aspects of the same scene will come into view. And the perception is not limited to sensory stimuli, but also includes emotional and judgmental reflections. In this broad definition, all the four types of construal change in Croft and Cruse (2004): attention, judgment, situatedness, and constitution, should be included.

The study of perspectivization is most widely known to be associated with Langacker's Cognitive Grammar (1987a, 1991), particularly the notion known as "subjectification." Langacker suggests that semantic change, focusing mainly on synchronic data, tends to become grounded by the speaker's subjective point of view, where subjectification refers to the change from an "objective" Optimal Viewing Arrangement (OVA) (with a noticeable conceptual distance between the conceptualizer

and the conceptualized) to a "subjective" Egocentric Viewing Arrangement (EVA) (when the conceptualizer becomes part of the conceptualized scene). If subjectification takes place, the semantic content of a lexical item is enriched with explicit strengthening of the speaker's commitment of a proposition. In various studies, subjectification is found to be relevant to semantic extension and grammaticalization (Bybee et al. 1994; Heine et al. 1991; Traugott and Dasher 2002; Langacker 1999, just to cite a few).

Langacker's theory of subjectification is criticized by Tarugott and Dasher (2002) for its not being able to account for interactional considerations of language use. In conversation, the hearer plays an important role which the speaker's attention should turn to in order to make a successful speech interaction, but the role of the hearer is, nevertheless, deemphasized in Langacker's framework.[1] Traugott and Dasher (2002) suggest that in addition to objectification and subjectification, a third possible way of perspectivization, i.e. intersubjectification, should be singled out to gracefully account for the process of semantic extension. There are, therefore, at least three possible ways of perspective-taking: objectification, subjectification, and inter-subjectification, with features of each listed below (Traugott and Dasher 2002: 22-23):

---

[1] The notion of "ground" proposed by Langacker (1985) subsumes both the speakers' and hearer's stance, and the different roles of speaker and hearer have not been addressed.

(124) **a. Objective expression:**

     i)   declarative, minimally-marked with regard to modality,

     ii) all participants in an event structure are expressed in surface structure,

     iii) lexical items are minimally concerned with the interlocutor's perspective (i.e. minimally deictic),

     iv) the Quality-heuristic predominates, i.e. contexts for meanings are provided so that interpretation is strongly determined, and what is not said is implied not to be the case.

b. Subjective expression:

     i)   overt spatial, and temporal deixis,

     ii) explicit marker of speaker's attitude to what is said, including epistemic attitude to the proposition,

     iii) explicit markers of speaker's attitude to the relationship between what precedes and what follows, i.e. the discourse structure; many aspects of discourse deixis are included here,

     iv) The Relevance-heuristic predominates.

c. Intersubjective expression:

     i)   overt social deixis,

     ii) explicit markers of speaker's attention to addressee, e.g. hedges, politeness markers, and honorific titles,

     iii) the Relevance-heuristic predominates, i.e. what is said implies more is meant.

According to Traugott (2003), there is a tendency for non- or less subjective language use to become more subjective and for subjective ones to become intersubjective. The tendency can be written as in (125):

(125) less subjective → subjective → intersubjective

Traugott (2003) argues that subjectification is prerequisite to intersubjectification: In a broad definition, intersubjectification can be deemed a subtype of subjectification because speaker's attitude toward the addressee is a function of the perspective of the speaker.

In the following sections, we will examine the theory of perspectivization from intra-language and inter-language aspects by the findings of the three case studies. The examination may help us understand the way of conceptualization from different points of view.

7.3. Intra-lingual evidence: From less subjective to (inter-)subjective

As we have mentioned, a broad definition of perspectivization should include attention and perspective. In our three case studies, attention and perspective are the major driving forces of the semantic extension. The rich manifestation of perspectivization in the three cases can help us understand the phenomena of perspectivization more clearly. Following the features summarized in (124), we focus on the elements that are marked when different perspectives are taken, which can be summarized in Table 7.1.

Table 7.1. Marked elements for different perspectives

	objective (less-subjective)	*subjective*	*intersubjective*
marked elements	minimally-marked	speaker's attitude to the proposition	speaker's attention to addressee
		speaker's attitude to the discourse structure	

According to Table 7.1, we re-examine the semantic extensions of the three delimited lexical items by classifying the senses into three stages. The results are shown in Table 7.2.

Table 7.2. (Inter-)subjectification and semantic change[2]

	less-subjective	*subjective*	*intersubjective*
nanaw		LIMITATION → (NEG_EXTREME) ↓ CONTINUATION	
nahaen	[recurrence] → ↓ ↘ [addition]	[continuation] SUCCESSION ⟶	PRECEDENCE
ma'	[similarity] → ↗ ↘	(CONNECTION) ↑ CONJUNCTION [counter-expect] [listing] ↓ [indefiniteness] [collectiveness] [extremity]	[periphery]

---

[2] In this table, we include senses, forces, and idiomatic senses of each lexical item. The convention of notation is: SENSE, (IDIOMATIC_SENSE), and [force].

This table manifests two facts that have been suggested in Traugott (2003) and Traugott and Dasher (2002). First, the change is from a less subjective expression to more (inter-)subjective expressions. For *nanaw*, the meanings are always subjective. Its prototypical meaning 'only' involves subjective expectation of a presupposed standard, and denotes that a quantity is less than the expectation. Extensions from the prototypical meaning involves higher subjective reasoning based on shared knowledge of the society, perceived evidences, or personal volitions. However, the semantic development of *nanaw* does not extend to the stage intersubjectification; its meanings are limited to the speaker's attention to his own epistemic stance. In the other two cases, the development is clearly right-ward and uni-directional.

The other fact attested in this table is that intersubjectification seems to be based on subjectification. The SUCCESSION sense of *nahaen* is a subjective generalization from strict activity repetition to loosely-connected action sequence, and it instantiates the PRECEDENCE sense which is entering the stage of intersubjectivity. By tactfully employing the sequential event structure, the speaker mitigates possible face-threats in conversation by making his invitation or suggestion an insignificant prelude to a presupposed upcoming activity.

Similarly, for the case of *ma'*, its [counter-expect] force is a subjective construing of the [similarity] force: The speaker assumes the typical situations for a specific activity to take place, but now we come across an activity that is carried out regardless of a non-favorable condition. The non-typical resemblance of [counter-expect] has accounted

for the emergence of [periphery] force by which a situation only passably meets a standard. The [periphery] force is often employed to denote the downgrading of a quality, particularly for politeness considerations, making it an intersubjective use of *ma'*.

In sum, the theory of intersubjectification (Traugott 2003; Traugott and Dasher 2002) can gracefully explain the family resemblance of the three lexical items under our investigation. In the process of semantic extension, the meaning of a lexical item is prone to become more and more engrossed with the speaker's attention to his own epistemic stance, and to the viewpoint of the hearer.

7.4. Inter-lingual evidence: Social cognition

The story of perspectivization does not end at the stage of speaker-hearer intersubjectification. The notion of intersubjectification will have to be brought to a higher social level in order to account for language-specificity of conceptual categorization. In this section, we will first present the structuring of concepts, semantics, and syntax in view of Croft's (2001) Radical Construction Grammar (RCG). The process of conceptual categorization involves verbalization of concepts via semantic partitioning and syntactic representation. From an inter-language perspective, we find that concepts are prone to be partitioned differently across languages, and the speaker in fact has to conform to the socially-agreed way of verbalization, with very limited subjective choice of construals.

7.4.1. A view from Radical Construction Grammar (RCG)

Verbalization is a process of categorization and meta-representation (Levinson 1997). Each language represents a specific way of viewing the world: Perceived experience is categorized into conceptual groups that make sense to the speakers, and the conceptual groups are labeled with linguistic symbols. In view of this, describing one language (L1) with another language (L2) involves possible distortion of L1's categorization system in order to be fit into the system of L2.

Croft's (2001) Radical Construction Grammar (RCG) best captures the essence of language-specific conceptual categorization. In Chapter 2, we have briefly introduced the tenets of RCG, which employs the notion of "semantic map" to show the functional distribution of particular contrastive linguistic forms, and also to look for cross-linguistic common tendencies of functional extension, in an attempt to sort out the cognitive mechanisms underlying human conceptualization.

RCG makes another important stipulation: the distinction of conceptual, semantic, and syntactic structures. In a given conceptual space, relevant functions are assumed to be connected, but it allows alternative semantic conceptualizations and syntactic representations. The leveling of the three structures can be illustrated in Figure 7.1.

Figure 7.1. The relation between semantics, syntax and conceptualization (Croft 2001)

Syntactic structure

⇓ ⇑

Semantic structure

⇓ ⇑

Conceptual structure
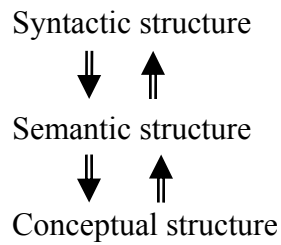
Under competitions of different motivations, the choice of a specific expression is to some extent "arbitrary" and subject to change. In different languages, speakers may highlight different elements of this conceptual space, resulting in cross-linguistic typological variations. The mapping between semantic structure (meaning) and syntactic structure (form) follows the spirit of "construction grammars" (Fillmore et al. 1988; Lakoff 1987; Langacker 1987a; Goldberg 1995).

Still, syntactic structures conventionalized in a language can affect its semantic structure which in turn links to its conceptual structure. This explains the conventionalized ways of conceptualization in a speech community, and the interactions among different levels of structures are thus deemed bi-directional. In Figure 7.1, the two-way influences between levels of structure are represented by bi-directional arrows.

In Croft's RCG framework, "semantic map" can clearly show the relation between conceptualization and verbalization in different languages. By looking at the diverged syntactic devices, linguists are likely to see the semantic subdivisions of a conceptual space, in an attempt to construct a universal semantic map on which human thinking is based. This is different from traditional Construction Grammar (especially Goldberg

1995), which centers on how a structural meaning of a linguistic structure determines the interpretation of an utterance. This claim of RCG facilitates lexical studies that are not necessarily construction-based.[3]

7.4.2. Language-specificity of semantic partitioning

Following RCG, there is a process of conceptualization between the conceptual structure and the semantic structure. The way of partitioning a conceptual space is usually language specific, and mismatches of semantic partitioning is clearly reflected in the process of direct translation.

Nanaw

As we have shown in Chapter 4, the inconsistent Mandarin glosses of Saisiyat *nanaw* in NTU-Formosan is resulted from the mismatch of semantic partitioning in Saisiyat and Mandarin. Related meanings coded by the same lexical item *nanaw* are glossed as many seemingly unrelated Mandarin words, because the meanings of NANAW that is connected as a conceptual space in Saisiyat is partitioned into more than one category in Mandarin, coded by different linguistic forms.

---

[3] Case studies of RCG are primarily construction-based, such as passives, complementation, case marking, and so on. However, its methodology is more flexible, making it possible to be utilized for small-scale investigations that are not construction-based.

Among the senses of *nanaw*, LIMITATION is expressed in Mandarin with *zhǐ* 'only,' as shown in (126). When a quantity is involved, Mandarin speakers employ *zhǐ.yǒu* 'only.exist' and *cái* 'just; only,' as shown in (127).

(126)   *ta*   **zhǐ**   *chi*   *yú*
           3SG   **only**   eat   fish
           'He eats only fish.'

(127)   *wǒmėn*   **zhǐyǒu/cái**   *san*   *gė*   *rén*
           1PL   **only.exist/only**   three   CLF   human
           'There are only three of us.'
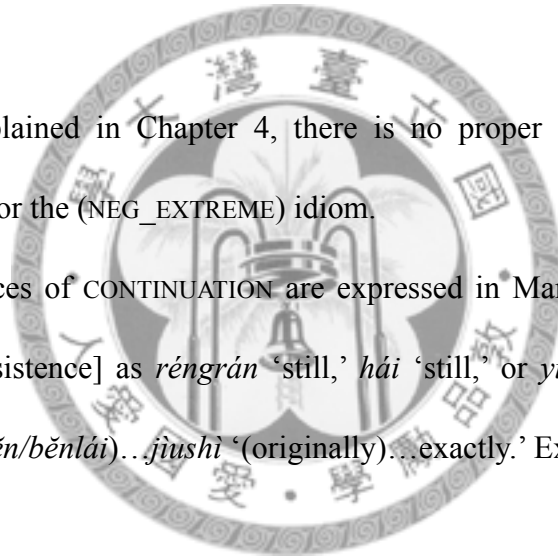
As we have explained in Chapter 4, there is no proper semantic and syntactic Mandarin equivalent for the (NEG_EXTREME) idiom.

And the two forces of CONTINUATION are expressed in Mandarin by two different linguistic forms: [persistence] as *réngrán* 'still,' *hái* 'still,' or *yizhí* 'continuously,' and [inherence] as (*yúanběn/běnlái*)…*jiushì* '(originally)…exactly.' Examples are provided in (128) and (129).

(128) Persistence of an activity
        yǔ   **yizhí/hái/réngrán**   *bù*   *tíng*
        rain   **continuously/still/still**   NEG   stop
        'The rain still did not stop.'

(129) Inherence of a natural disposition
        xǔe   **yúanběn / běnlái**   *jiushì*   *hùi*   *rónghùa*
        snow   **originally/originally**   exactly   able   melt
        'Snow melts by nature.'

The way of partitioning NANAW in Mandarin can be summarized as in (130).

(130) Semantic partitioning of Saisiyat NANAW in Mandarin

    (A) In Saisiyat, LIMITATION, CONTINUATION, and (NEG_EXTREME) are conceptualized as in the same linguistic category, labeled by *nanaw*.

    (B) LIMITATION is coded by *zhǐ(yǒu)* 'only' or *cái*.

    (C) CONTINUATION is partitioned in Mandarin into "persistence of an activity" and "inherence of a specific disposition," respectively coded by two sets of morpho-syntactically unrelated linguistic forms.

    (D) There is no consistent way of expressing (NEG_EXTREME) in Mandarin.

Based on (130), we see there are mismatches of semantic partitioning. And the way of partitioning NANAW in Mandarin can be sketched as in Figure 7.2.

Figure 7.2. Semantic maps of Saisiyat NANAW partitioned in Mandarin
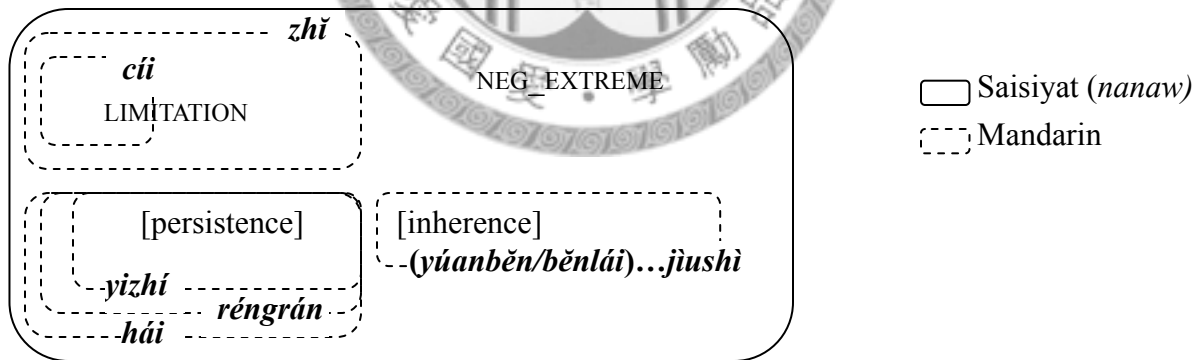
This figure shows that the concepts that are grouped in the same category coded by the same or related linguistic form in Saisiyat may be conceptualized differently in Mandarin Chinese. This explains why the informants have to use more than one Chinese

translation to explain the Saisiyat linguistics form. In fact, translating Saisiyat utterances into Chinese is in a way forcing the informants to conform to Chinese ways of semantic partitioning.

Nahaen

The case of *nahaen* also manifests language-specificity of semantic partitioning. The meanings that are conceived as related and coded by the same linguistic form *nahaen*, constitute a connected conceptual space in Saisiyat. This conceptual space is partitioned in Mandarin into at least five categories. *Nahaen* is a "family" that contains three senses: REPETITION, SUCCESSION, and PRECEDENCE. In Mandarin, repetition of an activity is partitioned into two categories, respectively labeled as *zài* and *yoù*, and they differ in terms of vantage points. See Figure 7.3. *Zài* and *yoù* both represent A and A' ordered in sequence with a gap in between, and they both focus on A' as the repetition of A. However, *zài* refers to the speaker's position at time t-1, but *yoù* signals the speaker's position at time t, which results into realis repetition (past/present) and irrealis repetition (future) respectively.

Figure 7.3. The viewing angle of *zài* and *yoù*
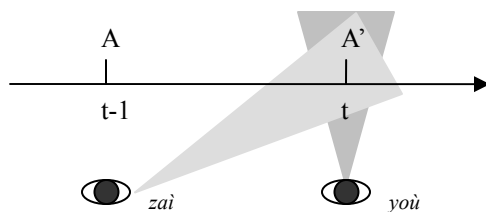
This difference has been pointed out in Li and Thompson (1981). *Zaì* and *yoù* are in complementary distribution in terms of the contexts they occur in. In (131), observe how *zaì* is not compatible in realis contexts, and *yoù* is anomalous in irrealis contexts.

(131) a. nǐ   **yoù/*zaì**   laí   lė
     2SG   **again/again**   come   CRS
     'Here you go again.'
  b. nǐ   **zaì/*yoù**   laí   yicì,   wǒ   jiù   dǎ   nǐ
     2SG   **again/again**   come   once   1SG   CONJ   beat   2SG
     'If you go like this again, I will beat you.'

However, in cases where some activity is expected to happen, one can only use *yoù* rather than *zaì*, as illustrated by (132) below. It is interesting that *yoù* and *zaì*, although typically viewed as realis/irrealis contrastive pair, are not of equal status in terms of subjectivity. In fact, *yoù* is prone to be used in cases where subjective expectation or judgment is involved, which is evidenced by the fact that when the utterance contains a CRS (current relevance state) marker *lė*, *zaì* is almost always incompatible.

(132) Mandarin expected repetition
    tamen   *mingtian*   **yoù/*zaì**   *yao*   *damajiang*   *le*
    3PL   tomorrow   **again/again**   want   play.majiang   CRS
    'They will have a meeting again tomorrow.'

Mandarin *hái* 'still,' commonly known as a continuous marker, also denotes repetition of an activity peripherally. Two kinds of repetition events are denoted by *hái*: a) repetitions that are counter-expected, as shown in (133), and b) repetitions motivated by speaker's volition, as shown in (134). However, these two functions denoted by *hái* are

not "plain" repetition of an activity, and they are also not the prototypical meaning of *hái*.

(133) Repetition that is counter-expected
jìao	*ta*	*búyào*	*lái,*	*ta*	***hái***	*lái*
call	3SG	NEG	come	3SG	**still**	come

'(I) have told him not to come, yet he came again.'

(134) Repetition that shows personal volition
zhèlǐ	*dongxi*	*hěn*	*hǎo*	*chi,*	*wǒ*	*míngtian*	***hái***	*xǐang*	*lái*
here	thing	very	good	eat	1SG	tomorrow	**still**	think	come

'The foods are good here, and I wish to come again tomorrow.'

In Mandarin, the relais/irrealis contrast of repetition has also been found in succession of activities. Again, there is a vantage difference made in Mandarin, i.e. *yòu* denotes past/present realis succession (135a), whereas *zai* denotes future irrealis event succession (135b).

(135)a.
ta	*zhúo*	*lě*	*jǐ*	*kǒu*	***yòu***	*xìang*	*qían*
3SG	peck	CRS	some	mouth	**again**	toward	front

fei	*yi*	*dùan*	*lù*	***yòu***	*zhǎodào*	*lě*	*shíwù*
fly	one	CLF	road	**again**	find	PFV	food

'(The bird) pecked, and then flied a little forwards, and then found some foods.'

b.
yǒuji	*féilìao*	*bìxu*	*(xían)*	*faxìao*	*fǔ-shóu*	*hòu*	***zài***	*shǐyòng*
organic	fertilizer	have.to	first	ferment	rot-ripe	after	**again**	use

'Organic fertilizers have to be fermented first before use.'

In Mandarin, continuity of an activity is primarily labeled by *hái* 'still' and *réngrán* 'still.' An example is given in (136). The use of *hái/réngrán* presupposes that this activity extends from the time prior to the reference time.

(136) *lǐang* *gè* *rén* *suirán* *fenkai* *lè*
two CLF human although divide CRS
dànshì **hái/ réngrán** *bǎochí* *yǒuyí*
but **still/still** maintain friendship
'Although the two separated, they still maintain their friendship'

In Mandarin, in addition to *lìngwài* 'another,'the meaning of addition is labeled also by *hái*. Like Saisiyat *nahaen*, Mandarin *hái* denotes additive sense only in the listing or elaborative contexts. There is a strong tendency of lexicalization of additive *hái* with existential *yoǔ* 'exist.' *Hái -yoǔ* 'still-exist' in a sense represent that something exists in addition to the others.

Prelude of an activity in Mandarin is expressed by the lexical item *xian* 'first' or a form similar to irrealis succession, i.e. *xian…zai…* 'first…then…' When used as a suggestion or advice in polite manner, examples of *xian* 'first' often collocates with V-*(yi)ge*-N, *yi-xia* 'a while', *yi-dian* 'a bit', *yi-xie* 'some,' etc. They are the constructions that are used to trivialize an action for purpose of politeness (Biq 2004; Chen 2006).

(137)a. **xian** *shùi* *gè* *jìao* **zài** *zhěnglǐ* *dongxi* *ba*
   first sleep CLF sleep **again** clean.up thing PAR
   'Have a nap first before cleaning up!'
  b. *nǐ* **xian** *zài* *zhèr* *děng* **yixìa** *wǒ* *qù* *nàbian* *mǎi* *pìao*
   2SG **first** LOC here.DIM wait **a.while** 1SG go there buy ticket
   'Wait here! I go there to buy tickets.'

If we neglect the functional overlapping of peripheral extensions of these Mandarin lexical items, the ways of partitioning Saisiyat NAHAEN in Mandarin can be summarized as the followings.

(138) Semantic partitioning of Saisiyat NAHAEN in Mandarin

(A) In Saisiyat, REPETITION, SUCCESSION, and PRECEDENCE are conceptualized as in the same linguistic category, labeled by *nahaen*, but they are grouped into more than one category in Mandarin with different linguistic coding.

(B) REPETITION and SUCCESSION are partitioned in Mandarin into realis and irrealis, but in Saisiyat, such distinction is not made.

(C) PRECEDENCE takes the form of irrealis SUCCESSION or by *xian* 'first.'

Based on (138), a semantic map can be constructed as in Figure 7.4.

Figure 7.4. Semantic map of Saisiyat NAHAEN partitioned in Mandarin
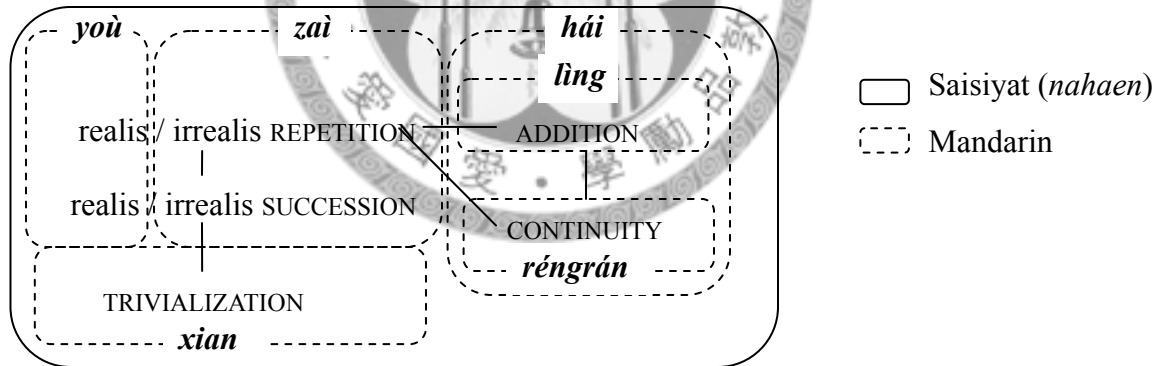
The case of *nahaen* is much more complex than *nanaw*, for the semantic partitioning in Mandarin is not a clear-cut one, and the fuzzy boundaries as well as complex functional overlap cannot be faithfully represented in this figure.

Such categorical mismatches hints at the danger of lexis-to-lexis translation: A fieldworker may easily transplant the semantic categorization from one language to another when the condition is in fact not so straightforward. Our interpretation of a lexical item is influenced by the knowledge of the meta-language, and direct translations reflect little of the categorization in the target language.

Ma'

The case of *ma'*, as we have shown in Chapter 6, shows a high degree of cross-linguistic predictability. Even in such a case of high cross-linguistic semantic predictability, the mismatch of semantic partitioning still holds. As we have shown in Section 6.6.2, Saisiyat *ma'* has three senses: COMPARISON, INCLUSION, and (CONNECTION), but the boundaries of semantic partitioning of Saisiyat *ma'* and Mandarin *yě* do not match perfectly, and some of the functions of Saisiyat *ma'* should be expressed by *yòu, hái* or *dou*. The semantic partition of MA' in Mandarin that we have presented in Chapter 6 (cf. Table 6.2) can be summarized in (139).

(139) Semantic partitioning of Saisiyat MA' in Mandarin
  (A) In Saisiyat, COMPARISON, INCLUSION, and (CONNECTION) are conceptualized as in the same linguistic category, labeled by *ma'*. A large part of the MA' conceptual space matches the YĚ conceptual space in Mandarin, but there are some marginal mismatches.
  (B) For the COMPARISON sense, the [similarity] force that compares different activities carried out by the same person can be expressed by *yòu* in addition to *yě*. The [counter-expect] force should be expressed by *hái*.

(C) For the INCLUSION sense, the [indefiniteness] force can be expressed by *dou*, and for affirmatives, only *dou* is acceptable, but not *yě*.

(D) For the (CONNECTION) idiomatic sense, there is no consistent equivalent in Mandarin.

Based on (139), we sketch a semantic map in Figure 7.4 to show how the "conceptual category" MA' in Saisiyat is partitioned in Mandarin. Although the majority functions of Saisiyat *ma'* can be expressed with *yě* in Mandarin, in cross-linguistic comparison of polysemy, mismatches are always expected.

Figure 7.4. Semantic maps of Saisiyat *ma'* partitioned in Mandarin
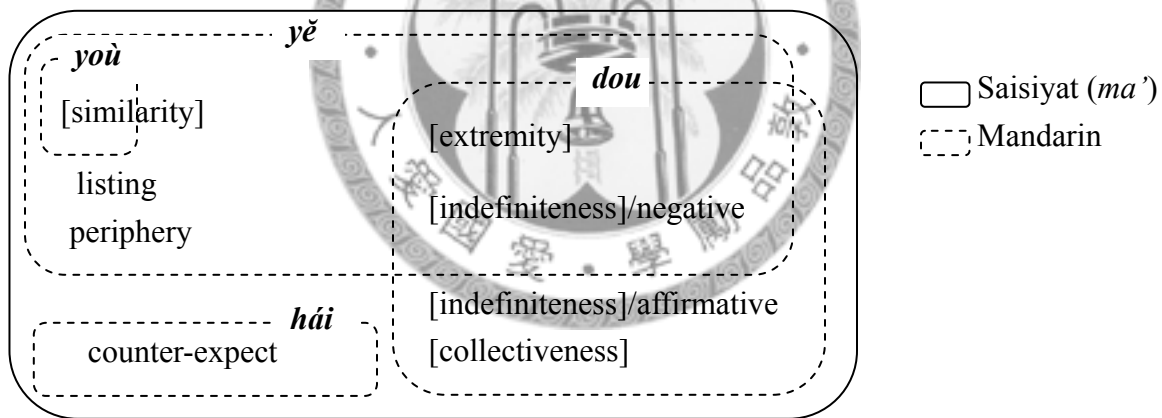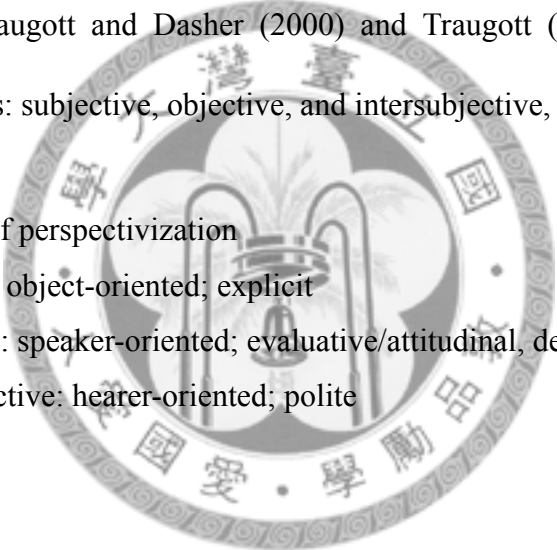
In all the three cases, the comparison shows that each lexical item has a particular way of categorizing and associating concepts. In every language, the way of categorizing the conceptual structure for the purpose of verbalization is unique.

7.4.3. Perspectivization: A socio-cultural view

The "conventional ways of conceptualization" in each language invites us to reconsider the relation between conceptual structure and semantic/syntactic structure in the framework given by Croft in Figure 7.1. We are also prompted to rethink the notion known as perspectivization in cognitive linguistics. Between the conceptual space and the semantic and syntactic structures, there is a perspective taken by language users for verbalization.[4]

According to Traugott and Dasher (2000) and Traugott (2003), there are three alternative perspectives: subjective, objective, and intersubjective, as summarized below:

(140) Three ways of perspectivization
    (A) Objective: object-oriented; explicit
    (B) Subjective: speaker-oriented; evaluative/attitudinal, deictic
    (C) Intersubjective: hearer-oriented; polite

With the three ways of perspectivization, language structures are supposed to be rich and idiosyncratic. The present study, nevertheless, shows that in a speech community, there is usually a preferred way of expression, which we will call the "collective" tendency of perspectivization. This is not a new claim, but can rather be traced back early in Saussure who regarded language as a social product — an agreed way of

---

[4] The term "perspectivization" has been used in many different senses. It is used elsewhere to mean the "vantage point" one takes to view a temporal or spatial configuration, e.g. "event-internal view" vs. "event-external view" as well as "realis vantage point" vs. "irrealis vantage point." This kind of perspectivization is referred to as "vantage point" or "viewpoint" in the present study. The term "perspectivization" in this study is used strictly to refer to the linguistic empathy/alignment with a specific conversational element in a communicative setting.

symbolization in a speech community for the purpose of communication. Such a perspective can be added to the above list:

(141) The fourth way of perspectivization
    Collective: culturally-oriented; ritualized; unmarked

Speakers' free choice is largely constrained when there has been a "collectively-oriented" way of expression in the speech community (Gyori 2002). Unless it is necessary to employ a "marked" way of expression, the conventional perspectivity is chosen as the priority for felicity of conversation. In a broad sense, constructions and fixed collocations are products of such collective perspectivization. They are agreed linguistic patterns that have to be "learned" or "memorized" for mastery of the language. As suggested by Graumann (1990), communicative competence of perspectivity is not only the simulation of others' perspective, but also the ability to grasp a social group's perspective.

Traugott (2003) claims that intersubjectification is in essence the function of subjectification, and here we would like to suggest that "collectivization" can be regarded as intersubjectification in essence. Rather than taking a speaker-internal view, the speaker in real conversation has to take a speaker-external in order to appeal to the hearer following a social-cultural convention of speech. Like intersubjectification, collectivization appeals to the speaker-external factors of language, but with a wider scope covering the backgrounding information of communication. In fact, such an

extensive view of "intersubjectification" is not our new proposal. In Nuyts (1998), the term "intersubjectification" is defined as the appeal of the evidence known to or accessible to a larger group of people who share the same conclusion as the speaker. In summary, in the three-level structure of language suggested by Radical Construction Grammar (Croft 2001), mapping between levels are geared to collective perspectivization, a broad view of intersubjectification, which suggests the incorporation of semantic-pragmatics and syntactic-pragmatics in linguistic studies.

7.5. Between language and thought

The Semantic Map approach has an ultimate assumption to rebuild the blueprint of universal human conceptualization (Haspelmath 2003), but according to this approach, the way to approximate language universality is by observation of language diversity. Presumably, semantic categorization and syntactic manifestations of the "same" concept tend to vary in different languages (Croft 2001). RCG correctly predicts that meaning is language-specific as well as construction-specific. A notion (a conceptual space) can substantiate various alternative ways of conceptualization, and different languages highlight different aspects of it with different semantic categorization and grammatical constructions. If the presupposed universal blueprint of human conceptualization does exist, we still prone to agree with Haspelmath (2003: 220) in saying that there are language-particular and construction-specific subdivisions of the universal semantic structure, which requires special attention and specific terminology to deal with each

language.

Employing this approach, it seems inevitable for us to face the year-long debate of universalism vs. relativism, although cross-linguistic semantic mismatches make no claim on their impacts on human thinking process. We would like to support a modest view of relativism here — acknowledging both "conceptual universality" and "grammatical relativity."

Let us review briefly the debate of language universality versus relativity. For a very long time, linguists have been ambitious to claim the relation between "speech" and "thought." It is argued that by studying the forms in languages of the world, we may uncover the thinking process of speakers in different speech communities. This postulation has invited considerable debates between relativists and universalists, and it remains an untied-knot ever since. Linguistic relativism claims that "meaning" lies in the form of a language. When forms diverge from one language to another, meanings are supposed to be culturally relative (Malinowski 1938; Boas 1940). Radical followers of Sapir-Whorfian hypothesis thus suggest "linguistic determinism," i.e. the form of a language affects the thinking capacity of its speakers (Whorf, cited from Carroll 1956). The rise of Chomsky's "cognitive approach" in 1960's is in a sense a challenge to relativism. Chomskian approaches hypothesize the innateness, autonym and universality of all human languages. By looking at the regularities of various grammatical patterns, we may find universal modularity in languages all over the world, e.g. the study of word order in Greenberg (1963). This view has inspired mushrooming syntactic studies in

search of the shared cognition of in every speech community. Even in the semantic field, there are efforts to counteract the relativist view. For example, the Natural Semantic Meta-language (NSM) suggests lexicon or grammar is constrained by the linguistic repertoire conventionalized or formalized in a language, and the search of "meaning" cannot be fully embedded in it (Wierzbicka 1996; Goddard 2001). With rigorous methodologies, NSM breaks "meaning" into fundamental building blocks, the so-called "primes," which are claimed to be universal. If interpreted in a radical way, it implies that human beings in all over the world can theoretically think in the same manner.

Linguistic determinism and absolute universalism represent two radical poles in claiming the relation between "speech" and "thought." Recently, theories in anthropology, psychology, and linguistics hold an intermediate view between extreme relativism and extreme universalism. While acknowledging the universal basis of language as a discourse-interactional pattern, the importance of social-environmental contexts in each culture has also been emphasized (Gumperz and Levinson 1996). Goddard and Wierzbicka (2004) and Goddard (2002) in their recent works of NSM put much more premium on the significance of social-cultural factors: The semantic primes are still suggested to be universal semantic cores, but the exponents of those primes and cultural scripts may vary in different languages.

In our study, we found that the Mandarin glosses given by the informants are too often inconsistent because there are language-specific considerations which are hardly translatable. Su and Huang (2006) thus suggest including cultural-specific information in

the so-called ethnolinguistic notes in corpus documentation system. This will not only facilitate the understanding of corpus users, but also help to preserve precious information that is of particular value to the study of endangered languages, such as Formosan languages. We will elaborate this point in Chapter 8.

Language specificity is attested in our investigation, but in view of "collective perspectivization" discussed in 7.4, the language-specific categorization of a conceptual space should NOT be taken as a piece of evidence in support of the claim that formal differences should have any effect on the speaker's thinking. We would like to emphasize that the collective view of linguistic perspectivization limits us to talk about "social cognition" (in the sense of Tomasello 2003) but not "personal cognition." The specific ways of semantic partitioning presented in the present dissertation reflect a shared viewpoint of the speech community, and do not have direct implication to the speaker's cognitive ability. An insightful observation in Slobin (1996) suggests that the relation between "speech" and "thought" can be more correctly interpreted if rephrased as "speaking" and "thinking." That is, the conventional ways of conceptualization in a speech community will definitely constrain the way of expression, and a speaker is required to "become sensitive" to the grammatical coding stipulated in that language. In this view, the relation between "speaking" and "thinking" is a dynamic process, and linguistic structure affects the thinking of its speakers primarily at the final stage of verbalization.

7.6. Summary

Carefully examining the semantic extension of the three delimited lexical items, we find that the speaker is becoming more and more aware of his epistemic stance, and also attempt to align with the hearer's point of view in order to be politeness, a tendency known as "intersubjectification" (Traugott and Dasher 2002). From a cross-linguistic point of view, however, the way of perspective-taking does not merely reflect the speaker's subjective viewpoint, nor does it entirely constrained by the speaker's want to appeal to his immediate speech participants. Instead, the speaker's construal of a scene for the purpose of verbalization has to conform to the semantic partitioning and grammatical system of the language, a phenomenon known as "social cognition" (Tomasello 2003). This collective view of linguistic perspectivization urges us to take a modest view of language determinism: The lack of a specific linguistic device to express a notion may affect the thinking of its speakers at the final stage of verbalization, but it does not have any direct implication to the thinking capacity of its speakers.

Chapter 8   Implications for Corpus Documentation

8.1. Preliminaries

When a linguist encounters a novel word or an utterance in a language, it is important to know what it means. It being a common sense, there is little consensus to how the meaning can be ascertained in linguistic fieldwork (Samarin 1967). The most extensively-used way of meaning elicitation is fairly intuitive: asking the informants to translate the linguistic form into another language that is familiar to the researcher. Via the analysis of polysemes, we find that direct translation may bring about two unexpected outcomes. First, the translation represents a mixture of semantic and pragmatic considerations, and the sheer dependence on direct translation will lead to multiple meta-language glosses, sacrificing accuracy, economy, and consistency, the three features highly valued in corpus documentation. Second, polysemy reflects language specific ways of conceptual categorization, and cross-linguistic mismatches of semantic categorization are common. Translating a target lexical item into the meta-language may distort the way of semantic partitioning specific to the target language, and the relation between the meanings of a lexical item in the target language is likely to be misconstrued if a researcher fails to recognize their mismatches.

As it is advisable to gloss a polyseme consistently with a cover term (Lehmann 1982), in this chapter, we suggest employing notions of categorization theory, particularly the concept of "prototype," to look for a cover term of a polyseme. In Section 8.2, the

advantages of maintaining a consistent for a polyseme in corpus documentation will be explicated. We suggest that a consistent gloss of a polyseme can help corpus users make associations of the functions in a way much like those made in Saisiyat. In Section 8.3, we will illustrate via our three case studies how the cover term of a polyseme can be ascertained in view of the prototype theory. In Section 8.4, we argue for the need to treat direct translations carefully by presenting some shortcomings of direct translation in linguistic fieldwork.

8.2. Glossing a polyseme with a cover term

The boundary between semantics and pragmatics is not clear-cut (Taylor 2003), a claim supported from our investigation which shows inseparability of semantics and pragmatics of the informants' direct translation of the texts. It then appears paradoxical when we insist presenting a consistent semanto-syntactic translation in the line of morphemic gloss, and pragmatic translation in that of free translation. This section reviews the advantages of consistent glossing, presenting Lehmann's suggestions, coupled with the insights we get from our empirical investigation.

8.2.1. Insights from Lehmann

The stipulation to maintain the distinction between semantic and pragmatic meanings is proposed in Lehmann's (1982). According to Lehmann, a linguistic form that has many related meanings should be coded by a cover term for it gives the corpus users

a clearer picture of which elements are responsible for the lexical meaning, phrasal meaning, constructional meaning, and sentential meaning respectively. Glossing a polyseme with a cover term does not mean making a clear distinction between semantics and pragmatics. The effect is quite the opposite: It drives the corpus users to ponder how meanings are forged — enriched, impoverished, adjusted — in contexts.

According to Lehmann (1982), the second advantage of consistent glossing is that it can facilitate the user's identification of a linguistics form. When a polyseme is translated consistently into a single lexical item, corpus users can more easily understand that all instances are in fact cognitively related as a polysemous network. On the contrary, if related instances of a linguistic form are coded by different translations, they are likely to be misconstrued as homonymy.

8.2.2. Further support

In addition to the reasons given by Lehmann (1982), in our investigation, we find more reasons to maintain a consistent gloss for a polyseme. First, glossing a polyseme with one cover term can meet one important demand, i.e. economy, of any corpus design. Secondly, consistent glossing with the best equivalent can avoid transferring the ways of semantic partition of the meta-language into the target language. It thus leaves the data least manipulated.

To begin with, the coding in a corpus should be economic, and it is made possible only when the gloss strictly follows semantic-pragmatic distinction. Although inferences

emerging from contexts can be many, but it is theoretically not probable, as correctly observed in Riemer (2001), to enlist a new sense for a lexical item every time when it is used in a novel way. It has been suggested in the Leipzig Glossing Rules that when a linguistic form has more than one equivalent in the meta-language, these equivalents can be listed in the morphemic gloss, separated by semi-colons. Nevertheless, one would find it rather bothersome to enlist all possible equivalents of Saisiyat *nanaw* in utterances like (142).

(142) *sia*   *s<om>i'ael*   ***nanaw***
       3SG.NOM   eat<AF>   **only;still;originally;definitely**
       ' He kept eating.'

For the purpose of economy, a polyseme should thus be glossed consistently by one cover term.

The second advantage of consistent glossing is that it can facilitate cross-linguistic search of a polyseme in a vast amount of data. When making cross-linguistic comparison, a researcher who attempts to look for the Saisiyat lexical item meaning 'again' can only get instances of *nahaen* translated and glossed as 'again' in the corpus. The researcher has to make extra efforts to discover the fact that *nahaen* is not always glossed as such. Only when many-to-one mappings of meaning and form are detected can the researcher analyze the relations between the instances. And a corpus user that does not have expertise in linguistics may never discover this fact.

In addition to the economy and efficacy concern, glossing a polyseme with a cover

term has a third advantage: It can help the corpus users realize the kinds of relations made by the speech community, and in this way appreciate how the speech community utilizes an existing linguistic form to denote relevant scenarios in the world. A target linguistic form glossed with multiple meta-language equivalents creates an illusion that the meanings do make much difference in the target language. In fact, the conceptual distinction made in the meta-language does not necessarily make any difference in the target language. For this reason, we suggest the target linguistic form should be glossed with a cover term. It offers corpus users a chance to appreciate the possible associations available in the target language, and is preserving in a sense the unique way of human conceptualization. One of the aims of preserving endangered languages is to preserve the unique ways of conceptualization. As has been pointed out by the UNESCO Ad Hoc Expert Group on Endangered Languages, "Language diversity is essential to human heritage. Each and every language embodies the unique cultural wisdom of a people. The loss of any language is thus a loss for all humanity." Mithun (2001) suggests that the way a linguist records the linguistic data sometimes shapes the data as to reflect unexpected ideology. If a polyseme reflects the way concepts are categorized in a language, to gloss it by context-situated multiple translations may transfer the semantic partitioning of the meta-language into the target language.

8.3. Prototype and equivalence

Given that a polyseme should be glossed consistently in a corpus, the question is:

How the cover term is to be ascertained. When a linguistic form (F1) in the target language (L1) is assigned an equivalent (F2) in the meta-language (L2), it is often a case that many of F1's functions are not covered by F2, and in addition, F2 itself may have many functional extensions that do not correspond to that of F1. In the first situation, direct translation results in multiple equivalents, according to the contexts. In the second situation, the researcher is running a risk of obtaining an equivalent whose typical meaning is not corresponding to the semantic core of the target linguistic form. When the target is glossed with an equivalent that overlaps only peripherally with it, the gloss would be entirely misleading.

When a researcher considers glossing a polyseme with a more precise, consistent cover term, we suggest that we should identify the prototype of the polyseme and find the proper gloss on basis of it. In this way, we can avoid the two negative outcomes mentioned above. Glossing a polyseme on the basis of its prototype makes sense because the prototype is a) the most stable part of a category, b) the most representative part of a category, and c) the part that is most likely to support the inference of other category members (Croft and Cruse 2004).

When looking for the best equivalent, we pay little attention to category boundaries mainly because category boundaries are usually fuzzy and fluid (Langacker 1991) that cannot stand for the most stable part of a category. Matching of boundaries is considered optional in the search for the best equivalent. Croft and Cruse (2004) suggest that there are cases in which cross-linguistic equivalents match the prototypes yet differ in their

boundaries, for example, French *corde* and English *rope*. When F1 does not have an

equivalent that matches it both in prototype and in boundaries, a researcher should be

content with prototype matching and neglect the boundary mismatch, accepting *rope* as

the best equivalent of *corde*. Only when 1) more than one F2 fulfills the prototype

matching criteria mentioned above, or 2) no F2 fulfills the prototype matching criteria,

should we look for the best boundary-matches. Matching of boundary in our study is

determined by a) most extensions of F1 covered by F2, and b) less extensions of F2

unpredicted by F1.

8.3.1. Prototype at a low level

In our analysis, we assume the cognitive significance of the so-called "sense"

(following Cruse 1986, 2000; Tyler and Evans 2001; Croft and Cruse 2004; among

others). The instances are first classified into "senses" according the autonomy tests and

then the relations between the senses are identified. However, as we have mentioned in

Section 6.4.1., the most salient meaning for each of our three case studies is "smaller"

(more restricted) than "sense." For example, when the speakers are asked to make up

sentences containing *nanaw*, they unanimously produce utterances in which *nanaw* is

collocated with numeral predicates, which belongs to one type of the LIMITATION sense.

The same situation is also found in the other two case studies. Also, in inter-language

comparison, it is often at the specific levels that we can talk about mismatches of

cross-linguistic equivalents. In order to reflect and to explain cross-linguistic

complexities, the Semantic Map approaches are usually tuned to very fine-grained scale due to. For example, Kemmer (1993) make distinctions between "spontaneous events associated with animate beings" and "spontaneous events associated with inanimate beings." However, such a distinction is too subtle to be considered two "senses" in most human languages. If the prototype is highly specific, what is its implication for categorization theories?

We believe what we have come across pertains to the problem known as "schematicity" widely discussed in cognitive linguistic researches. Stimulated by insights from philosophy and psychology, some linguists hold that speakers have a tendency to form schemata that are generalized via their experience of using the language.[1] Taylor (1990) criticized that linguists, when attempting to encompass all instances in a category, often seek for schemata that are too highly abstract to account for the speakers' conceptual processing. The attempt to construct a minimalist schema can lead to "rule fallacy" (Langacker 1987a) or "generality fallacy" (Croft 1998). Langacker's (1990) notion of "granularity" may thus be deemed as a solution to the dilemmatic position between schematicity and specificity: the granularity is often tuned to the specificity end, yielding the so-called "low-level schema," although the tendency to build schemata is conspicuous.

---

[1] The term "schema" has been adopted to mean different notions in different fields. In philosophical fields, the term is used by Kant to mean "any one of certain forms or rules of the 'productive imagination' through which the understanding is able to apply its 'categories' to the manifold of sense-perception in the process of realizing knowledge or experience." In neurological and psychological fields, schema is known as "an automatic, unconscious coding or organization of incoming physiological or psychological stimuli, giving rise to a particular response or effect." (Oxford English Dictionary)

Nevertheless, we find Taylor's (1990) solution might exhibit more explanatory power: a schema might co-exist with a representative member in a category. For example, it is not conflictive to think of DOG as a category that encompasses members that have generalized formal and behavioral traits and meanwhile as a category that is organized by a representative type (such as "golden retriever") and its resemblance to other types of DOG. In this respect, we take a "split-prototype" view wherein "representative member" and "schema" are essentially different but do not conflict in a category. The informants' tendency to make up utterances collocating with numeral predicates invites us to hypothesize the existence of a specific representative member in the category NANAW. On the other hand, the capability of *nanaw* to instantiate formally or semantically similar instances in different context to some extent implies the existence of an abstracted schema.[2] The division between a low-level prototype and a possible schema invites one question: Should the lexical item be glossed on the basis of a schema, or on the most salient meaning?

We suggest that for semantic elements, the search of a gloss should be on the basis of a low level prototype that directly reflects their most prominent meaning at a level that is most retrievable to native speakers. For grammatical elements, the search of a gloss should be based on a generalized schema that can represent their functional abstractness. In the next section, we further explicate the division between semantic glossing and grammatical coding.

---

[2] The existence of a schema is also evidenced in the fact that a group of forces may be jointly responsible for a semantic extension.

8.3.2. Semantic glossing vs. grammatical glossing

In addition to the "split prototype" problem, the other question is also crucial to language glossing. Two types of glossing are often distinguished in corpus documentation: semantic and grammatical. Semantic glosses are used for "content" elements, and syntactic glosses are used for "functional" elements. Observe an example given in the Leipzig Glossing rules as reprinted here:[3]

(143) Lezgian

Gila	*abur-u-n*	*ferma*	*hamišaluǧ*	*güǧüna*	*amuq'-da-c*
Now	they-OBL-GEN	farm	forever	behind	stay-FUT-NEG

'Now their farm will not stay behind forever.'

By convention, syntactic elements, such as oblique marker, genitive marker, future marker, and negator in (143), are represented in small capitals. On the other hand, semantic elements are in regular fonts. The distinction is not a rigid one, though. For example, a researcher may like to treat *abur* as a grammatical marker by glossing it as "3PL" (third person plural marker). Therefore, the Leipzig Glossing Rules suggest that a researcher may choose different treatments of his data depending on the purpose of his analysis. For example, in (144), two types of glossing are both acceptable.

---

[3] This example originally appears in one of the works of Martin Haspelmath. It is adopted by the manual of the Leipzig Glossing Rules to illustrate the different ways to present grammatical elements as opposed to semantic elements, and we cite this example from the manual of the Leipzig Glossing Rules.

(144) Russian

My	*s*	*Marko*	*poexa-l-i*	*avtobus-om*	*v*	*Peredelkino*
1PL	COM	Marko	go-PST-PL	bus-INSTR	ALL	Peredelkino
we	with	Marko	go-PST-PL	bus-by	to	Peredelkino

'Marko and I went to Perdelkino by bus.'

The choice depends on two concerns: a) the degree of schematicity of the lexical item, and b) the language experience of the researcher. On one hand, grammatical elements are more abstract in meaning whereas semantic elements are more concrete and attainable. On the other hand, a researcher's increasing experience of a specific form may render used for different functions may change his treatment. When the functions are too diversified and cannot be encompass with a concrete label, he may consider change a semantic treatment to a grammatical treatment.

The decision of the researcher may vary, and for our studies, we maintain treating a lexical item semantically for the following reasons:

(145) Grammatical coding vs. semantic coding
(a) Functional elements usually do not change the meaning of an utterance, but the three lexical items under our investigation changes the meaning of an utterance, and should be treated as content elements.
(b) Functional elements are required by syntax, but the three lexical items under our investigation are not required by syntax, and should be treated as content elements.
(c) Functional elements are usually highly productive, but the three lexical items under our investigation are not or less productive, and should be treated as content elements.
(d) Functional elements usually constitute paradigms or compositional sets with

other related linguistic forms, but the three lexical items under our investigation do not constitute paradigms or compositional sets with other linguistic forms, and should be treated as content elements.

(e) The functional category is a "closed" category, and any addition to this category should be treated with much more prudence for it implies the grammatical system of a language. On the other hand, treating a syntactic element as a semantic element does not trigger such consideration in terms of the entire language system.

The search for the cover term, though not explicated in Lehmann (1982), is attempted in the present dissertation. We employ insights from Contrastive Linguistics, especially Krzeszowski (1990b), claiming that the best equivalent should meet two necessary criteria:

(146) Criteria of prototype matching
(a) Mutual matching in the prototype area
   The target form (F1) in the target language (L1) should be covered as extensive as possible by the equivalent (F2) in the meta-language (L2).
(b) Wide coverage of the prototype of the target
   The parts of F2 that match F1's prototype should be F2's prototype.

In this section, we will apply the notion of prototype equivalence to look for the most proper meta-language equivalent(s) for inter-language glossing.

8.3.3. Empirical studies

The three case studies we presented in Chapter 4, 5, and 6 exhibit different degrees of cross-linguistic predictability, giving rise to different results when direct translation is used. The case of *nanaw* shows a low degree of cross-linguistic predictability. The informants tend to use different Mandarin glosses to explain the utterances containing *nanaw*, and those Mandarin glosses look irrelevant in Mandarin. Using direct translations as the gloss of *nanaw* creates a "homonymy illusion," i.e. the relation between the instances of *nanaw* cannot be detected by corpus users. In the case of *nahaen*, similar semantic extensions are more common in the cross-linguistic aspect. Regardless of the multiple Mandarin glosses, corpus users can more or less grasp the relation between the instances of *nahaen* because those Mandarin glosses have many functional overlaps in Mandarin. However, glossing *nahaen* by multiple Mandarin glosses fail to show how the meanings of *nahaen* are conceptualized as related in Saisiyat. In the case of *ma'*, due to the high cross-linguistic predictability of semantic hierarchy, the informants are more likely to give a consistent semanto-syntactic translations of *ma'* rather than its pragmatic translations.

If a researcher considers looking for a cover term to gloss a polyseme, we propose a set of procedure to find out the cover term that can properly represent a specific lexical item. In this section, we will illustrate how the procedures of prototype equivalence proposed in Chapter 3 can be applied in this respect via our three case studies.

8.3.3.1. The cover term of *nanaw*

The prototypical function of *nanaw* is to denote limitation of a quantity. Among the direct translations of *nanaw*, *zhǐ* is usually listed in Mandarin dictionaries to denote limitation of quantity. Based on three dictionaries, we synthesize the order of the meanings of *zhǐ* in Mandarin as shown in Table 8.1.[4]

Table 8.1. The meanings of Mandarin *zhǐ* as compared to Saisiyat *nanaw*

	The meanings of Mandarin *zhǐ*	A meaning of *nanaw*?
1	limited scope on specific things/activities (adv.)	Yes (prototype)
2	to do one thing with all one's might and neglect other things (adv.)	Yes
3	a disjunctive marker (collocating with copula *shì*) (adv.)	—[5]
4	to have no other choice but to… (collocating with *hǎo* 'good') (adv.)	—

As shown in Table 8.1, the first enlisted meaning of Mandarin *zhǐ* 'only' is also the most salient meaning of Saisiyat *nanaw*. Based on the criteria of prototype matching, Mandarin *zhǐ* 'only' is a proper equivalent for semantic glossing of Saisiyat *nanaw*. We would like to point out one important issue: Lexicalization often yields a specific meaning when a lexical item is associated with fixed elements in a particular grammatical structure, such as idioms and fixed expressions. They are typically language-specific. For example, *zhǐ* 'only' is morphologically compounded with *shì* (a copula) to denote disjunctive mood,

---

[4] Please refer to Appendix E to look for the synthesis of the meanings and the ordering of meaning of Mandarin *zhǐ(yǒu)* by consulting three Mandarin dictionaries: Jìaoyùbù Gúoyǔ Xǐao Zìdǐan (Miniature Dictionary of the Ministry of Education), Gúoyǔrìbào Cídǐan (Mandarin Daily News Dictionary), and Shìyi Xúesheng Cídǐan (Student Dictionary published by Shiyi Bookstore).

[5] The "—" symbol signals that *nanaw* is not found to have comparable functions.

and compounded with *hǎo* 'good' to express concession. In fact, it also frequently collocates with *yǒu* 'exist' to indicate limitation of a quantity. We suggest that a candidate meta-language gloss should preferably be less morphologically or syntactically laden, such as *zhǐ* 'only,' to avoid the impacts of language-specific lexicalization.

In addition to *zhǐ* 'only', another Mandarin translation *cái* is constantly adopted, when the informants are asked to translate *nanaw* in the sense of quantitative limitation. When we examine the meanings of Mandarin *cái* against the meanings of Saisiyat *nanaw*, we obtain the following result, as in Table 8.2.

Table 8.2. The meanings of Mandarin *cái* as compared to Saisiyat *nanaw*

	The meanings of Mandarin *cái*	A meaning of *nanaw*?
1	a moment ago (adv.)	—
2	a quantity less than expectation (adv.)	Yes (prototype)
3	limited condition for something to be realized (adv.)	—
4	an emphatic marker to counter a statement (adv.)	—

The result shows that the prototypical meaning of *cái*, i.e. the first enlisted meaning in three Mandarin dictionaries, is not found in *nanaw*. The more typical meaning of Mandarin *cái* is to indicate immediate temporal precedence, as illustrated in the Chinese example below:

(147) *wǒ*   ***cái***   *chi*   *shǔigǔo*
       1SG   **CAI**   eat   fruit
       Reading 1: 'I ate only fruits.'
       Reading 2: 'I just ate fruits not long ago.'

In this respect, *cái* is not as good an equivalent of Saisiyat *nanaw* to be used as its semantic gloss. We conclude that all instances of Saisiyat *nanaw* should be glossed as *zhǐ* cross-textually.

8.3.3.2. The cover term of *nahaen*

The most salient function of *nahaen* to native speakers is to denote recurrence of activity. Four Mandarin counterparts are adopted by the informants to translate the instances of [recurrence]: *hái* (cf. (133) and (134)), *zài*, and *yòu*. We look up those lexical items in dictionary, and synthesized their meanings as well as the ordering of their meanings, summarized in Table 8.3. According to this table, the prototypical meanings of *zài* and *yòu* correspond to that of *nahaen*. The prototypical meaning of *hái* is one meaning of *nahaen*, but not as prototypical. In addition, Mandarin *hái* has much more extensions that are not covered by *nahaen* (i.e. four of nine meanings). In this respect, the periphery of *hái* does not match that of *nahaen*, following the criteria of periphery matching (§ 3.4.3).

Table 8.3. The meanings of Mandarin *yòu, zài, hái* as compared to Saisiyat *nahaen*

Lexemes	Meanings enlisted	A meaning of *nahaen*?
yòu	repetition of an activity (adv.)	Yes (prototype)
	coordination of several activities/properties (adv.)	—
	emphatics of negations (adv.)	—
	succession of activities (adv.)	Yes
	a higher degree	—
	addition of a fraction to an integer (adv.)	Yes
zài	repetition of an activity (adv.)	Yes (prototype)
	a higher degree (adv.)	—
	succession of an activity (adv.)	Yes
	to introduce an activity to happen in the future (adv.)	Yes
	continuation of an activity (adv.)	Yes
hái	continuation of an activity (adv.)	Yes
	higher degree (adv.)	—
	addition of more activities (adv.)	Yes
	not yet (adv.)	Yes
	to remain in situ (adv.)	Yes
	offering several choices (collocating with copula *shì*) (adv.)	—
	repetition of an activity (adv.)	Yes (prototype)
	to passably meet a standard (adv.)	—
	to indicate a surprise (adv.)	—

As *yòu* and *zài* both fulfills the criteria of prototype matching, one of the solutions is to list both of them as the gloss of *nahaen*. The other solution is to look for the equivalent that also better captures the boundaries of Saisiyat *nahaen*. The criteria for periphery matching are restates below:

(148) Criteria of boundary matching

    (a) More coverage of F1's extensions

        Most F1's extensions are also F2's senses.

    (b) Less extensions that are not covered by F1

        F2 may have many extensions of itself, and the extensions should better be

        the senses of F1.

According to our analysis, *nahaen* has three senses: REPETITION (including [recurrence], [continuation]) and [addition]), SUCCESSION, and PRECEDENCE. It appears *zài* covers more senses and forces of *nahaen,* and it has fewer extensions that are not covered by *nahaen*. In this respect, if periphery matching is taken into consideration, *zài* would be chosen as the gloss for Saisiyat *nahaen*.

In (149), we suggest two solutions to gloss Saisiyat *nahaen* with maximum accuracy, consistency, and economy.

(149) *sia*   *s<om>i'ael*   ***nahaen***
      3SG.NOM   eat<AF>   **again**
      3SG.NOM   chi<AF>   **yòu;zài**   **(Solution 1)**
      3SG.NOM   chi<AF>   **zài**   **(Solution 2)**
      'He is still eating.'

Solution 1 takes into consideration of prototype matching, and Solution 2 further incorporates the consideration of periphery matching. A corpus developer may choose from either of them, but the way of glossing is supposed to be consistent cross-textually.

8.3.3.3. The cover term of *ma'*

Based on the informants' direct translation, there is high consistency in choosing Mandarin *yě* as the translation of Saisiyat *ma'*. Our observation also confirms that many functional extensions of Saisiyat *ma'* are also found in Mandarin *yě*. Given the high consistency and cross-linguistic predictability, Mandarin *yě* should be taken as the gloss of *ma'*. Nevertheless, we will take this chance to examine our procedures of equivalent

searching. Following the same procedure that we have employed in the analyses for *nanaw* and *nahaen*, we consulted Mandarin dictionaries to synthesize the meanings of *yě*. The result is shown in Table 8.4.

Table 8.4. The meanings of Mandarin *yě* as compared to Saisiyat *ma'*

Lexemes	Meanings enlisted	A meaning of *ma'*?
yě	an expression resembling a preceding one (adv.)	Yes (prototype)
	all (collocating with interrogatives) (adv.)	Yes
	to passably meet a standard (adv.)	Yes
	a disjunctive marker signaling counter-expectation (adv.)	Yes
	emphatics (collocating with *yi-dǐan* 'a bit') (adv.)	Yes
	to be not too bad (adv.)	Yes
	coordination of two clauses to mitigate a judgment (adv./conj.)	Yes
	a marker that makes to conjoined expressions resonate and coherent (conj.)	Yes

Mandarin *yě* and Saisiyat *ma'* not only match in their core meanings, but also in extended meanings. In terms of categorization, the prototype and the periphery of Mandarin *yě* roughly match those of Saisiyat *ma'*, making it a proper equivalent to Saisiyat *ma'*. All the functions of Mandarin *yě* enlisted in the three dictionaries can find an equivalent way of expression with *ma'* in Saisiyat. Although perfect matching of cross-linguistic equivalents is not expected, degree of their correspondence is surprising. [6]

---

[6] When we look into the Formosan Language Digital Archive, Academia Sinica, we find instances of Saisiyat *nanaw* are consistently glossed by *zhǐ* 'only.' Instances of *ma'* are glossed as *yě* 'also.' We did not find tokens of nahaen in the Archive. (http://formosan.sinica.edu.tw/ch/intro.htm)

8.3.4. Prototype: A contextual view

With categorization theory as the foundation of our semantic glossing, we maintain that the glossing of a lexical item should reflect the prototype of that lexical item. It is the most stable part of the meaning of a lexical item which is capable of instantiating other related meanings of the delimited lexical item. We also maintain that the meaning of a lexical item will exhibit dynamicity in its immediate contexts, and the dynamicity should be reflected in free translation.

In assigning the meaning of a lexical item, our informants rely heavily on contextual clues — linguistic or extra-linguistic — to advance an interpretation that makes sense in context. The contextually-driven interpretation may sometimes override the prototype. This explains why Cruse (1986: 1) adopts a contextual approach of semantics, holding that the semantic properties of a lexical item are fully reflected in appropriate aspects of the relations it contrasts with actual and potential contexts. How the notion of "prototype" is sensitive to its contexts is in fact widely discussed in studies of categorization. For example, Ungerer and Schmid (1996) claim that prototype is situated in the context, as illustrated by (150):

(150) He opened the door to face a pretty young woman with a *dog* in her arms.

The kind of dog that first comes to our mind in this situation is probably a small lapdog, e.g. Pekingese, though a golden retriever may be more prototypical for the

concept of 'dog.' When the situational context is taken into consideration, our understanding of a lexical category will be modified so that our interpretation will make sense, reducing potential ambiguity or vagueness in any natural, cooperative communication.

The so-called "context" should not be limited to those linguistic or situational in nature. Instead, "context" also takes a socio-cultural meaning in theories of categorization. For example, Lipka (1987, cited in Ungerer and Schimid 1996) suggests that the prototype of DESK is contextually-defined: A prototypical desk in European culture is higher than the prototypical desk in traditional Japanese culture. The former may have drawers whereas the later does not. If prototype is context-specific, the notion "context" here should be understood as cultural models.

By showing the dynamicity of the interpretation of a lexical item in different contexts, we wish to advocate an interactional view of semantics, i.e. understanding semantics in view of pragmatics. Wittgenstein suggests the meaning of a word is flexible and subject to the "intention" of the speaker. Among many possible meanings, we are looking for the one that can also be attained by the hearer in the contexts. In this sense, the interpretation of a polysemous word is to arrive at a tacit agreement between the speaker and the hearer in terms of their mutual understanding. If indeed nearly all lexical items are polysemous, as suggested in Deane (1988), communication is greatly dependent upon the speaker's ability to stimulate the viewpoint and the knowledge background of the hearer.

8.4. Shortcomings of direct translation

We insist that a linguistic form should be glossed consistently with the best equivalent cross-textually, but the informants' direct translation hardly answer to this expectation. Matthewson (2004) warns that linguists should be cautious when they attempt to ask the informants for a translation. She gives two suggestions that are of particular importance. First, translations should always be treated as a clue rather than a result. Second, do not expect the consultants to conduct analyses. Throughout our investigation, these two points are proven to be genuinely wise. In this section, we will show the features of the informants' responses that we have observed in fieldwork elicitation, and by singling out the features, we may thus prevent some possible mistakes of linguistic analysis.

8.4.1. Adherence to global coherence
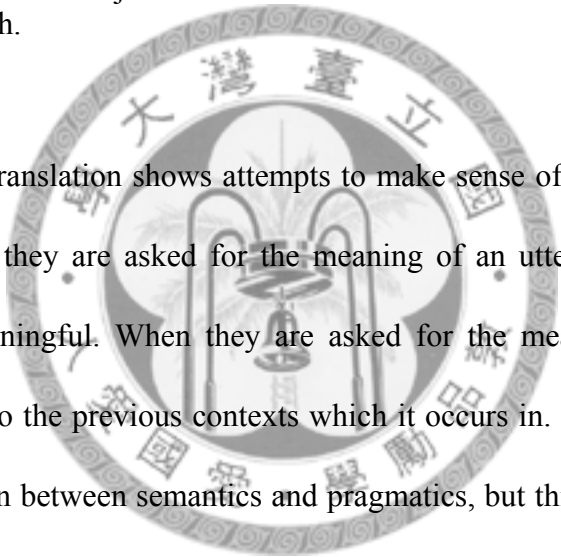
As we have shown, NTU-Formosan often contains multiple meta-language equivalents for one lexical item. One of the reasons is that the meaning of a word or an utterance is dependent upon its surrounding environment, and the informants are sensitive to pragmatic interpretations to make a word or utterance contextually meaningful. The informants are prone to stick to the pragmatic interpretation even when we ask the meaning of a single word in isolation. In (151), we present an observed interaction in which the researcher (S) attempted to ask an informant Gao (G) for direct translation of the word *nanaw*. (151a) and (151b) take place on different days.

(151) a. Process of direct translation on 2007/11/08

R: *'obay somi'ael nanaw ka 'aelaw*
    shì shémė yìsi?
G: 'Obay zhǐyǒu chi yú éryǐ.
R: *Nanaw* shì shémė yìsi?
G: Jiushì 'zhǐyǒu.'

R: What does *'obay somi'ael nanaw ka*
    'aelaw means?
G: It means 'Obay ate only aelaw.[7]
R: What's the meaning of *nanaw*?
G: It means 'only.'

b. Process of elicitation on 2007/11/10

R: *hahoera' kama'esem nanaw* shì
    shémė yìsi?
G: Jiushìshuo xuě běnlái jiù huì
    rónghùa, rúgǔo bèi tàiyáng
    zhàodào.
R: *Nanaw* shì shémė yìsi?
G: Jiushì 'běnlái jiù huì zhèyàng.'
    Jiushìshuo xuě běnlái jiù huì
    rónghùa dė ah.

R: What does *hahoera' kama'esem nanaw*
    mean?
G: It means the snow will melt naturally,
    probably heated by the sun.

R: What's the meaning of *nanaw*?
G: It means 'supposed to be.' The snow
    will of course melt.

The informants' translation shows attempts to make sense of the lexical item in the global context. When they are asked for the meaning of an utterance, they think of a scene to make it meaningful. When they are asked for the meaning of a word, they constantly refer back to the previous contexts which it occurs in. A researcher is usually aware of the distinction between semantics and pragmatics, but this distinction is foreign to the informants. If they are given a context in advance, they will stick to the context and their interpretation will coheres necessarily with this context.

8.4.2. Sensitivity to experience

Direct elicitation of meaning can be a real challenge to the linguists: The informants usually make judgment according to their real world experience rather than the sense of

---

[7] *'aelaw* means 'fish.'

230

an utterance. An utterance may have "sense" to the informants, but it lacks a "referent" in their culture or social context. If so, informants usually would opt to negate the existence of such an utterance. For example, in (152), the informants insist that 'a while' be assigned to the word *nahaen*. Although the repetition reading is also possible, the informants reject such a reading.

(152) *'obay   ki   kizaw   lasiwazay   **nahaen***
  PN COM PN divide   **again**
  "Obay and Kizaw separate on a temporary basis.'
  ??"Obay and Kizaw separate (> divorce) again.'

Later, we came across the same lexical item occurring in negative construction, as in (153). The informants, reject, as before, the repetition meaning of *nahaen*, insisting that it can only mean 'a while.' In (154), we place a perfective marker *ila* in this utterance, trying to limit the interpretation to repetition, but the informants think this statement is incorrect.

(153) *'izi'   'ip-lasiwazay   **nahaen***
  NEG NEG-divide **Again**
  'Let's not divorce on a temporary basis.'
(154)* *'izi'   ila   'ik-lasiwazay   **nahaen***
  NEG PFV NEG-divide **again**
  Intended meaning: 'Don't divorce again.'

Only when we encounter (155) of the same structure yet acceptable can we realize why (154) is not good. An experiential basis underlines: *lasiwazay* 'divorce' cannot be repeated because it is unusual for two persons to divorce twice.

(155) *'izi'*   *ila*   *'i'-raoe'*   **nahaen**   *ka*   *pinobae:*
       NEG   PFV   NEG-divide   **again**   ACC   wine
       'Don't drink alcohol anymore.'

One informant, when asked to explain the asymmetry of (154) and (155), reveals that Saisiyat is a monogamous society by convention, and divorce is uncommon, not to mention divorcing for the second time. Though they understand the meaning of (154), they won't take it because no "reference" can be located according to their world experience.

When informants reject an utterance, the reasons may range from mispronunciation, incorrect lexical choice, to pragmatic incompatibility (Mithun 2001). Direct translation offered by the informants in fact reflects considerations at the semantics, syntactic, or pragmatic levels. It is the researcher's responsibility to think of a way — comparison of the same lexical item in different structures, or of different lexical items in the same structure — to sort out the reason behind.

8.4.3. Incapability to explain textual functions

In our investigation, we learned that the informants find it difficult to explain the textual functions of a linguistic form.

Constructions such as *ma' isaa* and *isaa ma'* serve as markers to make discourse cohesion and to keep a smooth narrative flow. Informants often cannot explain the textual functions associated. It has been suggested in Traugott and Dasher (2002) that linguistic elements often undergo shifts of functions from propositional to expressive, which may
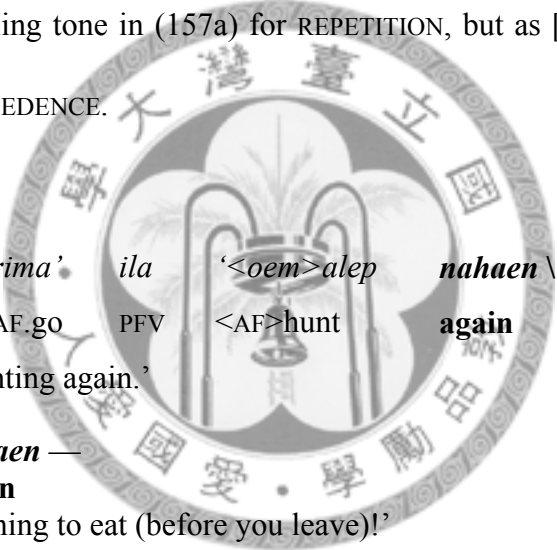
or may not go through a textual stage. Their suggested way of semantic change can be illustrated in (156):

(156) propositional > ((textual) > expressive)

The informants have as a rule no difficulty explaining propositional meanings. They often do so by giving us a condition for it to be true, including the necessary entities, activities, their sequences, their relations, and other relevant attributes. They also have no difficulty explaining expressive meanings. They do so by pointing out the intention, evaluation and speech acts of the utterance, or providing interactional situations in which the linguistic form will make sense. Textual functions of a linguistic form are on the other hand often neglected by the informants. When the informants are asked for the meaning of *ma' isaa* and *isaa ma'*, they often report that these strings do not have meanings. We thus hypothesize that those structures can often be omitted in non-interactional genre, making an illusion of their functional emptiness. In addition, to explain the textual function of a linguistic form involves meta-linguistic reasoning, i.e. describing language with language. Meta-linguistic reasoning involves sophisticated training and abstract reasoning. Matthewson (2004) points out that we cannot rely on the informants to do the analyses for us. To this, we would like to add a note: For the textual elements, it is almost impossible to find an answer from the informants.

8.4.4. Intricacy of volunteered analysis

Informants may volunteer to provide to provide explanations and analyses, but their input should only be used as reference, not as the answer. For polysemy, the seemingly irrelevant meanings and apparent sharing of the same form can be puzzling to the informants. In an attempt to make sense of the paradox, a young informant states that *nahaen* as REPETITION and *nahaen* as TRIVIALIZATION should be differentiated. He further argues for their difference by pointing out their phonologically different: It is pronounced as [nahæn\] with a falling tone in (157a) for REPETITION, but as [nahæn—] with a level tone in (157b) for PRECEDENCE.

(157) a. REPETITION

sia	*rima'*	*ila*	*'<oem>alep*	***nahaen*** \
3SG.NOM	AF.go	PFV	<AF>hunt	**again**

'He went hunting again.'

b. PRECEDENCE

si'ael	***nahaen*** —
eat	**again**

'Have something to eat (before you leave)!'

This phonological distinction is, however, not obligatory according to other informants. The functional division is also not as clear-cut. Two elder informants reported that *nahaen* in (157b) meaning PRECEDENCE can in fact be pronounced with a falling tone.

Sometimes, the informants give us their analysis out of good intention. To make a non-native speaker learn how a lexical item is to be used, it is better to making finer

divisions than to lump the seemingly irrelevant instances together. However, a researcher lacking experience and knowledge of the language may end up treating *nahaen* and *nahaen—* as two separate entries.

8.5. Some suggestions of using direct translations

Given those features of direct translation, a linguist should then be very cautious when he attempts to elicit the meaning of a lexical item via direct translation strategy. With this consideration in mid, in this section, we would like to point out three principles of utilizing informants' translations for linguistic analyses: a) treating direct translation as reference, b) treating any gloss as working gloss, and c) taking ethno-linguistic notes.

8.5.1. Treating direct translation as reference

In some cases such as *nanaw*, glosses such as 'only' and 'still' do not have a clear conceptual link between each other in many human languages, and the corpus users might have difficulty reconstructing their relation as a "family." The failure to recognize the semantic relations prevents a corpus user from appreciating the flexibility of human mind and the cognitive ability to relate concepts in a creative way.

But direct translation is still valuable in that it may serve as a reference for the researcher to discover the relations between various uses of a linguistic form. In NTU-Formosan, it is easy to find cases of multiple meanings associated with a single form. For example, in (158), *sinraehoe'* refers to 'breeds of animals,' 'types of things,'

and 'clans of people.'

(158)
a. Election

  17. B: ..   *E*
                     FIL
                     FIL

  18. M: ...   *piza'*   **sinraehoe'**   *Hiza*   *koSa'en*   *ka*   *wa'ae'*

		how.many	**species**	there	FIL	NOM	deer
		duoshǎo	**zhǒnglèi**	Nàlǐ	FIL	NOM	lù

'How many species (of deer) are there?'

'Nà lù zǒnggòng yǒu jǐ zhǒng ah?'

b. Holiday

  66. K: ...   *tata'a==*
                   chicken
                   ji

  67. K: ...   *'aeyam*
                   pork
                   zhūròu

  68. K: ...   *<L2haiyou==L2>*
                   [Man]
                   háiyǒu

  69. K: ...(0.8) *<L2tonikakuL2>*   *<L2samsen*   *laL2>*   *hoN*

		anyway[Jap]	three.kinds[Man]	[Man]	DM
		fǎnzhèng	Sānshēng	[Man]	DM

'Chicken, pork, or anything, as long as there are three kinds.'

'Ji ah, zhūròu ah, háiyǒu…fǎnzhèng sānshēng jiù shì lah.'

  70. C: ..   *o=*
                   BC
                   BC

  71. K: ..   *hara==*
                   like
                   xiàng

  72. K: ..   *to:o'*   **sinraehoe'**

		three	**kinds**
		sān	**zhǒng**

'Three kinds of offerings.'

'Zhūrú sānshēng lah.'

c. Holiday

 116. C: ...   *o==*   *to:o'*   **sinraehoe'** *hini*   *rini*   *raremewan*

		FIL	three	**clan**	here	here	PN
		FIL	sān	**shìzú**	zhèlǐ	nàlǐ	PN

'There are three clans here in Xiangtianhu.'

'Xiāngtiānhú zhèlǐ yǒu sān gė xìng.'

236

In the case of *sinrahoe'*, the various meanings constitute a semantic networking that is attainable even for non-native speakers. Most corpus users can infer their relation even if three different glosses are given. Direct translations with apparent semantic relations such as (158) are sometimes helpful for researchers to find out the relations between different uses of a form.

8.5.2. Treating any gloss as a working gloss

For language documentation at the early stages, inconsistent glossing is nevertheless unpredictable because a researcher cannot locate the core meaning of a lexical item given the multiple translations offered by the informants. In addition, a through analysis is not easy with insufficient linguistic data.

Bickford (1997) suggests a "rich text model" of interlinear glossing that includes "word gloss" in addition to "morpheme gloss." Example (159) is extracted from his paper to illustrate the design of such a model.

(159) A rich text model of interlinear glossing of Seri[8]

\po	hacx	ant	tahcniiixo	/ yoque	cmam	quih.
\ew	somewhere	**land**	it.was.poured	/ it.is.said	**cooked**	the
\mr	'akX	'ant	t- aa'-aknɟiiX -o	/ yo-ka-ææãSRõ	k-mam	k'
\em	somewhere	**earth**	Rl-Pv-pour-AdvS	/ Dt-US-say	SN-**ripe**	DefU
\et	cooked food was dumped out (there was so much).					

---

[8] The abbreviations at the beginning of each line stand for: po (practical orthography), ew (English word gloss), mr (morphemic representation), em (English morpheme), and et (English free translation). The original text contains Spanish glosses and Spanish translation. They are removed in this paper.

According to Bickford, the double layers of interlinear glosses give us a clearer picture of the semantic structure of the language: between the semantic core and contextual elaborations. For example, *ant* means 'land' in contexts and 'earth' in isolation. The same contrast is also manifested by *cmam*, which means 'cooked' in the context of (159) yet 'ripe' as its denotational sense. Bickford's design can "bridge between the highly analytical morpheme glosses and the free translation." Non-linguists can see how each morphemic element contributes to the meaning of words, and how each word accounts for the meaning of the sentence shown in free translation. The sudden leap from morphemic level to sentential/episodic level is mitigated. In addition, Bickford advances that this additional line of word glosses can allow native speakers to participate more actively in the work of language documentation without professional linguistic training.

Bickford suggests the word gloss can be a tentative solution especially for languages with complex morphologies. When enough language data are accumulated, researchers can then do morphological analysis based upon the word gloss by examining cross-textual distribution and structure of a specific linguistic form.

As we have shown in our studies of lexical polysemy, informants often offer glosses situated in its immediate context. The initial translation of a text may reflect "word gloss" rather than "morpheme gloss." But **the glosses can be regarded as working glosses** to be constantly checked and compared when more data are accumulated to see whether a new analysis is possible.

8.5.3. Necessitating ethno-linguistic notes

The present design of NTU-Formosan is following Lehmann (1982) and the Leipzig Rules, presenting only morphemic glossing and free translation. The gap between morphemic level and the phrasal level, as suggested in Bickford (1997), sometimes is so great that it demands a "conceptual leap" to bridge the gap. This leap may prompt the corpus users to appreciate the conceptual link made in Saisiyat, a point we try to argue in this chapter. A leap, when too big, can become an obstacle instead. For this, we may remedy by providing relevant contextual information, helping corpus users to make inference. Matthewson (2004) states also the need for contextual information to be noted down, especially under the following three situations: a) when the sentence will seem felicitous to the consultant unless a discourse context is provided, b) when the sentence being offered to the consultant is ambiguous, and c) when dealing with context-sensitive phenomena such as presupposition. In addition to linguistic contextual information, we suggest it is also necessary to note down cultural-specific information crucial for the understanding of a linguistic form. Currently, NTU-Formosan has set up a space to store ethno-linguistic information noted in the course of fieldwork. For example, (160) is an excerpt followed by an ethno-linguistic note to bridge the gap between the morphemic glossing and the free translation.

(156) Life

9. F: ...(2.4) t<om>angtang   ay   ./ .. ((CLEARS THROAT))
                  <AF>dig   Q
                  <AF>wa   Q
    'Do you do any digging?'
    'Nǐ yǒu-méi-yǒu wa tǔ?'

10. M: ..   eh==
                FIL
                FIL

11. F: ((CLEAR THROAT))

12. M: ...(0.9) t<om>angtang noka== .. kama-si'ael ka   **samiyan**
                  <AF>dig   use   NMZ-AF.eat   ACC   **oil;gas**
                  <AF>watǔ   yòng   NMZ-AF.chi   ACC   **yóu**
    'I use a machine.'
    'Wǒ yòng jicì.'
    Note: In SaiSiyat, `samiyan' is used to refer to any entity of super-natural or
    unexplainable power; therefore, `samiyan' is also used to refer to `god'.

This note (underlined in this excerpt) explains why the lexeme used to refer to God

or supernatural power is also used to denote gas or fuel.[9] Ethnological information is

important for anthropological studies, a view that has been advocated since Malinowski

(1938), who suggested three kinds of information be noted: the context of culture, the

context of situation, and the context of language, a view rarely followed in modern

development of the corpus. We suspect this is so partly because the contextual

information will grow infinite and linguists will end up doing the job of ethnographers.

Not all linguistic forms demand, however, this kind of attention, as advanced in Samarin

(1967). Only the structurally or culturally particular linguistic expressions are noted down

for NTU-Formosan. In addition, many Formosan languages are on the verge of language

---

[9] This excerpt is taken from NTU-Formosan as it is, but *samiyan* should in fact be glossed as 'god' after
our analysis.

death as well as cultural loss, recording the ethnolinguistic information is to preserve not only their way of speaking, but also the cultural wisdom that underlies.

8.6. Summary

Consistent glossing of a polyseme is advantageous in many aspects, but if the cover term fails to grasp the central meaning of a delimited polyseme, consistent glossing can still be misleading to corpus users. In this chapter, we employ the notion of "prototype" to look for the best equivalent to be used as the gloss of a lexical item. Three case studies are examined here as preliminary experiments. Direct translation is likely to assign inconsistent meta-language equivalents to a linguistic form, which sometimes fails to reflect conceptual categorization of the target language. The only way to determine the central meaning of a linguistic form is by "careful study of an abundance of examples and comparison with other forms" (Samarin 1967: 208). By pointing out the shortcomings of direct elicitation, we also suggest linguists be very cautious when using direct translation to elicit the meaning of a linguistic form. As correctly observed in field-related studies, such as Samarin (1967), Vaux and Cooper (1999) and Matthewson (2004), informants' responses are helpful, but should not be taken at face value.

Chapter 9   Conclusion

9.1. Recapitulation

This study approaches polysemy empirically from the problems we have encountered when documenting linguistic data of Saisiyat. Using direct translation approach, we find that a polyseme in one language often receives multiple equivalents in the meta-language. Consistent glossing can facilitate linguistic researches in locating target form in the corpus, but also help preserve the language-specific conceptual categorization unique to the target language. If enough data are accumulated and a polyseme can be compared and analyzed, a researcher may consider glossing it by a cover term. By theories of categorization, we suggest that a polyseme can be glossed on the basis of its prototype gestalt.

When analyzing the Saisiyat polysemous lexical items, we discover the impact of cross-linguistic semantic predictability on the use of direct translation for data analysis. A polysemy network with a low degree of cross-linguistic predictability is often found to be typologically-unimportant semantic extension manifested only in one language: With low degree of abstraction, and inconceivable in the relations between its meanings. A researcher is very likely to obtain via direct translation a group of translations that are seemingly irrelevant in the meta-language. If those direct translations are used in corpus documentation, the corpus users might neglect the relation between the uses, and are thus less likely to appreciate the specific way of semantic partitioning of the target language.

This is indeed the problem in the case of Saisiyat *nanaw*. When a semantic network has a higher degree of cross-linguistic predictability, similar semantic development is more likely to be found in other genetically-unrelated languages. The polyseme might have a large number of meta-language equivalents, but those direct translations are likely to have functional overlaps in the meta-language. When those direct translations are used as glosses in corpus documentation, corpus users, through careful observation, may be able to discover the relations between the glosses. However, multiple glosses of a single lexical item fail nevertheless to faithfully represent the ways of semantic partitioning of the target language. Furthermore, it is not economical for the corpus users to spend time thinking about the relations between the various uses of the lexical item, when the differences in uses are exaggerated due to multiple meta-language glosses. Our study of Saisiyat *nahaen* can be taken as one good example. We may also come across cases that are of a very high degree of cross-linguistic predictability, yielding similar ways of conceptual categorization in genetically-unrelated languages, and direct translation is likely to give rise to a higher consistency of meta-language equivalents. Semantic networks of this type are often associated with a high degree of abstraction, with the polyseme displaying a low degree of semantic contents and a high degree of dependence upon its collocating linguistic structures, which dictates the researcher to pay much more attention to constructional patterns. Our study of Saisiyat *ma'* is one of the examples.

Nevertheless, polysemy is the product of conceptual categorization, and because different motivations constantly compete with each other, semantic extensions are bound

to vary in different languages (Croft 2001). In other words, language diversity is supposed to be more common than similarity. We agree with Lehmann's (1982) claim that a polyseme should be glossed by a cover term cross-textually. This ties together its various uses, which is a good way to help the corpus users think about the semantic partitioning that might be unique to the target language. With the notion of "prototype equivalent" in Contrastive Linguistics, we suggest that the cover term used as the gloss should reflect the prototype of the polyseme. In view of this, we propose a set of procedure to look for the proper gloss of a lexical item on the basis of its prototype, which can serve as a guide for language documentation. As has been rightly observed in Samarin (1967), language glossing in earlier stages of corpus documentation is supposed to be inconsistent and imprecise, and the researcher needs to accumulate more data to find out the pattern by comparing the tokens of a polyseme occurring in different contexts.

Besides, when analyzing the meanings of polysemy, we find that a polysemy network often exhibits subjectification from the center to the periphery. The meaning of a lexical item is prone to be motivated by the speaker's subjective speculation, sympathy, evaluation, etc. Moreover, the speaker is further motivated by interactional considerations to align his viewpoint with that of the hearer in order to appeal to his hearer, a tendency referred to as "intersubjectification" of perspective-taking in discourse. From a cross-linguistic perspective, the so-called intersubjectification has to be redefined: The speaker's use of a language does not only reflect his free choice of construal, nor the

consideration of his immediate speech participants. In a cross-linguistic perspective, the speaker has to take into consideration a collective way of verbalization shared by the speech community. Individual speakers may not be aware of how a construal is chosen and why it is chosen. Instead, the choice of construal reflects a collective viewpoint taken by the entire community, which might be a result of constant negotiations of meaning accumulated cross-personally through a long time of language evolution. When we advocate for preserving language-specificity of semantic partitioning by proper methods of corpus documentation, we refer to this collective meaning of perspectivization rather than individual subjective choice of construal.

9.2. Implications and contributions

Empirically, we propose a set of standard that can be used as a guide for language documentation of polysemy. Combining the insights from Lehmann (1982) and our empirical investigation, we conclude that glossing a polyseme with a cover term cross-textually may yield five advantages:

a) Fundamental division between semantic contents of a lexical item and its pragmatic elaborations
b) Facilitating users' identification of a linguistic item
c) Economy and precision of data representation
d) Facilitating search of corpus data
e) Preservation of conceptual categorization of the target language

We suggest that when a researcher wishes to gloss a polyseme by a cover term, he can exploit the notion of "prototype." By employing theories of categorization that have been proven to be cognitively fundamental to human beings, we hope to make the corpus users appreciate the way of conceptual categorization of the target language, regardless of whether or not the users are familiar with linguistic theories. The prototype theory depicts the way human beings make inference in everyday life, and is potential to help us to achieve this goal in view of "backstage cognition," i.e. in an unconscious manner. The procedure to some extent shows that we believe **polysemy has to be investigated in a case-by-case manner**. Nevertheless, although most lexical items are polysemous and language-specific, the ways of partitioning are not always as complicated as the cases we have presented in this dissertation.

Theoretically, by investigating language-specific semantic partitioning, we advocate for re-thinking of previous lexical semantic assumptions. We find that the basic unit of meaning "sense" has little implication in empirical uses. In intra-language investigation, when our informants are asked to recite the meaning that is the most salient of a lexical item, their responses show little awareness of "sense" boundaries. Instead, their responses often reflect the level where basic grammatical relations between fundamental grammatical categories are defined and selectional restrictions and co-occurrence relations are stated. In cross-linguistic investigation, "sense" also has a shaky cognitive status. Due to the richness of inferential links and competition of cognitive mechanisms, language specificity often yields subtle groups of semantic units in cross-linguistically.

Consequently, comparative approaches often employ a large amount of subtle divisions that are contrastive not in terms of "cognitive autonomy" but in terms of different cross-linguistic manifestations. This explains why Semantic Map approach often ends with a large amount of usage "labels" (in the sense of Hapselmath 2003). The multiple labels are seemingly redundant and everlasting, but this is the only way to show the contrasts that make different formal manifestations in different language. This helps to explain why Haspelmath (2003) dispels the choice between "polysemy" and "monosemy," and adopts instead another term "multi-functionality" to refer to the sharing of the same form of many different uses.

Because our study tackles language-specificity of conceptual categorization, we inevitably face the debate of language universality versus language relativity, and also the relation between language and thought. The collective view of perspectivization that we embrace in the present dissertation clearly shows our stand: Language diversity reflects a collective way or habit of thinking, not the thinking ability of each language user. We do not exclude the possibility that subtle semantic divisions in one language may prompt the speakers to be more sensitive to a specific aspect of conceptual categorization, or it may make the speakers more insensitive to inference the relations between the concepts that are verbally coded in different ways. However, language is just one aspect of human cognition. Our thinking ability may not be entirely dependent upon speech, and the lack of one specific linguistic device to express a concept, in our view of collective perspectivization, does not impair the thinking ability of the speakers in a particular

speech community.

9.3. Further studies

We suggest that the notion of "prototype" can be exploited when a researcher attempts to look for a gloss that can be used consistently to gloss a polyseme. The present study is however limited to three cases of lexical polysemy. In the future, studies of different word classes (other than adverbials) should be incorporated to verify the practicability of this procedure. Studies of more different languages may also help us to examine and modify the present procedure. In addition, the present study focuses on strategies of semantic glossing, and may differ from principles of grammatical glossing, which is an issue to be followed up in the future.

Moreover, language comparison of three case studies also leads us to discover some preferred ways of verbalization that can be generalized to a number of lexical items in a language. We find that in our case studies of three adverbials, there might be a tendency of "convergence" in Saisiyat. In lexicographic studies, "convergence" is defined by the rendering of two or more words in the practice of translation (Hartmann and James 1998).[10] In languages with fewer vocabularies, there are supposedly richer ways of "conceptualizations" in order to verbalize newly-encountered situations. Lexical borrowing is one way to cope with vocabulary deficiency. In our investigation, we found that Saisiyat has borrowed a large amount of lexical items from Japanese, Hakka,

---

[10] For example, Dutch *slak* renders two English equivalents *slug* and *snail*. The opposite direction is "divergence."

Mandarin Chinese, and Taiwanese. Besides borrowing, various types of "construals" (in the sense of Croft and Cruse 2004) are also employed in this end. For example, Saisiyat *samiyan* primarily refers to "god," but it also means "gasoline," because they both have power beyond their scientific knowledge. As another example, Saisiyat does not have full-fledged causal connectors such as English *although*, *unless*, *because*, etc. so speakers are expected to rely greatly upon pragmatic inference, conceptual association, or interactive cues to reason for causal concepts. The rich ways of conceptualization in Saisiyat reflect not only the amazing imagination of human beings, but also the link between language and its cultural background. How Saisiyat speakers infer the intended meaning in verbal interaction with limited linguistic devices will be an issue for our further investigation. Whether concepts are prone to be grouped into "larger" families in Saisiyat is an issue that needs to be followed up in further studies. Here we would like to emphasize that language-particular tendencies of verbalization or conceptual categorization deserve our special attention. Glossing a polyseme with a cover term is a strategy to avoid presumptive conceptualization of semantic boundaries transferred from the meta-language.

References

AIKHENVALD, ALEKSANDRA IUREVNA. 2000. *Classifiers: A typology of noun categorization devices*. Oxford; New York: Oxford University Press.

AUSTIN, JOHN. 1961. The meaning of a word. *Philosophical papers*, ed. by J.O Urmson and G.J. Warnock, 23-43. Oxford: Clarendon Press.

BICKFORD, J. ALBERT. 1997. A Rich Model for Presenting Interlinear Text. *SIL Electronic Working Papers* (1997-003). Available on http://www.sil.org/silewp/1997/003/.

BIERWISCH, MANFRED. 1983. Semantische und konzeptuelle Reprïsentation lexikalischer Einheiten. *Untersuchungen zur Semantik*, ed. by Rudolf Ruzicka and Wolfgang Motsch, 61-99. Berlin: Akademie-Verlag.

BIQ, YUNG-O. 1989. *Ye* as manifested on three discourse planes: Polysemy or abstraction. *Functionalism and Chinese grammar*, ed. by James H-Y. Tai and Frank F. S. Hsueh, 1-18. South Organge, New Jersey: Chinese Language Teachers Association.

BIQ, YUNG-O. 2004. Construction, reanalysis, and stance: 'V yi ge N' and variations in Mandarin Chinese. *Journal of Pragmatics* 36.1655-72.

BLANK, ANDREAS. 2003. Polysemy in lexicon and discourse. *Polysemy: Flexible patterns of meaning in mind and language*, ed. by Brigitte Nerlich,Zazie Todd,Vimala Herman, and David D. Clarke, 267-93. New York: Mouton de Gruyter.

BLUST, ROBERT. 1999. Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. *Selected papers from the Eighth International Conference on Austronesian Linguistics*, ed. by Elizabeth Zeitoun and Paul Jen-kuei Li, 31-94. Taipei: Academia Sinica.

BLUST, ROBERT. 2002. Notes on the history of 'focus' in Austronesian languages. *The history and typology of western Austronesian voice systems*, ed. by Fay Wouk and Malcolm Ross, 63-78. Australia: Australian National University.

BOAS, FRANZ 1940. *Race, language, and culture*. New Work: Macmillan.

BOTNE, ROBERT. 2005. Cognitive schemas and motion verbs: COMING and GOING in Chindali (Eastern Bantu). *Cognitive Linguistics* 16.43-80.

BOWERMAN, MELISSA, and SOONJA CHOI. 2003. Space under construction: Language specific spatial categorization in first language acquisition. *Language in mind*, ed. by Dedre Gentener and Susan Goldin-Meadow, 387-428. Cambridge: MIT Press.

BOWERMAN, MELISSA. 1996. Learning how to structure space for language: A crosslinguistic perspective. *Language and space*, ed. by Paul Bloom, Mary A.

Peterson, Lyn Nadel, and Merrill F. Garret, 385-436. Cambridge: MIT Press.

BROWN, PENELOPE, and STEPHEN C. LEVINSON. 1987. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.

BRUGMAN, CLAUDIA. 1988. *Story of over: polysemy, semantics, and the structure of the lexicon*. New York: Garland.

BYBEE, JOAN; REVERE PERKINS; and WILLIAM PAGLIUCA. 1994. *The evolution of grammar: Tense, aspect and modality in the languages of the world*. Chicago: University of Chicago Press.

CAMPBELL, LYLE. 1998. *Historical linguistics: An introduction*. Edinburgh: Edinburgh University Press.

CARROLL, JOHN B. (ed.) 1956. *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. Cambridge: MIT Press.

CHAFE, WALLACE. 1994. *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.

CHANG, HENRY Y. 2004. The syntax of adverbial modifiers in Kavalan and Tsou. Paper presented at IsCLL-9, National Taiwan University, Taipei, November19-21.

CHEN, YI-TING. 2006. *Yi-dian* as a politeness device in Mandarin directives. Presented in International Association for Dialogue Analysis (IADA). Mainz, Germany, 9/25-29.

CHESTERMAN, ANDREW. 1998. *Contrastive functional analysis*. Amsterdam: John Benjamins.

CLAUSNER, TIMOTHY C., and WILLIAM CROFT. 1999. Domains and image schemas. *Cognitive Linguistics* 10.1-31.

COLEMAN, LINDA, and PAUL KAY. 1981. Prototype and semantics: the English *lie*. *Language* 57.26-44.

CROFT, WILLIAM, and D. ALAN CRUSE. 2004. *Cognitive linguistics*. Cambridge: Cambridge University Press.

CROFT, WILLIAM, and ESTHER J. WOOD. 2000. Construal operations in linguistics and artificial intelligence. *Meaning and cognition: A multidisciplinary approach*, ed. by Liliana Albertazzi, 51-78. Amsterdam; Philadelphia: John Benjamins.

CROFT, WILLIAM. 1991. *Syntactic categories and grammatical relations*. Chicago: Chicago University Press.

CROFT, WILLIAM. 1998. Linguistic evidence and mental representation. *Cognitive Linguistics* 9.151-73.

CROFT, WILLIAM. 2001. *Radical construction grammar*. Oxford: Oxford University Press.

CRUSE, D. ALAN. 1986. *Lexical semantics*. Cambridge: Cambridge University Press.

CRUSE, D. ALAN. 1990. Prototype theory and lexical semantics. *Meanings and prototypes: Studies in linguistic categorization*, ed. by S. L. Tsohatsidis, 382-402. London; New York: Routledge.

CRUSE, D. ALAN. 2000. *Meaning in language: An introduction to semantics and pragmatics*. Oxford: Oxford University Press.

DEANE, PAUL D. 1988. Polysemy and cognition. *Lingua* 75.325-61.

DOMINIEK, SANDRA, and SALLY RICE. 1995. Network analyses of prepositional meaning: Mirroring whose mind — the linguist's or the language user's? *Cognitive Linguistics* 6.89130.

DOWTY, DAVID. 1979. *Word meaning and Montague Grammar*. Dordrecht: Reidel.

DUBOIS, JOHN W.; STEPHAN SCHUETZE-COBUM; SUSANNA CUMMING; and DANAE PAOLINO. 1993. Outline of discourse transcription. *Talking data: Transcription and coding for language research*, ed. by J. A. Edwards and M. S. Lampert, 46-89. Hillsdale, NJ: Lawrence Erlbaum.

EVANS, VYVYAN, and MELANIE GREEN. 2006. *Cognitive linguistics: An introduction*. Edinburgh: Edinburgh University Press.

EVANS, VYVYAN. 2003. *The structure of time*. Amsterdam; Philadelphia: John Benjamins.

EVANS, VYVYAN. 2005. The meaning of *time*: Polysemy, the lexicon, and conceptual structure. *Journal of Linguistics* 41.33-75.

FAUCONNIER, GILLES, and MARK TURNER. 2002. *The way we think*. New York: Basic Books.

FAUCONNIER, GILLES. 1985. *Mental spaces: Aspects of meaning construction in natural language*. Cambridge, MA: MIT Press.

FILLMORE, CHARLES J. 1982. Frame semantics. *Linguistics in the morning calm*, ed. by I. Yang, 111-37. Seoul: Hanshin.

FILLMORE, CHARLES J., and B. T. ATKINS. 1992. Toward a frame-based lexicon. *Frames, fields, and contrasts*, ed. by Adrienne Lehrer and E. F. Kittay, 75-102. Hillsdale: Lawrence Erlbaum.

FILLMORE, CHARLES J.; PAUL KAY; and MARY CATHERINE O'CONNOR. 1988. Regularity and idiomacity in grammatical constructions: The case of *let alone*. *Language* 64.501-38.

FORD, CECILIA; BARBARA A. FOX; and SANDRA A. THOMPSON. 2003. Social interaction and grammar. *The new psychology of language, Vol. 2*, ed. by Michael Tomasello, 119-43. London: Lawrence Erlbaum.

FRITZ, GERD. 1998. *Historische Semantik*. Weimer: J. B. Metzler.

GEERAERTS, DIRK. 1989. Introduction: Prospects and problems of prototype theory.

Linguistics 27.587-612.

GODDARD, CLIFF, and ANNA WIERZBICKA. 2002. Semantic primes and universal grammar. *Meaning and universal grammar: Theory and empirical findings*, ed. by Anna Wierzbicka and Cliff Goddard, 41-85. Amsterdam: John Benjamins.

GODDARD, CLIFF, and ANNA WIERZBICKA. 2004. Cultural Scripts: What are they and what are they good for? *Intercultural Pragmatics* 1/2.153-65.

GODDARD, CLIFF. 2001. Lexical-semantic universals: A critical overview. *Linguistic Typology* 5.1-65.

GODDARD, CLIFF. 2002. The search for the shared semantic core of all languages. *Meaning and universal grammar: Theory and empirical findings, Vol. 1*, ed. by Cliff Goddard and Anna Wierzbicka, 5-40. Amsterdam: John Benjamins.

GOFFMAN, ERVING. 1981. *Forms of talk*. Philadelphia: University of Pennsylvania Press.

GOLDBERG, ADELE E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

GOLDBERG, ADELE E. 2006. *Constructions at work*. Oxford: Oxford University Press.

GRAUMANN, CARL F. 1990. Perspectival structure and dynamics in dialogues. *The dynamics of dialogue*, ed. by I. Markova and K. Foppa, 105-27. New York: Springer.

GRAUMANN, CARL F., and WERNER KALLMEYER. 2002. Perspective and perspectivization in discourse. *Perspective and perspectivization in discourse*, ed. by Carl F. Graumann and Werner Kallmeyer, 1-14. Amsterdam: John Benjamins.

GREENBERG, JOSEPH H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. *Universal of Language*, ed. by Joseph H. Greenberg, 73-113. Cambridge, Mass: MIT Press.

GUMPERZ, JOHN J., and STEPHEN C. LEVINSON (eds.) 1996. *Rethinking linguistic relativity*. Cambridge: Cambridge University Press.

GYORI, GABOR. 2002. Semantic change and cognition. *Cognitive Linguistics* 13.123-66.

HAN, YUKUO. 2004. Hanyu fuci you de qiyi—jian tan fanchou yufa dui hanyu yianjiu de shiyongxing (The polysemous structure of Mandarin adverb *you*: the applicability of categorization theory in Mandarin Linguistic research) [in Chinese]. *Academic Journal of Yunnan Normal University* 2.64-69.

HARTMANN, R. R. K., and GREGORY JAMES. 1998. *Dictionary of Lexicography*. London: Taylor & Francis.

HASPELMATH, MARTIN. 1997. *Indefinite pronouns*. New York: Oxford University Press.

HASPELMATH, MARTIN. 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. *The new psychology of language, Vol 2*, ed. by Michael Tomasello, 211-42. Mahwah, NJ: Lawrence Erlbaum.

HEINE, BERND; ULRIKE CLAUDI; and FRIEDERIKE HUENNEMEYER. 1991. *Grammaticalization: A conceptual framework*. Chicago: University of Chicago Press.

HIMMELMANN, NIKOLAUS P. 1998. Documentary and descriptive linguistics Linguistics. *Linguistics* 36.161-95.

HIMMELMANN, NIKOLAUS P. 2002. Voice in western Austronesian: An update. *The history and typology of western Austronesian voice systems*, ed. by Fay Wouk and Malcolm Ross, 7-16. Australia: Australian National University.

HO, DAH-AN, and SIU-FANG YANG. 2000. Austronesian Languages and Formosan Languages. *Formosan Language Series*, 1-36. Taipei: Yuan-liu. [in Chinese].

HOPPER, PAUL J., and ELIZABETH CLOSS TRAUGOTT. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.

HORN, LAURENCE R. 1989. *A natural history of negation*. Chicago: The University of Chicago Press.

HSIEH, FUHUI, and SHUANFAN HUANG. 2006. The pragmatics of case marking in Saisiyat. *Oceanic Linguistics* 45.91-109.

HSIEH, FUHUI. 2007. *Language of emotion and thinking in Kavalan and Saisiyat*. Doctoral dissertation, National Taiwan University.

HUANG, LILLIAN M. 2000. Verb classification in Mayrinax Atayal. *Oceanic Linguistics* 39.364-90.

HUANG, SHUANFAN. 1998. Emergent lexical semantics. *Selected Papers from the Second International Symposium in Language in Taiwan*, ed. by Shuanfan Huang, 129-50. Taipei: Crane.

HUANG, SHUANFAN; LILY I-WEN SU; and LI-MAY SUNG. 2004. *A Functional Reference Grammar of Saisiyat*. NSC technical report (NSC 92-2411-H-002-077).

JACKENDOFF, RAY. 1987. *Consciousness and the computational mind*. Cambridge: MIT Press.

JASZCZOLT, KATARZYNA M. 2003. On translating 'What is said': Tertium comparationis in contrastive semantics and pragmatics. *Meaning Through Language Contrast*, ed. by Katarzyna M. Jaszczolt and Ken Turner, 441-62. Amsterdam: John Benjamins.

JOHNSON, MARK. 1990. *The body in the mind: The bodily basis of meaning, imagination, and reason*. Chicago: University of Chicago Press.

KEMMER, SUZANNE. 1993. *The middle voice*. Amsterdam: John Benjamins.

KOFFKA, KURT. 1935. *Principles of gestalt psychology*. London: Routledge and Kegan

Paul.

KRZESZOWSKI, TOMASZ P. 1984. Tertium comparationis. *Contrastive linguistics: Prospects and problems*, ed. by Jacek Fisiak, 301-13. Berlin: Mouton.

KRZESZOWSKI, TOMASZ P. 1990a. *Contrasting languages: The scope of contrastive linguistics*. Berlin: Mouton de Gruyter.

KRZESZOWSKI, TOMASZ P. 1990b. Prototypes and equivalence. *Further Insights into Contrastive Analysis*, ed. by Jacek Fisiak. Amsterdam; Philadelphia: John Benjamins.

LÖBNER, SEBASTIAN. 2002. *Understanding semantics*. London: Arnold.

LABOV, WILLIAM. 1973. The boundaries of words and their meaning. *New ways of analyzing variation of English*, ed. by Charles-James N. Bailey and Roger W. Shuy, 340-73. Washington D. C.: Georgetown University Press.

LAKOFF, GEORGE, and MARK JOHNSON. 1980. *Metaphors we live by*. Chicago: University of Chicago Press.

LAKOFF, GEORGE, and MARK TURNER. 1989. *More than cool reason: A field guide to poetic metaphor*. Chicago: Chicago University Press.

LAKOFF, GEORGE. 1968. Instrumental adverbs and the concept of deep structures. *Foundations of Language* 4.4-29.

LAKOFF, GEORGE. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: Chicago University Press.

LANGACKER, RONALD W. 1985. Observation and speculation on subjectivity. *Iconicity in syntax*, ed. by John Heiman, 109-50. Amsterdam: John Benjamins.

LANGACKER, RONALD W. 1987a. *Foundations of cognitive grammar 1: Theoretical prerequisites*. Stanford: Stanford University Press.

LANGACKER, RONALD W. 1987b. A usage-based model. *Topics in cognitive linguistics*, ed. by Brygida Rudzka-Ostyn, 127-61. Amsterdam: John Benjamins.

LANGACKER, RONALD W. 1990. Subjectification. *Cognitive Linguistics* 1.5-38.

LANGACKER, RONALD W. 1991. *Foundations of cognitive grammar 2: Descriptive application*. Stanford: Stanford University Press.

LANGACKER, RONALD W. 1998. Conceptualization, symbolism and grammar. *The new psychology of language: Cognitive and functional approaches to language structure*, ed. by Michael Tomasello, 1-39. Mahwah, N. J.: Lawrence Erlbaum.

LANGACKER, RONALD W. 1999. *Grammar and conceptualization*. Berlin, New York: Mouton de Gruyter.

LEHMANN, CHRISTIAN. 1982. Directions for interlinear morphemic translations. *Folia Linguistica* 16.199-224.

LEHMANN, CHRISTIAN. 1985. Grammaticalization: Synchronic variation and diachronic change. *Lingua e Stile* 20.303-18.

LEHRER, ADRIENNE. 1992. A theory of vocabulary structures: Retrospectives and prospectives. *Thirty years of linguistic evolution*, ed. by Martin Putz. Amsterdam: John Benjamins.

LEVIN, beth. 1993. Towards a lexical organization of English verbs. Chicago: University of Chicago Press.

LEVINSON, STEPHEN C. 1997. From outer to inner space: Linguistic categories and non-linguistic thinking. *Language and conceptualization*, ed. by E. Pederson and J. Nuyts, 13-45. Cambridge: Cambridge University Press.

LEVINSON, STEPHEN C. 2000. *Presumptive meanings*. Cambridge, MA: MIT Press.

LI, CHARLES N., and SANDRA A. THOMPSON. 1981. *Mandarin Chinese: A functional reference grammar*. Berkley: University of California Press.

LI, PAUL JEN-KUEI. 1978. A comparative vocabulary of Saisiyat dialects. *Bulletin of the Institute of History and Philology* 49.133-99.

LI, PAUL JEN-KUEI. 1997. *Taiwan nandao minzu de zucun yu ciansi (The ethnic group and migration of Formosan people)*. Taipei: Changmin.

LI, PAUL JEN-KUEI. 2006. Numerals in Formosan languages. *Oceanic linguistics* 45.133-52.

LICHTENBERK, FRANTISEK. 1985. Multiple uses of reciprocal situations. *Australian Journal of Linguistics* 5.19-41.

LIPKA, LEONHARD. 1987. Prototype Semantics or Features Semantics: An Alternative? *Perspectives on language in performance*, ed. by W. Lörscher and R. Schulze, 282-98. Tübingen: Niemeye.

LIU, FENG-HSI. 2001. The scalar particle HAI in Chinese. *Cahiers de Linguistique-Asie Orientale* 29.41-84.

LÖBNER, SEBASTIAN. 2002. *Understanding semantics*. London: Arnold.

LYONS, JOHN. 1977. *Semantics*. Cambridge: Cambridge University Press.

MACLAURY, ROBERT E. 1991. Prototypes revisited. *Annual Review of Anthropology* 20.55-74.

MALINOWSKI, BRONISLAW. 1938. The problem of meaning in primitive languages. *The meaning of meaning*, ed. by Charles K. Ogden and I. A. Richards, 296-336. New York: Harcourt, Brace & World.

MATTHEWSON, LISA. 2004. On the methodology of semantic fieldwork. *International Journal of American Linguistics* 70.369-415.

MITHUN, MARIANNE. 2001. Who shapes the record: The speaker and the linguist. *Linguistic fieldwork*, ed. by Paul Newman and Martha Ratliff, 34-54. Cambridge: Cambridge University Press.

NERLICH, BRIGITTE, and DAVID D. CLARKE. 2003. Polysemy and flexibility: Introduction and overview. *Polysemy: Flexible patterns of meaning in mind and*

language, ed. by Brigitte Nerlich,Zazie Todd,Vimala Herman, and David D. Clarke, 3-10. Berlin; New York: Mouton de Gruyter.

PALMER, FRANK. 1976. *Semantics: A new outline*. Cambridge: Cambridge University Press.

PALMER, FRANK. 2001. *Mood and modality* New York: Cambridge University Press.

PAN, WENGUO, and WAI MUN THAM. 2007. *Contrastive linguistics*. New York: Continuum.

PUSTEJOVSKY, JAMES. 1998. *The generative lexicon*. Cambridge: MIT Press.

RIEMER, NICK. 2001. Interpreting semantic extension: Metaphor and metonymy on different levels of lexical categorization. The 2001 Conference of the Australian Linguistic Society. Canberra, Australia

ROSCH, ELEANOR. 1978. Principles of categorization. *Cognition and categorization*, ed. by Eleanor Rosch and B. Lloyd, 27-48. Hillsdale: Erlbaum Associates.

ROSCH, ELEANOR. 1983. Prototype classification and logical classification: The two systems. *New trends in cognitive representation: Challenges to Piaget's theory*, ed. by E. Scholnick, 73-86. Hillsdale, NJ: Lawrence Erlbaum Associates.

ROSS, MALCOLM, and STACY FANG-CHING TENG. 2005. Formosan languages and linguistic typology. *Languages and Linguistics* 6.739-81.

RUHL, CHARLES. 1989. *On monosemy: A study in linguistic semantics*. Albany: State University of New York.

RUHL, CHARLES. 2002. Data, comprehensiveness, monosemy. *Signal, meaning, and message: Perspectives on sign-based linguistics*, ed. by Wallis Reid Ricardo Otheguy, and Nancy Stern. Amsterdam: John Benjamins.

SAEED, JOHN. 1997. *Semantics*. Oxford ; Cambridge: Blackwell.

SAMARIN, WILLIAM J. 1967. *Field linguistics: A guide to linguistic field work*. New York: Holt, Rinehart and Winston.

SANDERS, TED J. M.; WILBERT P. M. SPOOREN; and LEO G. M. NOORDMAN. 1992. Toward a taxonomy of coherence relations. *Discourse Processes* 15.1-35.

SANFORD, ANTHONY, and SIMON GARROD. 1981. *Understanding written language*. Chichester: John Wiley & Sons.

SCHANK, ROGER C., and ROBERT ABELSON. 1977. *Scripts, plans, goals, and understanding: An inquiry into human knowledge*. Hillsdale: Lawrence Erlbaum.

SLOBIN, DAN I. 1996. From "thought to language" to "thinking for speaking". *Rethinking linguistic relativity*, ed. by J. John Gumperz and Stephen C. Levinson, 70-96. Cambridge: Cambridge University Press.

SPERBER, DANIEL, and DEIRDRE WILSON. 1995. *Relevance: Communication and cognition*. Oxford: Blackwell.

STAROSTA, STANLEY. 1995. A grammatical subgrouping of Formosan languages. *Austronesian studies relating to Taiwan*, ed. by Paul Jen-kuei Li,Dah-an Ho,Ying-kuei Huang, and Cheng-hwa Tsang. Taipei: Academia Sinica.

SU, LILY I-WEN, and SHUPING HUANG. 2006. Polysemy and Categorization: Implications to Corpus Documentation. *Streams Converging into An Ocean: Festschrift in Honor of Professor Paul Jen-Kuei Li on His 70th Birthday*, ed. by Henry Y. Chang,Lillian M. Huang, and Dah-an He, 415-38. Taipei: Academia Sinica.

SU, LILY I-WEN. 1998. *Grounding and coherence in Chinese discourse*. Taipei: Crane.

SU, LILY I-WEN; LI-MAY SUNG; SHUPING HUANG; FUHUI HSIEH; and ZHEMIN LIN. forthcoming. NTU Corpus of Formosan Languages: A state-of-the-art report. *Corpus Linguistics and Linguistic Theory*.

SWEETSER, EVE. 1986. Polysemy vs. abstraction: Mutually exclusive or complementary? *Berkeley Linguistic Society* 12.528-38.

SWEETSER, EVE. 1990. *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press.

TALMY, LEONARD. 1977. Rubber sheet cognition in language. *Papers from the Thirteenth Regional Meeting, Chicago Linguistic Society* 13.612-28.

TALMY, LEONARD. 1978. The relation of grammar to cognition. *Proceedings of TINLAP-2: Theoretical issues in natural language processing* ed. by David Waltz, 14-24. Urbana: University of Illinois Coordinated Science Laboratory.

TALMY, LEONARD. 1983. How language structures space. *Spatial orientation: Theory, research, and application*, ed. by H. Pick and L. Acredolo, 225-82. New York: Plenum Press.

TALMY, LEONARD. 1988. Force dynamics in language and cognition. *Cognitive Science* 12.49-100.

TANNEN, DEBORAH, and CYNTHIA WALLET. 1993. Interactive frames and knowledge schemas in interaction. *Framing in discourse*, ed. by Deborah Tannen, 57-76. Oxford: Oxford University Press.

TAYLOR, JOHN. 1990. Schemas, prototypes, and models: In search of the unity of the sign. *Meanings and prototypes: Studies in linguistic categorization*, ed. by Savas L. Tsohatzidis, 521-34. London; New York: Routledge.

TAYLOR, JOHN. 1995a. *Linguistic categorization: Prototypes in linguistic theory, 2nd edition*. Oxford: Clarendon Press.

TAYLOR, JOHN. 2002. *Cognitive grammar*. Oxford: Oxford University Press.

TAYLOR, JOHN. 2003. Polysemy's paradoxes. *Language Sciences* 25.637-55.

TOMASELLO, MICHAEL. 2003. The key is social cognition. *Language in mind*, ed. by Dedre Gentener and Susan Goldin-Meadow, 47-57. Cambridge: MIT Press.

TRAUGOTT, ELIZABETH CLOSS, and RICHARD B. DASHER. 2002. *Regularity in semantic change*. Cambridge: Cambridge University Press.

TRAUGOTT, ELIZABETH CLOSS. 1999. From subjectification to intersubjectification. the Workshop on Historical Pragmatics (14th International Conference on Historical Linguistics). Vancouver, Canada

TRAUGOTT, ELIZABETH CLOSS. 2003. From subjectification to intersubjectification. *Motives for language change*, ed. by Raymond Hickey, 124-39. Cambridge, UK: Cambridge University Press.

TSAI, HUI CHIN. 2006. Additive adverbs and coordinators in Chinese: A case study of *ye* 'also', *erqie* 'and' and *he* 'and'. *UST Working Papers in Linguistics* 2.67-95.

TSOHATZIDIS, SAVAS L. (ed.) 1990. *Meanings and prototypes: Studies in linguistic categorization*. London; New York: Routledge.

TUGGY, DAVID H. 1993. Ambiguity, polysemy, and vagueness. *Cognitive Linguistics* 4.273-90.

TYLER, ANDREA, and VYVYAN EVANS. 2001. The relation between experience, conceptual structure and meaning: Non-temporal uses of tense and language teaching. *Applied cognitive linguistics I: Theory and language acquisition*, ed. by Martin Putz, Susanne Niemeier, and Rene Dirven, 63-108. Berlin: Mouton de Gruyter.

TYLER, ANDREA, and VYVYAN EVANS. 2003. *The semantics of English prepositions: Spatial scenes, embodied meaning and cognition*. Cambridge: Cambridge University Press.

TYLER, ANDREA, and VYVYAN EVANS. 2004. Applying cognitive linguistics to pedagogical grammar: The case of *over*. *Cognitive linguistics, second language acquisition, and foreign language teaching*, ed. by Michel Achard and Susanne Niemeier, 257-80. Berlin: Mouton de Gruyter.

ULLMANN, STEPHEN. 1957. *The principles of semantics, 2nd ed.* Glasgow: Jackson.

ULLMANN, STEPHEN. 1972. *Semantics*. Oxford: Blackwell.

UNGERER, FRIEDRICH, and HANS-JORG SCHMID. 1996. *An introduction to cognitive linguistics*. Essex: Longman.

VAN DER AUWERA, JOHAN, and VLADIMIR PLUNGIAN. 1998. Modality's semantic map. *Linguistic Typology* 2.79-124.

VAUX, BERT, and JUSTIN COPPER. 1999. *Introduction to linguistic field methods*. Munchen: Lincom.

VERVAEKE, JOHN, and CHRISTOPHER D. GREEN. 1997. Women, fire, and dangerous theories: A critique of Lakoff's theory of categorization. *Metaphor and Symbol* 12.59-80.

WANG, LI. 1947. *Zhongguo xiandai yufa* (Modern Chinese grammar). Shanghai:

Zhonghua Bookstore.

WEINREICH, URIEL. 1962. Problems in lexicography. *International Journal of American Linguistics* 28.25-43.

WIERZBICKA, ANNA. 1972. *Semantic primitives*. Frankfurt am Mein: Athenäum.

WIERZBICKA, ANNA. 1996. *Semantics: Primes and universals*. Oxford: Oxford University Press.

WIERZBICKA, ANNA. 1999. *Emotions across language and culture: Diversity and universals*. Cambridge: Cambridge University Press.

WIERZBICKA, ANNA. 2002. Semantic primes and linguistic typology. *Meaning and universal grammar: Theory and Empirical Findings*, ed. by Cliff Goddard and Anna Wierzbicka, 257-301. Amsterdam: John Benjamins.

WILSON, DEIRDRE. 2000. Metarepresentation in linguistic communication. *Metarepresentations*, ed. by Daniel Sperber, 411-48. Oxford; New York: Oxford University Press.

WITTGENSTEIN, LUDWIG. 1963. *Philosophical investigations*. New York: Macmillan.

YEH, MARIE M. 2000a. *Saixiayu cankau yufa* (A Reference Grammar of Saisiyat). Taipei: Yuanliu [in Chinese].

YEH, MARIE M. 2000b. Nominalization in Saisiyat. *Paper presented at the Workshop on Nominalization in Formosan Languages*. Taipei: Academia Sinica.

YEH, MARIE M. 2003. A syntactic and semantic study of Saisiyat verbs, Doctorate thesis. Graduate Institute of English, National Taiwan Normal University.

YEH, MENG. 1998. On *hai* in Mandarin. *Journal of Chinese Linguistics* 26.236-80.

ZEITOUN, ELIZABETH, and CHING-HUA YU. 2005. The Formosan Language Archive: Linguistic analysis and language processing. *Computational Linguistics and Chinese Language Processing* 10.167-200.

ZEITOUN, ELIZABETH, and LILLIAN M. HUANG. 2000. Discussion on *ka-*: An overlooked marker of verbal derivation in the Formosan languages. *Oceanic Linguistics* 39.391-414.

ZEITOUN, ELIZABETH. 2000. Dynamic vs. stative verbs in Mantauran (Rukai). *Oceanic Linguistics* 39.415-27.

Dictionaries

Jìaoyùbù Gúoyǔ Xǐao Zìdǐan (Miniature Dictionary of the Ministry of Education)
   http://dict.mini.moe.edu.tw/cgi-bin/gdic/gsweb.cgi?o=ddictionary

Gúoyǔrìbào Cídǐan (Mandarin Daily News Dictionary)
   1978. Taipei: Mandarin Daily News.

Shìyi Xúesheng Cídǐan (Student Dictionary published by Shiyi Bookstore).
   2005. Taipei: Shiyi Bookstore.

Appendix A.

Typology of *tertium comparationist* (sketched in Fisiak 1990:8):

Contrastive Analyses

text-bound   systematic

-translatoinally   +translationally   structurally   functionally
equivalent texts   equivalent texts   constrained   constrained

quantitative   qualitative

systems   ⇔   constructions   ⇔   rules
(paradigmatic)

**statistical   translation   system   semanto-syntactic   rule   pragmatic
equivalence   equivalence   equivalence   equivalence   equivalence   equivalence**

phonological   lexical contrastive
contrastive analyses   analyses

substantial equivalence

Appendix B.
Coding of grammatical elements in NTU Corpus of Formosan Languages

Extracted from http://corpus.linguistics.ntu.edu.tw/coding.php

Code	中文		Note
1SG	1SG	1st person singular	
2SG	2SG	2nd person singular	
3SG	3SG	3rd person singular	
1IPL.NOM	1IPL.主格	1st person plural, inclusive, nominative	
1EPL.NOM	1EPL.主格	1st person plural, exclusive, nominative	
1PL	1PL	1st person plural	
2PL	2PL	2nd person plural	
3PL	3PL	3rd person plural	
NOM	主格	Nominative	
ACC	受格	Accusative	
AF	主焦	Agent Focus	
ASP	動貌	Aspect	
AUX	助動詞	Auxiliary	
BC	BC	Back Channel / Reactive Token	
BF	予焦	Benefactive Focus	
Ca	Ca 重疊	Ca Reduplication	
CAU	使役	Causative	
CLF	量詞	Classifier	
CLF.HUM	人量詞	Human Classifier	Kavalan: kin-zusa
CLF.NHUM	非人量詞	Non-human Classifier	Kavalan: u-zusa
COMP	補語詞	Complementizer	
COND	條件詞	Conditional Marker	
CONJ	連接詞	Conjunctor	
DAT	予格	Dative	
DEF	定指	Definite	
DET	限定詞	Determiner	
DIST	遠距	Distal	
DM	DM	Discourse Marker	
EVI	知識詞	Evidential	
EXCL	排除	Exclusive	
EXIST	存在	Existential	
EXPER	經驗	Experiential	

FIL	FIL	Pause Filler	
FS	FS	False Start	
FUT	未來	Future	
HAB	習慣	Habitual	
GEN	屬格	Genetive	
IF	工焦	Instrumental Focus	
IMP	祈使	Imperative	
INCL	包含	Inclusive	
INDF	不定指	Indefinite	
INS	工具格	Instrument	
INT	感嘆	Interjection	
INVIS	不可見	Invisible	
IRR	非實現	Irrealis	
LF	處焦	Locative Focus	
<L2H L2H>		L2H=Hakka	
<L2J L2J>		L2J=Japanese	
<L2M L2M>		L2M=Mandarin	
<L2T L2T>		L2T=Taiwanese	
LNK	連詞	Linker	
LOC	處格	Locative	
NCM	NCM	Non-common Name Marker	
NAF	非主焦	Non-agent Focus	
NEG	否定	Negative	
NEU	中性格	Neutral	
NMZ	名物化	Nominalizer/Nominalization	
NOM	主格	Nominative	
NRFUT	即將	Near Future	
OBL	斜格	Oblique	
PART	語助詞	Particle	
PF	受焦	Patient Focus	
PFV	完成	Perfective	
PN	人名/地名/專名	proper name/place name	
POSS	所有格	Possessive	
PROG	進行	Progressive	
PROX	近距	Proximal/Proximate	
Q	疑問	Question Marker	

QUOT	QUOT	Quotative	
REC	交互	Reciprocal	
RED	重疊	Reduplication	
REL	關係詞	Relativizer	
REFL	反身	Reflexive	
RF	指焦	Referential Focus	
TOP	主題	Topic	
VIS	可見	Visible	
VOC	呼格	Vocative	
X	X	Uncertain Hearing	
??	??	Uncertain Meaning	
this	這個		
that	那個		

Appendix C:
Coding of discourse elements in NTU Corpus of Formosan Languages

Extracted from http://corpus.linguistics.ntu.edu.tw/coding.php

Units	
Intonation Unit	a new line
Word	a blank space
Truncated word	- -
Speaker identity / turn start	:
Speech Overlap	[ ]
Transitional Continuity	
Final	.
Continuing	,
Appeal	?
Terminal Pitch Direction	
Fall	\
Rise	/
Level	_
Accent and Lengthening	
Primary accent	^
Secondary accent	`
High booster	!
Low booster	;
Lengthening	==
Tone	
Fall	\
Rise	/
Fall-Rise	V
Rise-fall	Λ
Level	_
Pause	
Long (>0.7sec)	...(N)
Medium (0.3-0.6 sec)	...
Short (<0.2 sec)	..
Latching	(0)
Vocal Noises	
Vocal noises	(CAPITAL LETTERS)
Inhalation	(H)
Exhalation	(Hx)
Glottal stop	%

Laughter	@
Quality	
Quality	<Y Y>
Laugh quality	<@ @>
Quotation quality	<Q Q>
Phonetics	
Phonetic / phonemic transcription	(/ /)
Transcriber's Perspective	
Researcher's comment	(( ))
Uncertain hearing	<X X>
Indecipherable syllable	X
Specialized Notations	
Duration	(N)
IU boundary	&
Accent unit boundary	|
Embedded IU	<| |>
Restart	{Capital Initial}
False start	< >
Code switching	<L2 L2>
Nontranscription line	$
Reserved Symbols	
Phonetic / orthographic symbols	'
Morphosyntactic coding	+ * # { }
User-definable symbols	" ~

Appendix D:

Definitions of semi-senses and non-discrete meanings in Croft and Cruse (2004)

SEMI-SENSES

Facets: Facets are distinguishing components of a global whole, yet not capable of being subsumed under a hyperonym. For example, *book* refers to a global concept BOOK which contains facets [TOME] and [TEXT]. Though the two facets are theoretically isolatable respectively in *a red book* and *an insightful book*, their "joint compositional properties" (Croft and Cruse 2004: 120) are evidenced in that there are many predicates that activate both facets simultaneously as in *to publish a book*.

Subsenses: Subsenses are also known as **microsenses**, which refer to "distinct sense units of a word that occur in different contexts and whose default construals stand in a relation of mutual incompatibility at the same hierarchical level" (Croft and Cruse 2004: 127). For example, *knife* denotes a collective notion that contains weapons as in *He threatened the couple with a knife* and cutlery as in *He asked the waiter for a knife and fork*. But such specificities can be subsumed under the same hyperonym as in *You can buy all sorts of knife in this shop*.

NON-DISCRETE SENSES

Ways-of-seeing: Ways-of-seeing are also called **perspectives**, which refers to different ways of looking at the same thing. For example, hotel refers to a global concept HOTEL, but we can view a hotel from its part-whole relation, kind-of relation, function, and its life history.

Active zones: Active zones and **semantic components** are also known as **local senses**. They are "units whose sole manifestation of autonomy is compositional; they do not constitute pre-meanings in the full sense" (Croft and Cruse 2004: 138). Active zones

are extensional components that are isolated for compositional purposes. For example, a *red pencil* may have a red lead or a red casing. And the lead and casing are conceptually extensional and referential (i.e., can be pointed to). On the other hand, semantic components are intensional properties. For example, *an overworked stallion* is either overworked as EQUINE or as MALE. Such properties are inclusive parts of a global whole which are compositionally active.

Contextual modulations: Contextual modulations have the least autonomy and the meaning modulation is situated only in a specific context. For example, *My cousin married an actress* immediately triggers the interpretation that the cousin is male. Such meaning is not coded, and can be said to be pragmatically motivated to make sense of an utterance.
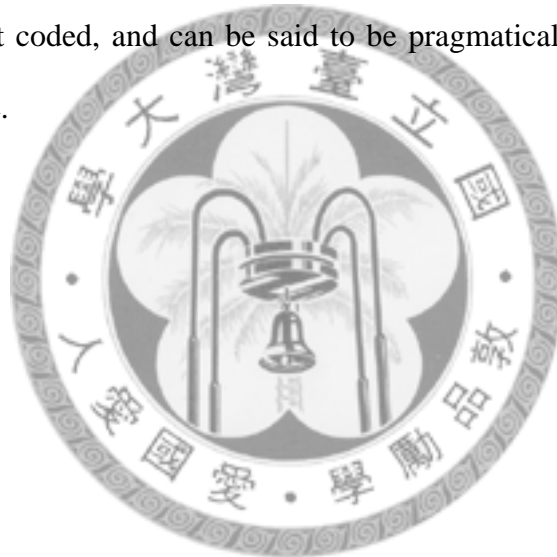
Appendix E.

Dictionary meanings of relevant Chinese lexemes (ordered by Roman alphabet)

I. *cái* **(adv.)**

	教育部國語小字典	國語日報辭典	世一學生辭典	**synthesized**
1	a moment ago (adv.)	a moment ago (adv.)	a moment ago (adv.)	**a moment ago (adv.)**
2	a quantity less than expectation (adv.)	a quantity less than expectation (adv.)	a quantity less than expectation (adv.)	**a quantity less than expectation (adv.)**
3	an emphatic marker to counter a statement (adv.)	limited condition for something to be realized (adv.)	limited condition for something to be realized (adv.)	**limited condition for something to be realized (adv.)**
4				**an emphatic marker to counter a statement (adv.)**

II. *hái*

	教育部國語小字典	國語日報辭典	世一學生辭典	**synthesized**
1	continuation of an activity/state (adv.)	not yet (adv.)	to remain in situ (adv.)	**continuation of an activity (adv.)**
2	higher degree (adv.)	continuation of a state/activity (adv.)	addition of more activities (adv.)	**higher degree (adv.)**
3	repetition of an activity (adv.)	higher degree (adv.)	continuation of an activity (adv.)	**addition of more activities (adv.)**
4	to passably meet a standard (adv.)	addition of more activities (adv.)	higher degree (adv.)	**not yet (adv.)**
5	to indicate a surprise (adv.)	offering several choices (collocating with copula *shì*) (adv.)	offering several choices (collocating with copula *shì*) (adv.)	**to remain in situ (adv.)**
6	offering several choices (collocating with copula *shì*) (adv.)			**offering several choices (adv.)**
7				**repetition of an activity (adv.)**
8				**to passably meet a standard (adv.)**
9				**to indicate a surprise (adv.)**

III. *ling(wài)*

	教育部國語小字典	國語日報辭典	世一學生辭典	**synthesized**
1	other (adj./adv.)	other, in addition (adv.)	other, in addition (adj./adv.)	**other, in addition (adj./adv.)**
2	separated(ly) (adj./adv.)		separated(ly) (adj./adv.)	**separated (adj./adv.)**

IV. *réngrán*

	教育部國語小字典	國語日報辭典	世一學生辭典	**synthesized**
1	as before (adv.)	as before (adv.)	as before (adv.)	**as before (adv.)**

V. (*shǒu)xian* (adv.)

	教育部國語小字典	國語日報辭典	世一學生辭典	**synthesized**
1	in advance, temporally preceding (adv.)	temporally preceding (adv.)	in advance (adv.)	**in advance, temporally preceding (adv.)**

VI. *yě* (adv.)

	教育部國語小字典	國語日報辭典	世一學生辭典	**synthesized**
1	an expression resembling a preceding one (adv.)	an expression resembling a preceding one (adv.)	an expression resembling a preceding one (adv.)	**an expression resembling a preceding one (adv.)**
2	all (collocating with interrogatives) (adv.)	all (collocating with interrogatives) (adv.)	all (collocating with interrogatives) (adv.)	**all (collocating with interrogatives) (adv.)**
3	to passably meet a standard (adv.)	to passably meet a standard (adv.)	to passably meet a standard (adv.)	**to passably meet a standard (adv.)**
4	emphatics (collocating with *yi-diǎn* 'a bit') (adv.)	a disjunctive marker signaling counter-expectation (adv.)	a disjunctive marker signaling counter-expectation (adv.)	**a disjunctive marker signaling counter-expectation (adv.)**
5	a disjunctive marker signaling counter-expectation (adv.)	to be not too bad (adv.)		**emphatics (collocating with *yi-diǎn* 'a bit') (adv.)**
6		coordination of two clauses to mitigate a judgment (adv./conj.)		**to be not too bad (adv.)**
7		a marker that makes to conjoined expressions resonate and coherent (conj.)		**coordination of two clauses to mitigate a judgment (adv./conj.)**
8				**a marker that makes to conjoined expressions resonate and coherent (conj.)**

VII. *yi-hǔir*

	教育部國語小字典	國語日報辭典	世一學生辭典	**synthesized**
1	instantly, in a short time instantly (adv.)	in a short time (adv.)	in a short time (adv.)	**instantly (adv.)**

VIII. *yòu*

	教育部國語小字典	國語日報辭典	世一學生辭典	**synthesized**
1	repetition of an activity (adv.)	repetition of an activity (adv.)	repetition of an activity (adv.)	**repetition of an activity (adv.)**
2	coordination of several activities/properties (adv.)	coordination of several activities/properties (adv.)	coordination of several activities/properties (adv.)	**coordination of several activities/properties (adv.)**
3	emphatics of negations (adv.)	succession of activities (adv.)		**emphatics of negations (adv.)**
4	a higher degree (adv.)	emphatics of negations (adv.)		**succession of activities (adv.)**
5	succession of activities	a higher degree		**a higher degree**
6	addition of a fraction to an integer (adv.)	addition of a fraction to an integer (adv.)		**addition of a fraction to an integer (adv.)**

IV. *yúanběn*

	教育部國語小字典	國語日報辭典	世一學生辭典	**synthesized**
1	formerly (adj./adv.)	formerly (adj./adv.)	formerly, at the beginning (adj./adv.)	**formerly (adj./adv.)**

X. *zài*

	教育部國語小字典	國語日報辭典	世一學生辭典	**synthesized**
1	repetition of an activity (adv.)	repetition of an activity (adv.)	repetition of an activity (adv.)	**repetition of an activity (adv.)**
2	succession of an activity (adv.)	to introduce an activity to happen in the future (adv.)	a higher degree (adv.)	**a higher degree (adv.)**
3	continuation of an activity (adv.)	a higher degree (adv.)		**succession of an activity (adv.)**
4	a higher degree (adv.)			**to introduce an activity to happen in the future (adv.)**
5				**continuation of an activity (adv.)**

XI. *zhǐ*

	教育部國語小字典	國語日報辭典	世一學生辭典	**synthesized**
1	limited scope on specific things/activities	limited scope on specific things/activities	limited scope on specific things/activities	**limited scope on specific things/activities**
2	to do one thing with all one's might and neglect other things	to do one thing with all one's might and neglect other things	a disjunctive marker (collocating with copula *shì*)	**to do one thing with all one's might and neglect other things**
3			to have no other choice but to…(collocating with *hǎo* 'good')	**a disjunctive marker (collocating with copula *shì*)**
4				**to have no other choice but to… (collocating with *hǎo* 'good')**

271