Current Studies in Chinese Language and Discourse

Edited by

Yun Xiao and Linda Tsung

10

John Benjamins Publishing Company

Current Studies in Chinese Language and Discourse

Studies in Chinese Language and Discourse (SCLD)

The *Studies in Chinese Language and Discourse* book series publishes works of original research on Chinese from a linguistic, cognitive, socio-cultural, or interactional perspective. We welcome contributions based on systematic documentation of language structure which displays fresh data and analysis from such areas as corpus linguistics, grammaticalization, cognitive linguistics, sociolinguistics, discourse and grammar, conversation analysis, and typological and comparative studies. Both monographs and thematic collections of research papers will be considered.

For an overview of all books published in this series, please see
http://benjamins.com/catalog/scld

Executive Editor

Hongyin Tao
University of California, Los Angeles

Co-editors

K.K. Luke
Nanyang Technological University

Li Wei
UCL Institute of Education

Volume 10

Current Studies in Chinese Language and Discourse
Global context and diverse perspectives
Edited by Yun Xiao and Linda Tsung

Current Studies in Chinese Language and Discourse

Global context and diverse perspectives

Edited by

Yun Xiao
Bryant University

Linda Tsung
The University of Sydney

John Benjamins Publishing Company

Amsterdam / Philadelphia

Table of contents

Contributors

Hock Huan Goh is currently the Chair of Research and Ethics Committee and Research Scientist at the Singapore Centre for Chinese Language Limited, a centre of Nanyang Technological University (NTU, Singapore). He obtained a Ph.D. degree in Education from the National Institute of Education (NTU, Singapore). He has led four tier-2 government-funded research projects since 2010, which involved corpora building, corpus-based text-grading and national curriculum evaluation. Apart from these projects, his academic interests include Chinese language education in multi/bilingual context, child Mandarin competence, corpus-based applications, and Chinese linguistics. He has compiled and co-compiled three frequency dictionaries of Singapore daily-used Chinese characters, words and sentence patterns, and published a book on Mandarin competence of Singaporean Chinese preschoolers plus various papers and book chapters on Chinese Mandarin in Singapore.

Chenhsuan Huang received her Ph.D. from National Chung Cheng University, Taiwan. Her research interests include sociolinguistics, cognitive semantics, corpus linguistics and language pedagogy. Her dissertation focuses on the Chinese classifiers variation from a cognitive sociolinguistics perspective. One of the chapters from her dissertation, "Semantic Variation on Chinese Classifiers for Vehicles: A Case Study on *liang* (輛), *bu* (部), *tai* (台)," was published in one of the prestigious journals in Chinese language teaching in Taiwan. Besides, she has also taught Chinese language at universities in Taiwan for many years.

Meishu He is a postgraduate research student in Centre for Education and Psychology Research on Ethnic Minorities at Southwest University, China. Her specialty is comparative education with the focus on multiculturalism and educational development in ethnic minority areas of Southwest China. Her research interests concentrate on culture change and education development of the ethnic minority communities in China.

Lisong Jiang is Associate Professor of anthropology at Southwest University, China. His academic interests cover historiography, historical anthropology, and cultural anthropology. His recent research focuses on historical anthropology, especially on the studies of literary texts of southwest China in early Qing Dynasty.

His recent conference paper is "The Female Images of Ethnics in Early Qing China: A Text Analysis on Poems of the Savages (Mandong Zhuzhici)", presented at the Conference of CASCA/IUES2017, Ottawa. His publications include *An Introduction of Cultural Anthropology* (Southwest Normal University Press, 2008) and several chapters in different books. He has published several journal articles in *Guizhou Minzu Yanjiu, Guangxi Minzu Daxue Xuebao, Xinan Minzu Daxue Xuebao, Xinan Daxue Xuebao*, etc.

Shuai Li is Associate Professor, coordinator of the Chinese Program, and director of undergraduate studies in the Department of World Languages and Cultures at Georgia State University. His research interests include second language acquisition of Chinese, language testing and assessment, interlanguage pragmatics, and the international popularization of Chinese. His research articles have appeared in applied linguistics journals such as *Language Learning, The Modern Language Journal, Language Teaching, System, Language Testing in Asia, Pragmatics and Cognition, Chinese as A Second Language, Chinese as A Second Language Research*.

Xiaoting Li is Associate Professor of Chinese linguistics at the Department of East Asian Studies, University of Alberta. Her research focuses on language in Chinese interaction and multimodal interaction. Her publications include a book monograph *Multimodality, Interaction and Turn-taking in Mandarin Conversation* (John Benjamins, 2014), a co-edited book volume (with Tsuyoshi Ono) *Multimodality in Chinese Interaction* (De Gruyter Mouton, 2018), and articles in journals and book volumes including *Journal of Pragmatics, Language Sciences, Journal of Chinese Language and Discourse* etc. She serves on the Board of Directors of the Chinese Language Teaching and Research Association in Canada, and the editorial boards of the journal *Chinese Language and Discourse* and the book series *Chinese Discourse Studies* (World Publishing Cooperation).

Ni-Eng Lim is Assistant Professor of Chinese Linguistics in the School of Humanities at Nanyang Technological University (NTU), Singapore. He received his doctoral degree in Applied Linguistics from the University of California, Los Angeles, with specialties in interactional linguistics, conversation analysis, multimodal discourse analysis, and corpus linguistics. His research interests are interactional, socio-cultural and cognitive operations at work. His most recent project focuses on how doctor-patient interaction works in the multilingual and multicultural setting of Singapore hospitals. He has published articles in journals such as *Chinese Language & Discourse, Journal of Chinese Language Teachers Association, Language and Dialogue, Discourse Processes, Journal of Pragmatics* and *International Journal of Bilingualism*.

Jinzhan Lin is the Head of Research and Senior Research Associate at the Singapore Centre for Chinese Language Limited, a centre of Nanyang Technological University, Singapore. He received his Ph.D. degree from the National University of Singapore. Since 2013, he has led two government-funded research projects about corpora construction and application for Chinese Language teaching and assessment. His academic interests include corpus linguistics, lexical semantics, and Chinese Language teaching resources. In addition to the publications of various papers, he has compiled and co-compiled three frequency dictionaries of Singapore daily-used Chinese characters, words and sentence patterns, as well as a thesaurus of modern Chinese.

Jie Luo is a language instructor of Chinese at the School of Language, Linguistics, Literatures and Cultures, University of Calgary. She teaches beginning and advanced Chinese courses. She finished her Master's degree at the Department of East Asian Studies, University of Alberta, with specialization in Chinese linguistics. Her research focuses on Chinese multimodal interactions. Her MA thesis is entitled *The interactional functions of yinwei-clauses in Mandarin conversations.*

Liang Tao is an Associate Professor in the Department of Linguistics at Ohio University. She received her Ph.D. in Linguistics and postdoctoral training in cognitive psychology. Her research follows discourse-functional studies of language, including the usage based theory and exemplar representations on grammar and syntax in discourse, conversation analyses, language development, and psycholinguistics. She currently holds a diversified research agenda, including language variation and grammaticalization, repair in conversations by native and learners of Chinese language, Chinese language pedagogy, and psycholinguistics in reading development. She is a fellow in the Psychonomic Society.

Fei Ren is Assistant Professor and coordinator of the Chinese language program at Georgetown University. Her research interests focus on temporality and modality in Mandarin Chinese. Her current research includes the temporal reference of Chinese modal auxiliaries, the interface of Chinese aspect and modality, semantics of Chinese state complements, and their pedagogical applications.

Linda Tsung is an Associate Professor in the faculty of Arts and Social Sciences, University of Sydney. Her research interests are multilingualism, multilingual education, language policy and cultural identity in Australia and Greater China. She has published widely on these topics. In particular, in recent years, she has published two sole-authored books: *Language Power and Hierarchy: Multilingual Education in China* (Bloomsbury 2014) and *Minority languages, Education and Communities in China* (Palgrave Macmillan 2009). Both have received excellent reviews by scholars in Asia, the UK and the USA. Professor Tsung has also

co-edited books entitled *Contemporary Chinese Discourses and Social Practice in China* (Benjamins 2015) and *Teaching and Learning Chinese in Global Contexts* (Continuum 2010) and provided research leadership in the discipline of teaching Chinese as a second language, Chinese discourse studies and multilingualism globally.

Wei Wang is a Senior Lecturer in School of Languages and Cultures at the University of Sydney, Australia. His research interests include discourse studies, sociolinguistics, translation studies and language education. His recent research concentrates on sociolinguistics and (critical) discourse analysis, especially on interdisciplinary studies of contemporary Chinese discourse. His publications include *Media Representation of Migrant Workers in China: Identities and Stances* (2017), *Contemporary Chinese Discourse and Social Practice in China* (2015) and *Genre across Languages and Cultures* (2007). His journal articles appear in *Discourse Studies, Applied Linguistics Review, Journal of Multicultural Discourses, Chinese Language and Discourse, Australian Review of Applied Linguistics, Translation and Interpreting* and other international academic journals. He also published book chapters with Mouton, Bloomsbury, Routledge, Benjamins, the University of Michigan Press, and Wiley-Blackwell.

Yun Xiao is Professor of Chinese language and linguistics in the Department of Modern Languages at Bryant University. Her research interests are Chinese second language acquisition and pedagogy, syntax and discourse, heritage language learning and teacher training. Her recent publications include over 30 journal articles and book chapters, and three co-authored/edited research monographs, which are *Fostering Rooted World Citizenry: Studies in Chinese as a Heritage Language* (2008), *Teaching Chinese as a Foreign Language: Theories and Applications* (2008, 2011), and *Current Issues in Chinese Linguistics* (2011). She is also the primary author of the award-winning 4-volume *Tales and Traditions: Readings in Chinese Literature Series* (2008, 2010).

Jia Yang is an Assistant Professor in the department of Global Languages and Cultures at the University of Dayton. Her current research interests include corpus linguistics, Chinese language pedagogy, computer-assisted language learning, pragmatics, and cross-cultural communication. Her recent scholarly work has appeared in *Chinese as a Second Language, International Chinese Language Education, Journal of Technology and Chinese Language Teaching, Journal of the National Council of Less Commonly Taught Languages*, and *Language Learning and Technology*.

Lubei Zhang is currently a lecturer at Southwest Jiaotong University, China. She obtained her Ph.D. degree from Department of Chinese Studies, the University

of Sydney in 2015 and spent one year as a post-doctorial fellow at University of California, Los Angeles, in 2017. Dr Zhang's major research interests are in multilingual education and language policies for ethnic minorities in China. She has conducted fieldworks in minority areas in southwest China with Tibetan and Yi students. As a principal researcher, she led the Chinese national social science foundation project entitled "A comparative study of Semantic representation of the mental lexicon of Tibetan-Chinese bilingual and Tibetan-Chinese-English Trilingual based on SOA and ERP" 2016–2018. Her publications include *Contribution of Working Memory, Orthographic and Sentential Processing to Chinese Text Comprehension by Tibetan and Yi Students* (*The Reading Matrix* 2017) and *Changing attitude for bilingual education in Liangshan, China* (Benjamins 2015).

Chunsheng Zhao is a Research Associate at the Singapore Centre for Chinese Language Limited, a centre of Nanyang Technological University, Singapore. He obtained his Master of Arts in Applied Linguistics from Nanyang Technological University, Singapore. His academic interests include teaching Chinese language as a second language, corpus-based language teaching/learning and Chinese etymology. He is the Principal Investigator of a research project on students' Chinese learning motivation. He has co-compiled three frequency dictionaries of Singapore Chinese words and syntaxes and published several journal articles.

Lin Zhu is a Ph.D. student in the Department of East Asian Languages and Literatures at the University of Oregon. His research interests include Chinese syntax and semantics and second language aquisition in Chinese. He also works as a teaching assistant for Chinese language classes and a tutor and domain coach for the Chinese Flagship Program at University of Oregon.

Chinese discourse from diverse perspectives

An introduction

Yun Xiao and Linda Tsung

Bryant University / The University of Sydney

Since the 1980s, globalization processes have intensified in which technology advances, communication expands, people move, and languages flow. Globalization thus exerts an impact on all walks of life. China, in particular, as it has rapidly grown to becoming a major player in world affairs, has seen this effect across most aspects of its life. Chinese discourse has evolved rapidly through the impact of both external and internal factors associated with globalization. As Fairclough (1992) pointed out, discourse not only reflects social changes but also is integral and constitutive of social changes. The speedy economic, social and political changes taking place in China are bound up with changing Chinese discourse, reflecting new practices, identities, ideologies, values, and attitudes. Chinese discourse has not only been an important force for social changes in China, but itself is also in a constant state of changing (Tsung and Wang 2014).

In linguistics, increased efforts have been spent not only on the "innate" structural changes but also on language use in the social/cultural context. With the increasing global interest in China and Chinese language, a large amount of new data has become available. Authentic Chinese spoken or written texts are computerized and organized by researchers as corpora, making it possible to identify patterns of language use in subfields such as the lexicon, grammar, and discourse. This book features a discourse empirical orientation from diverse perspectives and various methodologies, in which large-scale databases (including self-created written and spoken corpora), narratives, interviews and surveys, among others, are employed and analyzed to gain a better understanding of new developments and changes in the Chinese language and discourse – both in China and in the global context.

This volume comprises a collection of 13 articles, which present updated and exemplary work in some of the key areas of studies in Chinese language and discourse, and touches upon a range of interlocking themes. The authors are internationally engaged scholars, many of whom are recognized leaders in their particular

areas in the study of Chinese language and discourse. They employ updated approaches and methodologies from a variety of fields, including applied linguistics, functional linguistics, and social linguistics, to describe the structure of Chinese language and discourse and to identify and examine critical issues, many focusing on globalization-induced language developments and changes. With an empirically-based discourse/socio-cultural approach, the collection of articles as a whole addresses a number of issues, ranging from linguistic features and formation of new words, usage-induced grammaticalization and language developments, and the interactional functions in everyday conversations. By their research orientations, the articles are arranged in roughly four groups.

The first three articles in this volume address changes in new word formations, new constructions in standard Chinese (Mandarin) discourses, and new variations in relation to the social context in which the Chinese language is used. In Chapter 2, Yun Xiao analyzes the linguistic features and word formation processes of neologisms in *The List of New Words in Media 2015*. The results show that word length has largely increased in contemporary Chinese language use, with both 3- and 4-morpheme words hovering around 40% of the total and 2-morpheme words under 17%. The major formation processes involved in the 4-morpheme expressions are blending, abbreviation, coinage, and numerical formulae. Chapter 3 presents a new construction in Beijing Mandarin Chinese that currently co-occurs with its original form in both spoken language and written texts. By illustrating such usage-induced language change, the study provides evidence showing how grammatical changes are fostered and conditioned through everyday communication. Shuai Li in Chapter 4 focuses on the contextual variations (i.e., power and imposition) in mitigation production of Chinese request-making. The results show that both power and imposition exert significant influence on the internal and external modifications and associate with their modifiers. In addition, the preferred sequential organization of external modifications differs according to context types.

The following two articles in Chapters 5 and 6 adopt the Conversation Analysis (CA) approach to discourse and examine the interactional functions or goals in everyday Mandarin Chinese conversations. By examining 11 hours of conversational data, Chapter 5 by Xiaoting Li focuses on the interactional functions of *yinwei*-clauses. The examination of the data shows that the *yinwei*-clause has a variety of functions, such as providing accounts for a speaker's prior action such as disagreement and assertion and serving as parentheticals to communicate background information. In Chapter 6, Ni-Eng Lim examines the frequently-used Mandarin Chinese meta-language expression "I-say-to-you." Drawing on data from the large-scale CallFriend Mandarin Corpus, the analysis reveals that a core function of "I-say-to-you" expressions is to preface upcoming "delicate" matters,

such as dispreferred next action, disagreement or disaffiliative turns, and other actions that may be resistance-implicative for the recipient.

Using diverse methods from various perspectives, Chapters 7–11 examine some of the outstanding Chinese linguistic issues such as near-synonyms, adjectival phrase (AP) state complement, constraints on collocational behaviors, kinship metaphors, and time expressions. Chapter 7 revisits the Chinese near-synonyms *jian* (建), *zao* (造), *gai* (蓋) 'to build' by testing three independent variables, such as word length of the object noun phrase, preverbal locative phrase, and building purpose. The analysis shows that word length and building purpose can account for the differences among these verbs. Chapter 8 examines how Chinese near-synonyms associate with typical collocates. The results show that the collocational behaviors of the selected synonyms are constrained by their own semantic, grammatical, prosodic, stylistic and pragmatic features. Chapter 9 investigates the semantic and pragmatic constraints on the generic/episodic interpretation of Chinese sentences containing a state complement (SC) realized by an adjectival phrase (AP). The analysis reveals that the generic interpretation of such sentences is a result of the interaction of the semantics of the verb or verb phrase before 得 *de* (V/VP-得), the AP complement after 得 *de*, and pragmatic knowledge. Chapter 9 investigates two metaphorically used kinship terms in Modern Chinese, 父 *fu* 'father' and 母 *mu* 'mother.' The findings suggest that figurative meanings expressed by *mu* (mother) are more conventionalized and have higher metaphoricity than those expressed by *fu*. From a Systemic Functional Linguistics perspective, Chapter 11 explores the common patterns, system networks, and realizations of time expressions in Mandarin Chinese, one of the fundamental concepts of human cognition and communication. The results show that time expressions can be classified as extent or location, realized as nominal groups, adverbs and pre-verbal phrases.

The final two articles extend a wide range of research agenda that includes a sociolinguistic analysis of the *Kam* ethnic identity and specialized corpora in Singapore, respectively. Chapter 12 explores the relationships between *Kam* people's sense of membership in their ethnic community and social practices that define the membership and identity construction. Chapter 13 presents two specialized corpora for Chinese language education in Singapore, which comprise the Singapore Daily Written Chinese Corpus and the Singapore Primary School Children Spoken Chinese Corpus. While the Written Corpus provides information on Chinese characters, vocabulary and sentence structures used in daily written materials, the Spoken Corpus provides guidelines for attainable spoken proficiency at different academic levels.

Collectively, these studies make a valuable contribution to research on Chinese language and discourse from diverse perspectives in today's globalized world. As such, the book will serve as a sound reference for Chinese researchers

and educators. It will attract the attention of individuals working in diverse fields: Chinese language and discourse, Chinese linguistics and language education, Chinese multiculturalism, and other areas. Those working outside the field of Chinese studies will also find the book of great value. From an educational perspective, the book will be a good reader for the upper-level undergraduate or post-graduate students who major in Chinese studies and Chinese applied linguistics.

References

Fairclough, Norman. 1992. Discourse and Text: Linguistic and Intertextual Analysis within Discourse Analysis. *Discourse and Society* 3 (2): 193–217.

Tsung, Linda and Wang, Wei. 2014. Analysing Chinese discourse in the new era. *Chinese Language and Discourse* 5:1 (2014), 1–6.

New words in contemporary Chinese language use

Linguistic features and formation processes

Yun Xiao

Bryant University

This study analyzes the linguistic features and word formation processes of the new words in *The List of New Words Used in Media 2015*. The results show that the average word length of the 446 new words used in media 2015 is 3.34, with both 3- and 4-morpheme words hovering around 40% of the total and 2-morpheme words under 17%. The majority of the 2-morpheme new words parallel the Chinese syntactic patterns, such as [modifier + modified], [subject + predicate], [verb + object], and [verb + verb]. The major processes involved in the 4-morpheme word formation are blending, abbreviation, coinage, and numerical formulae. In the blending and abbreviation processes, large chunks of information are clipped off to maintain the [2+2] 4-morpheme word length pattern. In addition, like many other newly-created usages, the case of 互联网+ shows that, in language change, new words can be created through grammaticalization and various types of derivational morphology, involving the creation of new affixes.

Keywords: lexicalization, morpheme, word length, word structure, formation processes, blending, abbreviation, coinage, numerical formulae, clipping

1. Introduction

This study aims to obtain a better understanding of the contemporary Chinese language use through the lens of newly-created words/expressions. Data show that there is a massive rise of new usages since China thrived in the decades-long economic reform and international integration (Yang 2012). *Language Situation in China*《中国语言生活状况报告》 reports that 5,733 new words/expressions were created in the past 10 years (2005 to 2015). Some of them come and go, but some stay. For instance, 顶 *ding* 'support' was ranked #1 in the 100 most frequently-used Internet words in 2005 but dropped to #1942 in the frequently-used

usages in 2015. Likewise, 汗或寒 *hanhuohan* 'fearfully respect' was ranked #13 in 2005 but disappeared in 2015. On the other hand, 被-XX *bei*-XX 'by-XX,' which was invented by the Internet users in 2008 with the first instance being 被自殺 *bei-zisha* 'to be suicided,' was initially viewed being syntactically ungrammatical and semantically absurd. But it was coined immediately and expanded immensely, with *bei* being elected as the "hot" word of 2009 (*Language Situation in China* 2009: 257). Despite the fact that about a decade has passed since its inception, 被 *bei* continues to be "hot." Data from *Language Situation in China* show that, in the three major media outlets (printed, TV, and Internet), *bei* was ranked #37 in 2009 (token = 1,278,666), #36 in 2011 (token = 1,361,079), #35 in 2014 (to-ken = 1,293,904), and #32 in 2015 (token = 1,284,553). Such linguistic phenom-ena demonstrate that some new usages can stand the test of time while some die away instantly.

Along with the rapidly increasing new usages, a large number of semi-pre-fixes/suffixes are invented (Jiang 2009; Zeng and Wang 2011), which generate many big word families such as 被-XX, 大-XX, XX-/X-族 through grammati-calization and reanalysis (Xiao 2015, 2016). Unlike the early new usages in the Chinese Internet language (CIL) which were created and spread by grassroots net citizens and officially branded as "ordinary, low-standard, and vulgar" (*Language Situation in China* 2012: 205), some of the current counterparts are invented and spread by elites or political leaders to reflect on the present social life, such as 互联网+ *hulianwang*+ 'Internet plus,' 四个全面 *sige quanmian* 'the Four-Pronged Comprehensive Strategy,' 红通 (红色通缉令) *hongtong* 'arrest warrant for the most wanted criminal(s),' 一带一路 *yidai yilu* 'the Belt and Road Initiative.'

However, there has been, so far, little research to look into the structure of the new usages, such as their length patterns, internal compositions, and formation processes. Drawing on theoretical perspectives in language change and lexical-ization (Packard 2000; Brinton and Traugott 2005; Bybee 2015; among others), this study takes on a synchronic analysis and intends to answer questions such as: What are the linguistic features of new words/expressions in contemporary Chinese language use? What linguistic processes are involved in their forma-tion? Data for analysis is drawn from *The List of New Words Used in the Media in 2015* 《2015年度媒体新词语表》, included in Volume B of *Language Situation in China 2016*.

2.   Language change and lexicalization: Processes and tendency

Languages change all the time and in all aspects, in which the most obvious to lan-guage users are changes in words (Bybee 2015: 1). During social communications,

new words are formed from the users' experience, and established words change their meanings and patterns to apply to new experiences or ideas. Once a change gains momentum, it is unlikely that it can be stopped (p. 11). Moreover, most languages acquire new words easily through processes such as borrowing, derivation, and compounding (Matthews 1991). Such processes by which new items that are considered "lexical" come into being are defined as "lexicalization" (Brinton and Traugott 2005: 32). According to Brinton and Traugott, lexicalization is the production of neologisms (new entries in the inventory) that encompasses a wide variety of linguistic processes, which include the ordinary word formation processes, the fusion processes that result in a decrease in compositionality, and those of separation that result in an increase in autonomy (2005: 32). Specifically, these linguistic processes include compounding, derivation, conversion, clipping and ellipsis, blending, back formation, initialism/acronym, loan translation, and coinage. They create, modify, combine, or separate existing units (p. 61), resulting in synchronic and diachronic language change. Out of them, compounding is the general process that involves the unifying of two or more autonomous words to form a third, e.g., *black + board > blackboard* (Brinton and Traugott 2005: 34), and derivation involves the addition of a derivational affix to form a new word, e.g., *unhappy* (p. 35). On the other hand, clipping and ellipsis abbreviate existing forms at different syntactic levels. While clipping abbreviates a word, e.g., microphone > mike, ellipsis shortens a phrase, e.g., public house > pub (p. 40). Moreover, through compounding and clipping, blending creates "portmanteau word" by fusing words into a single lexeme, e.g., (web)b + log > blog (p. 41).

According to Brinton and Traugott (2005: 46), the aforementioned processes create new lexemes, spread them, and finally add them into the inventory of a language with a tendency, diagrammed as:

nonce formation > institutionalization > lexicalization

In the continuum, "nonce formation" indicates "a new complex word coined on the spur of the moment" to serve an immediate communicative need of economizing, filling in a conceptual/lexical gap, or creating a stylistic effect. Once the nonce formation starts to be accepted by other speakers, institutionalization occurs; when the formation becomes a new word of the language, or a neologism, with a fixed meaning independent of context, lexicalization occurs.

Such a breakdown of the lexicalizing processes and tendency provide a discerning perspective and effective tools to analyze new words/usages of a language, with which I am turning to the next section.

3.   Studies on Chinese word: Structure, length, and constraints

The English "word" is translated as 词 *ci* in Chinese, but the concepts they encompass bear more differences than similarities. In English, "word" has two levels of presentation: lexeme in abstraction and word forms in concrete meanings. According to Brinton and Traugott (2005), a lexeme refers to a word considered as an abstraction such as RUN rather than as its various concrete word forms such as *run, runs, ran,* and *running* (p. 10). In the dictionary, the lexeme is listed as an independent and fixed lexical item, with its associated word forms being included in the same entry. However, the Chinese word 词 *ci*, as a borrowed modern concept from Western languages, has yet a clear and satisfactory definition in Chinese linguistics (Lü 2004: 40). By contrast, the Chinese character 字 *zi*, the monosyllabic graphic unit, has a long history of establishment. In both Classical and Modern Chinese, the overwhelming majority of characters represent single morphemes, and the great majority of morphemes coincide phonologically with a monosyllable (Norman 1988: 154). Thus, these three labels (i.e., morpheme, syllable, and character) are often interchangeably used, and so will they in this study.

In the prestigious modern Chinese dictionaries such as *New Chinese dictionary* 《新华词典》 (2001) and *Modern Chinese Dictionary* 《现代汉语词典》 (6th edition 2012), word *ci* is defined as "the smallest unit in the language that can be freely used." Nonetheless, not all the items listed in these dictionaries are "smallest units." Take 人 *ren* 'person' in *Modern Chinese Dictionary* (2012) for example. It leads a series of expressions which include 168 polysyllabic forms, each starting with *ren.* Out of them, 84 are disyllabic (e.g., 人民 *renmin* 'people'), 26 tri-syllabic (e.g., 人民币 *renminbi* 'RMB'), 49 with four syllables (e.g., 人民公社 *renmin gongshe* 'the people's commune'), 2 with five syllables (e.g., 人民检察院 *Renmin Jianchayuan* 'The People's Procuratorate'), 4 with six syllables (e.g., 人民代表大会 *renmin daibiao dahui* 'National People's Congress'), and 3 with eight syllables (e.g., 人同此心，心同此理 *Ren tong cixin, xin tong cili.* 'People feel and think alike on this matter.'). As shown in the above list, except for *Ren* which is a monosyllabic smallest unit, all the other items are strings of syllables or characters, ranging from 2 to many. In addition, disyllabic forms take 50% of the total in this particular list. Some of these expressions may be words, and some may not. For instance, 人民公社 *renmin gongshe* 'the people's commune' is not a word but a phrase, as claimed by *Lü* (2004: 40). This particular list is, however, not exceptional but illustrates the notion how "word" is defined in the Chinese language and how "words" are listed in the dictionaries of this kind. It is beyond the scope of this study to define the Chinese word, but it is pertinent to review previous studies in this area, which have more or less answered questions such as: What do Chinese words look like? What are their internal structures? What constrains their length

and pattern? What is the favored word length and pattern in the language? And what are the major processes in their formation?

So far, a good number of Chinese studies have claimed that the majority of Chinese words are disyllabic compounds. Norman (1988: 156) notes that the greatest number of words in the Chinese dictionary are compounds, which consist of at least two morphemes neither of which is an affix. Drawing data from a corpus of 100-million characters entitled *Core Wordlist of Contemporary Chinese for Information Processing*, Dong (2004) concurs that the majority of Chinese words are compounds and that compounding is the leading word formation process in Chinese, which typically combines independent words, although most of the attention has been directed to root combinations in traditional Chinese linguistics (p. 41). With a comprehensive counting and computation of the *Modern Chinese Dictionary* (3rd edition, 1996), Zhou (2014: 90) finds that, out of the 58,481 items in the inventory, 39,548 (67.63%) are disyllabic, 4,910 (8.40%) tri-syllabic, 4,798 (8.20%) with 4 syllables, 218 (0.37%) with 5 syllables, 105 (0.18%) with 6 syllables, and 108 (0.185%) with 7 to 12 syllables. The data lead Zhou to claim that 2-syllable compounds are the majority of Chinese words (p. 90).

The aforementioned claims are largely based on traditional Chinese dictionaries from an etymological perspective, but they are, somehow, supported by research findings from other perspectives. For instance, with a phonological approach Duanmu (2012) probes the Chinese word-length preferences by drawing on data from the 1.3 million-character Lancaster Corpus of Mandarin Chinese (LCMC, McEnery and Xiao 2004). He notes that in the section of "biographies and essays," there are 17,141 tokens and 4,451 types of polysyllabic expressions, ranging from 2 to 8 syllables. After eliminating the errors and those of 5 syllables and beyond, Duanmu finds that [1+1] 2-syllable compounds are the majority, which occupies 55.8% in the token counts of [N N] and 62.8% in the token counts of [V N]. He further proposes that there are four possible length patterns in the two-word expressions, such as 2+2, 2+1, 1+2, and 1+1 (p. 91), out of which 1+2 is disfavored in [N N] and 2+1 is disfavored in [V N]. Based on the data, Duanmu argues that a phonology-based analysis can better account for Chinese language facts and make the best prediction (p. 109), in that the combinations are more motivated by phonological constraints than by syntactic or semantic requirements. To some extent, Duanmu's findings support Feng's "prosodic syntax" notion (2013), which proposes that phonology prohibits certain word orders and length patterns in Chinese word formation. According to Feng, 2-syllable combinations are prosodic words with the smallest and standard metrical foot (2013: 100), while 3-syllable combinations constitute a superfoot, which are derivative and defective, especially with 2+1 [V N] and 1+2 [N N] (p. 101). On the other hand, [2+2] 4-syllable combinations are seamless in that they are composed of two standard metrical feet (p. 103).

However, agreement has not reached as to what constrains the Chinese internal word structure. Zhou (2014) contends that, in Chinese, polysyllabic words are subject to multi-level constraints, such as syntactic, semantic, phonetic, and stylistic (p. 156). He maintains that out of the 32,346 [root + root] combinations in Mandarin Chinese (p. 147), 96.57% of them follow the syntactic patterns, such as [modifier + modified] e.g., 人心 *ren xin* 'people's conscience', 高龄 *gaoling* 'old age'; [subject + predicate] e.g., 人为 *renwei* 'people behave', 气短 *qiduan* 'breath is short'; [verb + object] 赛车 *saiche* 'to race cars', 下海 *xiahai* 'to venture into doing business'; [verb + complement] 纳入 *naru* 'to bring in', 缩小 *suoxiao* 'to narrow down', etc. In addition, Zhou holds that semantic constraints override prosodic requirements in Chinese compounding. For instance, the following items are all in standard metrical foot but those in Group A are established words listed in the dictionary while those in Group B do not exist in the language due to their semantic violations, as illustrated in Table 1.

Table 1. Semantic constraints on compounding

Group A	Group B	Notes
烈日 *lieri* 'the glaring hot sun'	*烈星 *liexing* *'hot and glaring stars'	The sun can be 烈 (hot and glaring), but stars cannot.
落后 *luohou* 'lagged behind'	*落前 *luoqin* *'trailing in front'	落 means "trail behind," so it cannot go with 前 (in front).
喷饭 *penfan* '(laugh so hard as) to spew one's food (out of the mouth)'	*射饭 *shefan* *'(laugh so hard as) to shoot one's food (out of the mouth)'	喷 means "emitting under pressure," while 射 means "shooting by force."
奸徒 *jiantu* 'crafty villain(s)'	*忠徒 *zhongtu* *'loyal villain(s)'	徒 (villain) is derogatory, which can go with 奸 (crafty) but not 忠 (loyal).

Source: Zhou (2014: 150–152)

As discussed above, ample Chinese studies have been focused on compounds, their length preferences, and their motivating constraints. Nonetheless, Packard (2000) notes that, in addition to compounding, there are a number of other word formation devices in Chinese, such as abbreviation, borrowing, shifting, and numerical formulae (p. 268). According to Packard, abbreviation is the second most important device only after compounding. And compounding is a form of reanalysis in which a juxtaposed, novel concatenation of morphemes becomes lexicalized as a single word with no preexisting form, while an abbreviated form has a clear preexisting longer word or phrase that contains all the constituents of the abbreviation (2000: 268). Moreover, multi-syllabic expressions tend to be abbreviated to

disyllabic words, in which the first morpheme of each of the constituent words is selected unless a choice is overridden by semantic considerations (p. 271).

Table 2. Abbreviation samples

Abbreviations	Forms abbreviated
北图 *Beitu* 'North-chart Beijing Library'	北京图书馆 *Beijing Tushuguan* 'Beijing Library'
劳保 *laobao* 'labor insurance'	劳动保险 *laodong baoxian* 'labor riskprotecting insurance'
劳改 *laogai* 'reform through labor'	劳动改造 *laodong gaizao* 'labor change-create reform'
门责 *menze* 'doorstep hygiene responsibility system'	门前环境卫生责任 *menqian huanjing weisheng zeren* 'doorstep environment hygiene responsibility'
外劳 *wailao* 'foreign workers'	外来劳务人员 *wailai laowu renyuan* 'foreign workers'
公车 *gongche* 'public vehicle, bus'	公共汽车 *gonggong qiche* 'public-common vapor-vehicle'

Source: Packard (2000: 271–272)

As illustrated in Table 2, 劳动保险 *laodong baoxian* is abbreviated to 劳保 (劳动~~保险~~) *laobao* 'labor insurance', in which the first morpheme of each constituent word is chosen in the abbreviation process; however, in 公共汽车→ 公车 (公共汽车) *gongche* 'bus,' the first and the last morphemes are chosen due to the fact that 车 *che* 'vehicle', instead of 汽 'vapor', carries the core meaning of 公车 *gongche* 'bus'.

The aforementioned studies provide valuable evidence for the length pattern, internal structure, and linguistic constraints of Chinese words, with which the next section will look into the linguistic features and formation processes of the new words/expressions in contemporary Chinese language use by analyzing *The List of New Words Used in Media 2015* 《2015 年度媒体新词语表》.

4.   The study

The List of New Words Used in Media 2015 (the List thereafter) is included in *Language Situation in China 2016,* which is one of the annual reports in a series published by China's State Language Commission since 2005. Each of the annual reports in this series consists of two volumes: Volumes A and B. While Volume A presents reports and issues on contemporary Chinese language use, policy, and planning, Volume B consists of quantitative databases. The List under analysis

contains 469 new words/expressions that reflect on the "new phenomena, new concepts, new situations, and new changes in people's lives in 2015" (*Language Situation in China* 2016: 202). It is selected from the 1.7-billion-character National Language Resource Corpus, which collects texts used in media outlets that involve 15 mainstream newspapers, 15 major broadcasting stations (TV stations and radios), and two leading Internet websites' news publications (http://news.sina.com.cn/, http://news.qq.com/) (Hou et al. 2016). Collectively, they employed 575,424,830 words by token counts and 2,418,415 words by type counts from January 1st to December 31st, 2015. Each item in the List is provided with definition, examples, sources, and frequency ranking, in which the newly-promoted 互联网+ *hulianwang*+ 'Internet plus' is at the top of the List, with 58,341 token counts drawn from 24,583 different texts.

After eliminating the letter words (e.g., IP, SC, TBO, SPOCs), proper names (e.g., 安利 *Anli*, 玛丽苏 *Mary Su*, 杰克苏 *Jack Sue*), and combinations of characters + multi-letters e.g., IP 影视 *IP yingshi* 'movies and TV with intellectual properties') or characters + numerical digits (e.g., 中国制造 2015 *Zhongguo zhizao* 2015 'the strategic plan to make China become a strong manufacture nation issued on May 8th, 2015') from the List, 446 items remain for analysis. Included are transliterated words from borrowing (e.g., 饿怒 *enu* 'hungry', 轰趴馆 *hongpaguan* 'home party'), characters + symbol such as 互联网+ *huliangwang*+ 'Internet plus' and multi-word phrases such as 天津港爆炸事故 *Tianjingang baozha shigu* 'the Tianjin Harbor blasting accident,' 液态金属机器 *yetai jinshu jiqi* 'liquid-metal machinery', etc. In the following sections, the word length, word length preferences, word-internal structures, and linguistic processes involved will be explored.

4.1   Word length

Table 3 presents the length of the 446 items included in the List, out of which the majority are combinations of 3 or 4 morphemes, with 3.34 as the average. Specifically, 39.46% of them consist of 3 morphemes, 36.10% comprise 4 morphemes, and 16.82% contain 2 morphemes. Those with 1 morpheme and 5 above are in scarcity.

Table 3. Word length of media new usages in 2015

Word length	1	2	3	4	5	6	7	Total	Average length
Token counts	2	75	**176**	**161**	26	5	1	446	3.34
Percentage	0.45%	16.82%	**39.46%**	**36.10%**	5.83%	1.12%	0.22%	100%	

As shown in Table 3, the finding does not well agree with the long-held conviction that 2-morpheme words are the majority in Chinese (e.g., Norman 1988; Packard 2000; Dong 2004; Duanmu 2012; Zhou 2014). Instead, the present data contains a large number of 3- and 4-morpheme expressions, with 2-morpheme counterparts being the minority. Likewise, the present results do not well agree with those reported in *Language Situation in China* (2012) that claim, out of the new expressions created in 2009–2011, 3-morpheme words are the majority (51.01% in 2009, 53% in 2010, and 51.77% in 2011), while 4-morpheme counterparts are fewer (17.42%, 22.60%, and 21.08% respectively) and 2-morpheme ones are even fewer (18.69%, 17%, 15.51% respectively) (p. 246). In addition, compared with these 3-year data (2009–2011), 4-morpheme expressions have largely increased in the present List, while 3-morpheme counterparts notably decreased, with both hovering around 40%.

Likewise, such results do not well support previous findings from various perspectives. Cao (2010) proposes that 2-morpheme compounds are the leading pattern in standard Mandarin Chinese, while 3-morpheme words are colloquial and idiomatic. From a historical perspective, Zhou (2014: 258) maintains that the majority of Chinese words were monosyllabic in the ancient language, which evolved to combinations of 2- and 4-morpheme units in Old Chinese to represent the elitist upper culture (雅文化 *yawenhua*). On the other hand, combinations of 3 morphemes did not rise in large numbers until Tang Dynasty (i.e., middle ancient Chinese) when the vulgar common culture of towns (市井文化 *shijing wenhua*) started to thrive. In other words, while 2- and 4-morpheme words are formal and standard in the Chinese language, 3-morpheme counterparts have a shorter history and represent the informal and idiomatic style.

4.2   Word length patterns

Given the fact that most of the 2-morpheme words have a single 1+1 length pattern and those of 5 and beyond are actually multi-word phrases, this section is focused on the length pattern of 3- and 4-morpheme new words. The data show that there are three word length patterns in the 3-morpheme words and 4 patterns in the 4-morpheme words, as illustrated in Tables 4 and 5.

Table 4 shows that the 3-morpheme new words have three length patterns, which are 1+2, 2+1, 1+1+1, out of which the majority are 2+1 [N + suffix] (68.75%), followed by 1+2 [Prefix + N] (28.98%). In the 2+1 pattern, content words such as 感 *gan* 'sense', 群 *qun* 'group', 客 *ke* 'person', 度 *du* 'degree', 险 *xian* 'insurance', 帝 *di* '-emperor', 族 *zu* 'group', 党 *dang* 'party, group', and symbol +, take the position of the traditional Chinese suffix and function as such. Likewise, in the 1+2 pattern, content words such as 大 *da* 'big, major', 小 *xiao* 'small, short', 慢 *man* 'slow',

Table 4. 3-morpheme word length patterns

Length patterns	Frequency and percentage	Examples
2+1	121 (68.75%)	获得感 *huode-gan* 'sense of gaining' 创客群 *chuangke-jun* 'maker and inventor group' 挂科险 *guake-xian* 'penalty for student's failing an exam' 文化+ *wenhua* + 'culture plus'
1+2	51 (28.98%)	大基建 *da-jijian* 'major infrastructure projects' 慢就业 *man-jiuye* 'gap year after college' 硬医闹 *ying-yinao* 'to disturb the medical practices with violence' 微暴力 *wei-baoli* 'minor violence with disrespect and abuse'
1+1+1	4 (2.27%)	董监高 *dongjiangao* (**董**事, **监**事, **高**级管理人员) 'Board directors, supervisors, and advanced managers' 调转促 *diaozhuancu* (**调**结构, **转**方式, **促**升级) 'adjust the structure, transform the model, and upgrade the industry' 苏百万 *subaiwan* （**苏**宁, **百**度, **万**达）'Suning, Baidu, and Wanda corporations'
Total	**176 (100%)**	

软 *ruan* 'soft', 微 *wei* 'minor', are in the prefix positions. In such linguistic contexts, these content words function as pseudo suffixes or prefixes while maintaining their semantic meanings, which may become "real" affixes over time through grammaticalization and reanalysis (Xiao 2015). On the other hand, the pattern of 1+1+1 [N N N] or [V V V] is novel and scarce (2.27%), which are created through abbreviations. For instance, 董监高 *Dongjiangao* is abbreviated through word-end clippings from a longer preexisting phrase of 董(事) '*dongshi* board directors', 监

Table 5. 4-morpheme word length patterns

Length pattern	Frequency and percentage	Examples
2+2	156 (96.89%)	自贸片区 *zimao pianqu* 'free trade zone' 接力腐败 *jieli fubai* 'officials corrupt one after another in the same post'
3+1	2 (1.24%)	互联网+, *hulianwang*-+ 'Internet plus' 扶老人险 *hulaoren-xian* 'helping the fallen elderly' insurance
1+3	2 (1.24%)	微养老院 *wei-yanglaoyuan* 'Day care for elderly' 反水货客 *fan-shuihuoke* 'anti-smugglers'
1+1+1+1	1 (0.62%)	高精尖稀 *gaojingjianxi* 'advanced, accurate, cutting-edge, and rare'
Total	**161 (100%)**	

(事) *jianshi* supervisors', 高(级管理人员) *gaoji guanli renyuan* 'senior managers'. After the clippings, the remaining morphemes are fused to a single word that retains the meaning of the original phrase.

Table 5 shows that, out of the 161 4-morpheme words, those in the length pattern of 2+2 [N N] are the overwhelming majority (96.93%). The remainder are scarce but signal new developments in the contemporary Chinese language use, such as the affixed 扶老人-险 and character + symbol combination 互联网 + (which will be discussed in details below). In addition, 高精尖稀 *gaojingjianxi* 'advanced, accurate, cutting-edge, rare' is abbreviated through word-end clippings from a longer preexisting phrase 高(级) *gaoji* 'advanced', 精(密) *jingmi* 'sophisticated', 尖(端) *jianduan* 'cutting-edge', 稀(缺) *xique* 'rare'. After the clippings, the remaining morphemes are fused to a single word that gains a new meaning, that is, "the government-promoted strategic plan for cultivating talented people".

4.3   Word-internal structures

As noted in the previous studies, the majority of 2-morpheme Chinese words are compounds that typically combine two independent words (Norman 1988; Dong 2004), which are constrained at the syntactic level with patterns, such as [modifier + modified], [subject + predicate], [verb + object], [verb + complement] (Zhou 2014). Taking such analysis as the point of departure, this section concentrates on the internal structure of the [1+1] 2-morpheme words to examine how they parallel the Chinese syntactic patterns. Table 6 presents the 4 types of word-internal structures found in the seventy-five [1+1] 2-morpheme compounds in the List, which are [modifier + modified] (78.67% of the total), [verb + object] (17.33% of the total), and [verb + verb] (2.67% of the total), [subject + predicate] (1.33% of the total), with [modifier + modified] being the leading structure.

As shown in Table 6, out of the 59 [modifier + modified] structures, 30.67% are in the pattern [N + N], 22.67% in [Adj + N], 20.00% in [Adv + V], and 5.33% in [Num + N], with [N + N] being the leading pattern. Such data do not well agree with Zhou's findings (2014), in that [V + V] and [Num + V] are found in the present corpus but not in Zhou's count, and [verb + complement] is included in Zhou's count but not in the present corpus.

Table 6. Structures of 2-morpheme words

Structure types		Frequency and percentages	Examples
Modifier + Modified	N + N	23 (30.67%)	度秘 *dumi* 'Baidu-promoted robot secretary' 狮妈 *shima* 'Lion Mom'
	Adj. + N	17 (22.67%)	顶豪 *dinghao* 'the most luxurious mansion' 非标 *feibiao* 'atypical credit assets'
	Adv. + Verb	15 (20.00%)	互撕 *husi* 'to fight a war of words with each other' 爆买 *baomai* 'to engage in a shopping spree'
	Num. + N	4 (5.33%)	三互 *sanhu* 'mutual aids in information, supervision, and law enforcement' 两单 *liangdan* 'power and responsibility lists'
Verb + Object		13 (17.33%)	督企 *duqi* 'to supervise the enterprises' 锁盘 *suopan* 'to close the stock transaction'
Verb +Verb		2 (2.67%)	担当 *dandang* 'carry and take responsibility' 挂骂 *guama* 'hang a sign of curse'
Subject + Predicate		1 (1.33%)	众扶 *zhongfu* 'All support from the government, enterprises, and people'
Total		**75 (100%)**	

4.4   Word formation processes

The present data reveals several major linguistic processes in the Chinese new word creation, such as compounding, derivation, abbreviation, blending, coinage (coining or invention), and numerical formulae. According to Quirk et al. (1985), clipping and blending are highly productive ways in English word-formation, while Packard (2000) notes that, in Modern Chinese, abbreviation and coining are important ways of creating words. According to Packard, multi-syllabic expressions tend to be abbreviated to disyllabic words that further form 4-morpheme combinations (p. 80), while coining overwhelmingly yields disyllabic forms (p. 267). To illustrate the linguistic processes involved in the formation of new words/expressions, this section concentrates on the [2+2] 4-morpheme expressions included in the List, as shown in Table 7.

Table 7. Formation processes of [2+2] 4-morpheme compounds

Processes	Frequency and percentage	Examples
Blending	78 (50.00%)	I. Literally: 二孩效应 *erhai xiaoying* 'the second child effect' 网红经济 *wanghong jingji* 'online celebrities or influences make economic gains on fans' II. Figuratively: 蚊蝇腐败 *wenying fubai* 'petty officials embezzle money and engage in corrupt practices' 人口悬崖 *renkou xuanya* 'population crisis with declining birthrate and increasing lifespan'
Coinage for new products	34 (21.79%)	滴滴打车 *didi dache* 'app to call a cab' 微众银行 *weizhong yinhang* 'online corporate-owned banks'
Abbreviation	31 (19.87%)	巴黎恐袭 (2015年巴黎恐怖袭击) *bali kongxi* 'Paris terrorists attack' 自贸片区 (自由贸易试验区的分片区域) *zimao pianqu* 'Free Trade Zone'
Numerical formulae	13 (8.33%)	三网一化 *sanwang yihua* 'three coalitions and one industrialization' 四个全面 *sige quanmian* 'the Four-Pronged Comprehensive Strategy'
Total	**156 (100%)**	

Table 7 presents the four formation processes of the [2+2] 4-morpheme compounds in the List, out of which 50.00% are formed by blending to express new ideas or new changes literally or figuratively, 21.79% by coinage to name new products, 19.87% by abbreviations to describe new situations or new events, and 8.33% by numerical formulae to recapitulate a series of parallel actions or concepts with numbers. Among them, blending is the leading process. Such results show a similar pattern with those reported in Quirk et al. (1985) that blending is the highly productive way in English word-formation, but they do not well support Packard's observation (2000) that abbreviation and coining are the most important ways of creating Chinese words.

In addition, although both abbreviation and blending involve clipping and ellipsis, they differ in the availability of preexisting longer forms on which they operate. On the one hand, abbreviation has a clear preexisting longer word or phrase that contains all the constituents of the abbreviation (Packard 2000). For instance, 自由贸易试验区的分片区域 'Pilot Zones for Free Trade' is abbreviated to 自贸

片区(自由贸易试验区的分片区域) *zimao pianqu* 'Free Trade Zone', which involves middle clipping (由 *you* 'free', 易 *yi* 'trade'), end clipping (域 *yu* 'zone/district'), and middle ellipsis (试验区的分 *shiyanqu-de fen*). To satisfy the [2+2] prosodic requirement, some important information such as 试验区的分 'individual experimental or pilot zones' is clipped. Likewise, in 巴黎恐袭 （2015年巴黎恐怖袭击） 'Paris terrorists attack,' the essential information such as the year of 2015 is clipped to maintain the [2+2] 4-morpheme pattern.

On the other hand, blending creates new expressions with or without preexisting forms, e.g., 全面二孩政策下，因抚养压力对生育第二个孩子感到恐惧的现象→二孩效应, in which an initial clipping and a long middle ellipsis are applied, and 现象 *xianxiang* 'phenomena' is paraphrased/reworded as 效应 *xiaoying* 'effect'. Then the remaining key word 二孩 and the surrogate word 效应 form a [2+2] 4-morpheme compound: 二孩效应 *erhai xiaoying* 'the second child effect' through a fusion process. As in abbreviation, a large chunk of information such as 因抚养压力对生育第二个孩子感到恐惧的 'fear the pressure for raising a second child' is clipped off in the blending process to maintain a [2+2] 4-morpheme word length pattern. Such results show that, in Chinese word formation, requirements for prosodic and word length patterns are crucial and in general met at the expense of contents or semantic meanings. Moreover, blending is shown to be a flexible and economical device for the nonce formation, in that it can freely modify or rewrite a given statement with clipping, ellipsis, paraphrasing, and semantic reduction or extension. Given the swift social changes in contemporary China, blending becomes the most productive process that creates new forms with the greatest ease to express the many new situations and new ideas arising at various social levels, as illustrated in Table 7.

With similar effect, the Numerical Formulae compresses a row of parallel information to a [2+2] 4-morpheme word by leaving out large chunks of information. For instance, 三网一化[1] *sanwang yihua* 'three coalitions one -ization' is condensed from a long statement such as 高速铁路，高速公路和区域航空大网络以及基础设施工业化 *gaosu tietu, gaosu gonglu he quyu hangkong dawangluo yiji jichu sheshi gongyehua* 'high-speed rail, highway, regional airline alliance, and infrastructure industrialization'. This expression was originally invented to describe/summarize the Sino-African Agreement signed by China and African Alliance on January 27th, 2015, for the purpose of promoting Africa's transportation and infrastructure. In the formation process, a large amount of detailed information is dismissing, even though the new form can be misleading to audience who are out of the context or lack the background knowledge. Specifically, the compressed form employs 网 *wang* to represent the three major branches of transportation

---

1. http://gb.cri.cn/42071/2015/01/28/6891s4856170.htm

(rail, highway, and airlines) and 化 *hua* to indicate industrialization infrastructure, which makes many shades of semantic extension. For 网 *wang* indicates network or Internet in Chinese but does not exactly mean "rail or highway," and 化 *hua* '-ization' is a suffix that can change a noun or an adjective to a verb but not a content noun denoting "industrialization."

In contrary to the belief that coinage is a rare or exceptional process that creates new inventions (Brinton and Traugott 2005: 42), it is found to be largely employed in the List to make names for the many new products brought about by China's new technology, new software, and new internet services in recent years. For instance, out of the 34 coined new names listed in Table 7, twelve are related to the newly-created cab-calling apps, such as 滴滴打车 *didi dache* 'didi calling a cab', 滴滴快的 *didi kuaidi*, 滴滴专车 *didi zhuanche*, 滴滴快车 *didi kuaiche*, 滴滴代驾 *didi daijia*, 滴滴出行 *didi chuxing*, in which a 2-morpheme onomatopoeic word 滴滴 *didi* (the sound of phone ringing) is combined with another 2-morpheme word of vehicles or transportation device to form a 4-morpheme compound that well meets the phonological requirement of [2+2] standard metrical feet.

Song (2015: 6) notes that, structurally, there are two types of [2+2] 4-morpheme expressions in Chinese: one is 四字语 *size yu* 'four-character phrases', and the other is 四字格 *sizige chengyu* 'four-character idioms'. 四字语 are phrases that are open to structural modifications, e.g., 经济复苏 *jingji fusu* 'economy recovery' → 经济的复苏 *jingji-de fusu* 'recovery of economy'; 现场解决 *xianchang jiejue* 'on-the-spot solution' → 在现场解决 *zai xianchang jiejue* 'to solve on the spot' (p. 11). On the other hand, 四字格 are fixed idioms in which any structural expansion would cause ungrammaticality (p. 6), e.g., 一衣带水 'close neighbors separated by only a strip of water' vs. * 一衣带的水 ；大呼小叫 'shout and wrangle' vs. *大呼和小叫. By this account, the new 4-morpheme expressions in the List are mostly phrases 四字语 in that structural expansion or change does not cause ungrammaticality, e.g., 二孩效应 'the second child effect' → 二孩的效应 'the effect of the second child'; 三网一化 'three coalitions one – ization' → 三网和一化 'three coalitions and one – ization'; 滴滴打车 '*didi* calling a cab' → 用滴滴打车 'calling a cab with *didi*'.

4.5 The case of 互联网+ 'Internet-plus'

In addition to the linguistic processes discussed above, the List also embraces processes that form unconventional combinations to express new events and new actions, such as "characters + symbol", e.g., 互联网+ *hulianwang*+ 'Internet plus,' 文化+ *wenhua*+ 'culture plus.' Unconventional as it is, 互联网+ 'Internet plus' rises up to the top in the List by token count and has been spreading and coined across

China and abroad. For instance, on October 8th, 2016, University of Leeds, UK, issued a call-for-paper for an international Chinese pedagogy conference, with the title being "2017 互联网+汉语国际教育专题研讨会 '2017 Internet+ Forum on Chinese International Education'." Likewise, a coined form such as "直播+"靠谱吗? *zhibo+ kaopu-ma* 'Is Live TV news+ reliable?' is used as a news headline in *People's Daily* (overseas edition, January 2nd, 2017).

Given such high popularity, 互联网+ is a case worth examining for its sudden rise and rapid expansion and coinage. Unlike the early usages in the Chinese Internet language (CIL) which were created and spread by grassroots net citizens, 互联网+ was invented by elites and subsequently promoted by political leaders. It was first coined by technology leaders in 2012 for the concept of "enterprise + Internet", which meant that Internet technology was an important element in the corporate economy (Ouyang 2015). It did not catch much attention until 2015 when the Chinese Premier *Li Keqiang* adopted the concept and employed it in the government strategic plan, named as 互联网+行动 *hulianwang+xingdong* 'Internet + action', in which the symbol '+' no longer means 'to add' but 'to lead and integrate', and 'action' stands for all walks of life such as agriculture (互联网+农业 *hulianwang+nongye*), manufacture (互联网+制造 *hulianwang+ zhizao*), transportation (互联网+交通 *hulianwang+jiaotong*), irrigation (互联网+水利 *hulianwang+shuili*), finance (互联网+金融 *hulianwang+jinrong*), energy (互联网+能源 *hulianwang+nengyuan*), services (互联网+服务 *hulianwang+fuwu*), etc. (Ouyang 2015: 36). Like some of the 'hot' CIL usages, 互联网+XX is swiftly spreading in China, taking a direction however opposite to the CIL. Its path can be diagrammed as: up from the technology elite circle, further up to the government platform, and down to the Internet and ordinary people. Data from CNKI (中国知网) shows that there were only 575 cases of 互联网+ in 2012, which grew up to 7.5 million in 2015 (*Language Situation in China* 2016: 231). A quick search in Google pops up 538 million instances in 0.45 seconds (retrieved on December 2nd, 2016).

Syntactically, 互联网 originally served as the initial head of a copulative compound, and the symbol "+" linked the subsequent head (e.g., 互联网+XX). With its widespread use and increasing recognition as 互联网+时代 *hulianwang+ shidai* 'Internet+ era', the subsequent head is dropped off, while 互联网+ stands alone to mean "internet as the core/engine that integrates and leads all the industries/businesses" (Sun 2016: 1, Shi 2016: 161). The lone 互联网+ is further analogized and generates a new pattern of "XX+" in CIL and beyond, such as 文化+ *wenhua+*, 'cultural plus,' 直播+ *zhibo+*, 'live TV news,' 太平洋+ *taipingyang+* 'Pacific plus,' Alibaba Buy+,[2] etc. The XX+ continues to be coined and extends to X+X in

---

2. http://technology.inquirer.net/56131/alibaba-launches-full-vr-shopping-experience-buy

recent days, such as 16+1 合作 *hezuo* '16+1 Cooperation' (*Li Keqiang* November 6, 2016),[3] 1+6 圆桌对话会 *yuanzuo duihuahui* '1+6 roundtable forum' (*Li Keqiang* July 22, 2016)[4] by the government leaders, in which 1 *yi* 'one' stands for China as the leader and the parallel figure as the number of participating foreign countries. This new pattern of X (num.) + X (num.) has been spreading to indicate two composing or opposing components of an entity, such as 粤港澳大湾区 9+2[5] *YueGangAu Dawanqu 9+2* '9+2 in the Guangdong-Hong Kong-Macau Big Bay Area', G20 在巴黎协定上分裂成 19+1[6] *G20 zai Balixiedingshang fenliecheng 19+1* 'The G-20 summit 2017 split into 19+1 on the Paris climate agreement'. In the former, "9+2" indicates the 9 cities and 2 regions that shape the Big Bay Area currently under a new initiative of assembly and development by the Chinese government, whereas in the latter "19+1" means that "at the G-20 summit, 19 member countries were against the U.S. on the climate agreement".[7]

Such language use arguably exhibits a linguistic ripple effect, which starts from the initial state 互联网+XX and spreads outwards incrementally in the forms of 互联网+, XX+, X+X, promoted by and spread at both the official and grassroots levels.

5. Discussion and conclusions

By analyzing the 2015 media new words for their linguistic features in word length, word length patterns, word-internal structures, and their formation processes, including the special case of 互联网+, the present study makes a number of meaningful findings. First, the results show that the average word length of the 446 words under analysis is 3.34, with both 3-morpheme and 4-morpheme words hovering around 40% of the total and 2-morpheme words under 17%. The finding does not support the long-held conviction that 2-morpheme words are the majority in Chinese (e.g., Norman 1988; Packard 2000; Dong 2004; Duanmu 2012; Zhou 2014). Secondly, the majority of the 2-morpheme new words parallel the Chinese syntactic patterns, such as [modifier + modified], [subject + predicate], [verb +

---

3. http://www.chinanews.com/gn/2016/11-06/8054452.shtml

4. http://www.fmprc.gov.cn/web/zyxw/t1383713.shtml

5. http://news.ifeng.com/a/20170413/50928202_0.shtml

6. http://news.sina.com.cn/w/hq/2017-07-09/doc-ifyhwefp0288635.shtml

7. https://www.usatoday.com/story/news/politics/2017/07/08/paris-climate-agreement-g-20-summit-its-19-against-trump-u-s/461795001/

object], and [verb + verb], which indicates that Chinese word formation is largely constrained at the syntactic level, as claimed by Zhou (2014). Thirdly, the remarkable increase of 3-morpheme words signifies that the contemporary Chinese new words are by and large informal and idiomatic. The data further show that over 97% of the 3-morpheme words are in 1+2 or 2+1 word length pattern, which indicates that they are formed through derivation with suffixes or prefixes. Such results confirm the report of *Language Situation in China* (2011–2012) and findings of previous studies that a large number of semi-prefixes and semi-suffixes are rising in the contemporary Chinese language use (Jiang 2009; Zeng and Wang 2011; Xiao 2015, 2016; among others). They further give evidence to support Bybee's observation that all languages have means to create new words from existing resources, one of which is various types of derivational morphology (2015).

Fourthly, the large amount of 4-morpheme words found in the List upsets the belief that 4-morpheme words are minority in the Chinese inventory. In contrary, the data show that over 36% of the new words are 4-morpheme forms, and out of them over 97% are in 2+2 length pattern, which, according to Feng (2013), are favored forms; namely, they well meet the Chinese phonological requirement. In addition, the major processes involved in the 4-morpheme word formation are blending, coinage, abbreviation, and numerical formulae, and blending is the most productive one, probably due to its ability to employ multiple formation devices with or without preexisting longer forms. It is assumed that such high flexibility makes it economical and possible to express the new ideas, new situations, and new changes brought about by the rapidly-changing contemporary China. Moreover, in the blending and abbreviation processes, large chunks of information are often clipped off to maintain the [2+2] 4-morpheme word length pattern, which suggests that, in Chinese word formation, requirements for prosodic and word length patterns are essential and often met at the expense of contents or semantic meanings. Furthermore, the new 4-morpheme expressions in the List are mostly phrases, which are open to structural expansion and syntactic modification. By Bybee's hypothesis on language change momentum (2015), such phrases may be solidified and become fixed 4-character idioms in the long run as they continue to be employed by language users.

Fifthly, the case of 互联网+ reveals a linguistic ripple effect in the lexicalization process, which starts from the initial state 互联网+XX and spreads outwards incrementally with more new expressions, unconventional and abnormal. The case shows that, to reflect on China's new changes, unconventional novel forms can take a top-down direction in the course of creation and expansion, i.e., from elites or government officials to the mainstream society. These "top" parties may be more powerful in spreading and retaining such forms than grassroots netizens whose CIL usages are officially branded as "low-standard, vulgar" and subject to

censorship at various levels. Linguistically, the likely reason that 互联网＋ is adopted by Chinese speakers and generates large word families is that it fits the Chinese [3+1(suffix)] word pattern after the verb + 'to add' in 互联网+XX is grammatized to become a quasi suffix in 互联网+. Like 被-XX which also starts with an unconventional novel but generates large word families with a [1(prefix)+2] word pattern, 互联网＋ involves grammaticalization and reanalysis that create a new affix.

Finally, by Brinton and Traugott's lexicalization scheme (2005), these new words are "*nonce formation*", which are in the stage of "*institutionalization*" (i.e., creating and spreading) at the present time. Can they reach the stage of "*lexicalization*" by entering the inventory of the Chinese language? With the evidence furnished in the study, the answer is positive in that the majority of these new forms conform the Chinese morphological, syntactic, prosodic, and semantic requirements, which well condition them to survive, spread, and thrive. Take 被-XX (*bei*-XX) as an example. It was invented by the Internet users in 2008 through coinage 被自殺 *bei-zisha* 'to be suicided', which was initially viewed as an unconventional and absurd nonce form. However, the usage is now adopted by the prestigious *Modern Chinese Dictionary* (2012: 58) to become part of the inventory. To the contrary, its peer usage 汗或寒 *hanhuohan*, 'fearfully respect' was ranked #13 in 2005 but died out by 2015. There may be many reasons for its demise, but the obvious one is that this novel form does not meet any of the Chinese requirements at the morphological, semantic, syntactic, prosodic, or semantic level.

The evidence and findings furnished in the study arguably suggest that, out of the many newly-created forms in contemporary Chinese language use, those which have favored patterns in word length and meet the various Chinese linguistic requirements will be accepted by the users and likely become neologisms of the language inventory in the long run, while those which do not may become obsolete or die out sooner or later.

Acknowledgements

References

Brinton, Laurel J. and Elizabeth Closs Traugott. 2005. *Lexicalization and Language Change*. Cambridge University Press.

Bybee, Joan. 2015. *Language Change*. Cambridge University Press, UK.

Cao, Wei曹炜. 2010. 《现代汉语词汇研究》(Modern Chinese etymology). Guangzhou, China: 暨南大学出版社 (Jinan University Press).

Chinese Department of Education Language and Script Resource Management Bureau 教育部语言文字信息管理司. 《中国语言生活状况报告》 (*Language Situation in China*), 2005–2016. 商务印书馆 Beijing, China: The Commercial Press.

Duanmu, San. 2012. "Word-length preferences in Chinese: a corpus study." *East Asian Linguistics* 21(1):89–114.

Dong, Xiufang董秀芳. 2004. 《汉语的词库与词法》 (Chinese lexicon and morphology). 北京大学出版社 (Beijing, China: Beijing University Press).

Feng, Shengli冯胜利. 2013. 《汉语韵律句法学》 (*Chinese prosodic grammar*). 商务印书馆 (Beijing, China: The Commercial Press).

Hou, Min侯敏, Zou Yu邹煜, Teng Yonglin滕永林, Cheng Nanchang程南昌. 2016. 《2015年度媒体新词语表说明》 [Description of the list of 2015 media new words]. In 《中国语言生活状况报告》 (Language Situation in China: 2016), Volume B. 商务印书馆 (Beijing, China: The Commercial Press).

Jiang, Zongxu蒋宗许. 2009. 《汉语词缀研究》 (Study on Chinese affixes). 成都：四川出版社集团巴蜀书社 Chengdu, China: Sichuan Publishing Corp.

Lü, Shuxiang吕淑湘. 2004. 《吕叔湘文集》第五卷 (Lü, Shuxiang collected works), Volume 5. 商务印书馆 (Beijing, China: The Commercial Press).

Matthews, Peter H. 1991. *Morphology*. Second edition. Cambridge University Press.

McEnery, Tony and Richard Xiao. 2004. "The Lancaster Corpus of Mandarin Chinese: A corpus for monolingual and contrastive language study." In *Proceedings of the fourth international conference on language resources and evaluation (LREC) 2004*, Ed. M. T. Lino, M. F. Xavier, F. Ferreire, R. Costa, and R. Silva. 1175–1178. Lisbon, May 24–30, 2004.

Norman, Jerry. 1988. *Chinese*. Cambridge University Press.

Packard, Jerome L. 2000. *The Morphology of Chinese: A Linguistic and Cognitive Approach*. Cambridge University Press.

Ouyang, Rihui欧阳日辉. 2015. "从"+互联网"到"互联网+"：技术革命如何孕育新型经济社会形态 [Changing from "Plus Internet" to "Internet +": How technological revolution brings about new economic and social patterns]. "《学术前沿》 (Frontiers) 2(1): 23–38.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. Longman.

Shi, Xing施行. 2016. "互联网+之浅见 [Remarks on Internet+]." 《科技传播》(Science and Technology Communications) 1(1): 161–162.

Song, Saiping宋赛萍. 2015. 《汉语的四字格》 On Chinese four-character idioms. 北京：北京语言大学出版社 Beijing: Beijing Language and Culture University Press.

Sun, Bo-lin孙柏林. 2016. "试评互联网+装备制造业 [On the "Internet + equipment manufacturing industry]." 《自动化技术与应用》 (Technicques of Automation & Application) 35(1): 1–7.

Xiao, Yun. 2015. "The rise of Chinese Internet language: Product and agent of globalization from below." *Global Chinese* 1(1): 111–137.

Xiao, Yun. 2016. "The Rise of Catch-Word "被 *bei*-XX:" Grammaticalization and re-analysis of *bei*." *Chinese as a Second Language* 51(1): 79–97.

Xiaoping Yang杨小平. 2012. 《当代汉语新词新语研究》 New words and expressions in contemporary Chinese. 北京: 中国社会科学出版社 Beijing: The Academy of Chinese Social Sciences Press.

Zeng, Liying and Wang Pingli. 2011. "Use a lexical database for identifying quasi-affixes in contemporary Chinese." *International Journal of Knowledge and Language Processing* 2(2):53–61.

Zhou, Jian周荐. 2014. 《汉语词汇结构论》 (On Chinese lexical structures). 人民教育出版社 (Beijing, China: People's Education Press).

现代汉语词典 (Modern Chinese Dictionary) 第6 版 (The 6th edition). 2012. 中国社会科学院语言研究所词典编辑室编 (Linguistics Research Institute of Chinese Academy of Social Sciences). 商务印书馆 (Beijing, China: The Commercial Press).

新华词典 (New Chinese dictionary), 修订版 (Revised edition). 2001. 商务印书馆 (Beijing, China: The Commercial Press): 商务印书馆辞书研究中心 (Dictionary research center of the Commercial Press).

CHAPTER 3

Usage based language change and exemplar representations in Beijing Mandarin Chinese

Liang Tao
Ohio University

This study offers support to usage-based studies to promote the importance
of everyday language use in language development and grammaticalization.
Specifically, the study presents a new construction in Beijing Mandarin Chinese
that currently co-occurs with its original form, both in spoken language and
written texts. The change is another instance of phono-syntactic conspiracy (Tao
2002, 2006, 2009). It starts from phonological reduction and ends in a syntactic
change of a highly frequently used rhetorical question 不是…吗: 'Isn't it the case
that….' However, the process differs from previous findings (Bybee 2010) in that
the grammaticalization process involves usage frequency as well as cognitive,
cultural and social factors. The findings further support the view that language
and grammar are fostered and conditioned through human communication.

Keywords: grammaticalization, frequency, chunking, phonological fusion,
Mandarin tones, phono-syntactic conspiracy, discourse-functional linguistics,
bu35shi…ma (不是…吗)?

1.   Introduction

Language is a complex adaptive system (Bybee 2010: 2) with variations and chang-
es at any given time. This study offers support to the proposal of usage-based
theory and exemplar representations (Bybee 2003, 2013; Pierrehumbert 2001) to
promote the importance of usage frequency (Bybee and Hopper 2001; Zipf 1968)
in language development and grammaticalization. The study also proposes ad-
ditional social factors that may have assisted the spreading of language change.

Specifically, the study discusses the impact of Beijing Mandarin (and possibly
northern Mandarin in general) to the grammaticalization processes in Mandarin
Chinese. The processes involve two steps, (a) phonological fusion caused by fre-
quent usage in conversational speeches, which eventually lead to phonological

simplification and language change; and (b) the status and impact of Beijing Mandarin to Mandarin Chinese, which is reflected in media (TV shows, etc.) and in written Chinese texts. The former is reflected in spontaneous conversations without speakers' conscious choice; whereas the latter could have resulted from conscious selection of a new pattern to reflect casual spoken Chinese. Both show the impact of everyday communication on language and grammar. Detailed discussions are presented in Section 4.

The basic premise of the usage-based theory is that experience with language creates and impacts the cognitive representations of language (Langacker 1987, 1999; Barlow and Kemmer 2000). Every token of a construction type impacts the memory as exemplar representations of the type (Bybee 2006, 2013; Pierrehumbert 2001). The cognitive representations of grammar consist of exemplar representations of token variations. Therefore, "grammar is the cognitive organization of one's experience with language" (Bybee 2006: 711; Thompson and Hopper 2001).

Grammaticalization concerns the creation of a new grammatical morpheme or a new construction, often out of an existing construction (Hopper 1998; Bybee 2003, 2006, 2010, 2013). The processes often start from language variations. When a variation moves away from its original environment, such as a grammatical construction A, to a new construction B, B is considered grammaticalized (Bybee 2000, 2003, 2006, 2010, 2013; Bybee and Scheibman 1999; Heine 2002; Hopper 1994, 1999). Grammaticalization often happens to an individual lexical item, yet a multitude of changes of the item in different constructions may result in some general changes to a language. For instance, 'I'm going to…' has been used highly frequently in spoken American English, and now 'going to' has become a single lexical item 'gonna' (Bybee 2006: 719), although in dictionaries it is considered a 'contraction' to 'going to' (Cambridge Dictionary online). This practice is an instance of domain-general processes of part of human cognition (Bybee 2010).

Many studies on grammaticalization have offered explanations of how a construction A developed into B historically, but the intermediate stages have been mostly buried in history (Heine 2002). Current discourse-functional studies offer synchronic documentations of the intermediate stages of language development (Hopper and Traugott 2003). For instance, the most widely used Mandarin Chinese classifier 个 *ge51* is now paired with a large variety of referents, such as humans and many objects. The original core meaning of 个 *ge51* denotes a single piece of bamboo, arrow or thing (Wieger 1965: 611; Lindqvist 1991: 231; Wang 1958: 238; also see discussions in Erbaugh 1986). This lexical item has developed into the function of a general classifier in Mandarin Chinese referring mostly to humans (Lü 1990). It has also extended its function to pair with nominal referents that have their own default classifiers (Erbaugh 1986; Lü 1990; Tao 2002, 2005, 2006). The above information has been inferred from written historical documents, but

no specific details as to how *ge51* expanded its realm into human beings could be pointed out. As a matter of fact, historical linguistics often defines language development based on written records that display usage changes over time (such as when such usages first appeared), but not how exactly such changes occurred. Current discourse-functional studies may offer instances of language change as it is being developed (e.g. Bybee 2006, 2010, 2013; Su 2016). Su (2016) illustrates the beginning and spreading of an emergent construction synchronically in individual instances of oral communication. Specifically, the study discusses how the *X type thinker* construction in English discourse was invented by an influential figure – Bill Gates, which was documented with modern technology (recorded DVD of an interview). Through widely circulated videos online, the new construction was repeated/reused gradually by other speakers, including political leaders (US president George W. Bush). The present study offers another instance of grammaticalization but the processes completely differ from Su's (2016) observation.

"Both grammatical meaning and grammatical form come into being through repeated instances of language use" (Bybee 2013: 712). Frequent usage of certain constructions in conversations may lead to variations, resulting from phonological fusion, eventually leading to morphological and grammatical changes (Bybee 2006, 2010, 2013; Hopper and Traugott 2003; Heine 1993, 2002; Tao 2002, 2005, 2006, 2009). The changes impact speakers' cognitive perception and processes of the new constructions (Bybee 2006; Tao 2009). Cognitive representations are built up as language users encode utterances and categorize them on the basis of phonetic form, meaning, and context. As incoming utterances are sorted and matched by similarity to existing representations, units such as syllable, word, and construction emerge.

Variations are often based on usage frequency. For instance, in American English, schwa reduction and disyllabification occur before sonorants /n/ and /r/, based on usage frequency (Pierrehumbert 2001). In high frequency words, disyllabification happens so the internal syllable of words such as 'ev<u>e</u>ry' is produced with the omission of the schwa. In rare words, the schwa is fully pronounced with its following /r/, such as in 'artill<u>ery</u>'. All of the variations form exemplar representations of the schwa-related syllable that are stored in the memory. These variations are potential candidates for language change. Furthermore, when two elements are used together frequently, phonological reduction may occur and two words may be fused into one morpheme, e.g., wanna (want to) (Bybee, Pagliuca, and Perkins 1991; Bybee, Perkins, and Pagliuca 1994; Bybee and Scheibman 1999).

Variations develop in a speech community from everyday conversations. Specific phonological characteristics in a speech community is built through interactional conversations, partially due to a pervasive practice of repetition. Studies find that there is a very high rate of repetition and reuse of expressions between

speakers in interactional conversations for the purpose of co-operation and stance building (Bybee 2006; Goodwin 2013; Goodwin and Goodwin 2004; Su 2016; Tannen 1987; Thompson and Hopper 2001). Repetition of commonly used expressions lead to high usage frequency (Bybee 2006, 2010; Bybee and Hopper 2001; Tao 2015). One of the consequences of high usage frequency is automatic imitation during word production (Fowler 1986). Experimental studies (Goldinger 1998, 2013) suggest that listeners are able to track speakers' articulatory gestures and "rapidly approximate those gestures while responding" (Goldinger 2013: 2). There are a myriad factors behind this behavior, and usage frequency plays a major role in this practice. This present study proposes that the combination of usage frequency and automatic word imitation could be one of the factors that help build phonological characteristics of a speech community, including phonological variations.

Presented in this study is an instance of phonological variation that is developing into a new grammatical construction in Mandarin Chinese, starting most likely from spoken Beijing Mandarin (hereafter, Beijing Mandarin) and moving to written Chinese, resulting in a new construction that may have modified the function of a lexical item, the negative particle 不 *bu51* in the construction 不是…吗 *bu35shi51…ma* 'Negative-be…Q: isn't it the case that…'?

It has been well documented that frequent usage may lead to sound erosions (e.g., Bybee 2003, 2013; Heine 2002; Heine and Kuteva 2002; Hopper and Traugott 2003; Tao 2002, 2005, 2006). Lass proposed (1984: 176) a cross-linguistic pattern of consonant changes: "stop > fricative > approximant > zero". The present study shows exactly how this developmental pattern happens in spoken Beijing Mandarin Chinese, and how it may have been adopted into standard Mandarin Chinese as a form of spoken language. This study proposes that the essence of this phonological development is everyday language use. Previous studies suggest that sound erosion often occurs when a verb loses its full verbal status while on its way to become a grammatical particle (Heine 1993: 56). In the present study sound erosion occurs for a different reason, usage frequency and phonological or prosodic units in everyday conversation.

The specific synchronic developmental pattern this study proposes concerns /sh/, an alveolar fricative in the syllable *shi51* 'be', whose changes follow the pattern: Fricative > approximant > syllabic > zero. Specifically, the variation occurs in a highly frequently used two-word unit 不是 *bu35shi51* 'Negative-be', and its variations have impacted the rhetorical construction 不是…吗 *bu35shi51…ma* 'negative-be…Q'. The phonological variation occurs on the verb 是 *shi51*, which, in conversational speeches is often uttered unstressed, losing its tone to become a syllabic rhotic /r/ with the vowel omitted. The rhotic /r/ may be dropped, resulting in a new construction 不…吗 *bu35…ma*. This new construction may lead to

ambiguous interpretations that have to be differentiated through context. A detailed explanation of this change is presented in the following section.

What is new in this report is that before being eliminated, the approximant (Ladefoged 1964) first becomes syllabic, standing alone for the verb 是 *shi51* 'be'. Therefore, with the deletion of the approximant, the verb is eliminated, which, when happening in a syntactic construction as mentioned above, leads to the development of a new construction. Furthermore, the current processes to grammaticalization may have been expedited by additional cognitive and social factors,[1] including cognitive processes of a reduced syllable and the influential status of Beijing Mandarin as the pronunciation foundation for 普通话 *Putonghua* 'common language', the official language in China, transmitted through the wide accessibility of media due to rapid development of technology.

The findings of this study may enhance our understanding and appreciation of the impact of everyday conversations on language changes. The findings also demonstrate how language development and grammaticalization actually take place synchronically through oral communication, with direct evidence of the connections from construction A to B (Heine 2002). It offers further support to the proposal (Chafe 2002) that everything in grammar has a reason. It could be for cognitive or social reasons (Su 2016), or historical reasons. In the present study, the discussions demonstrate both cognitive and social reasons for the development of grammaticalization.

In the following sections of this report, Section 2 introduces some background information about Mandarin Chinese grammar that is directly related to this study, including a brief introduction of the case study. Section 3 introduces a theoretic proposal, phono-syntactic conspiracy, which underpins the case study. And Section 4 presents the case study.

2.   Mandarin Chinese and the current study

2.1   Mandarin Chinese pertaining to the present study

The specific case concerns a two-word unit 不是 *bu35shi51* 'Negative-be (Neg-be)', whose variations have brought changes to a rhetorical question 不是…吗 *bu35shi51… ma* 'Neg-be…question particle (Q) isn't it the case that…' in spoken Beijing Mandarin.

---

1. Currently there is also an important economic factor that is promoting the use of Putonghua for better communication among highly mobilized business vendors and migrant workers.

The present study relies on two sources to determine the usage frequency ranking of 不是, (a) the *Lexicon of Common Words in Contemporary Chinese* (Research Team for Lexicon of Common Words in Contemporary Chinese[2] (hereafter: Lexicon 2008)), which contains 56,008 commonly used words based on written texts (e.g., literature, news report, and online discourse); and (b) the Online Corpus of Chinese Discourse by the Center for Chinese Linguistics at Peking University (CCL corpus, Zhan, Guo, and Chen 2003), which contains both written texts with diversified genres and transcripts of spoken Beijing Mandarin Chinese. One can search for the discourse environment and numbers of occurrences of any word from this corpus. The Lexicon (2008) lists usage frequency rankings based on both the broadness of usage in different genres and occurrence frequency of a word. It ranks the verb 是 *shi51* 'be' and the negative particle 不 *bu51* as #2 and #5 of the highly frequently and widely used words respectively (Lower numbers indicate higher usage frequency). It lists 不是 as a two syllable word, ranking at #564,[3] or in the top 1% of the commonly used words.

The CCL corpus (Zhan et al. 2003) is a relatively more balanced online corpus that combines both spoken and written Chinese discourse. From this corpus, the occurrence tokens of 是 (close to 2,500,000) and 不 (close to 2,200,000) are certainly very high. The total occurrence tokens of 不是 *bu35shi51* 'Neg-be' are 216,996. As a comparison, the two-syllable word 这些 *zhe51xie55* 'these' (usage frequency rank #85, Lexicon 2008) has the total occurrence token of 215,626 in this corpus. The word 关系 *guan55xi* 'relationship' (usage frequency rank #101) has a total occurrence token of 195,302. So it could be inferred that 不是 *bu35shi51* 'Neg-be' could be among the first 100 highly frequently used words or expressions.

The expression 不是 in the construction 不是…吗 *bu35shi51… ma* 'Neg-be… question particle (Q) isn't it the case that…' is currently produced in four phonological patterns, centering on the various phonological forms of the verb 是 *shi51* 'be'. The phonetic transcription in this study uses *Han51yu214 pin55yin55* (*pinyin*) 'Chinese spell-sound', the official Romanized system for Mandarin Chinese in Mainland China.

---

2. This book was commissioned by the State Language Commission of the Chinese Ministry of Education. It took some 30 renowned Chinese linguists 10 years to complete. The book provided 56,008 common words ranked by usage frequency.

3. In spoken Northern Mandarin, 不是 also functions as a noun meaning 'one's fault'. But this usage rarely occurs in written discourse unless reporting a conversation. So it is unlikely that this nominal function be ranked so high in the Lexicon (2008).

(1) 不是　　 … 吗?
　　 Negative-be … Q
　　 a.　不是… 吗　 *bu35shi51*　 … 　*ma*[4]
　　 b.　不是… 吗　 *bu35* ↲　　 … 　*ma*
　　 c.　不(?) … 吗　 *bu35*　　　 … 　*ma*
　　 d.　不… 吗　　 *bu51*　　　 … 　*ma*
　　 　　'Isn't it the case that…?'

This rhetorical question in general expresses speakers' attempt either to confirm or to emphasize some truth or the validity of a referent or an event, or to introduce new information in conversations. These functions are illustrated in Section 4.

　　 The four patterns presented above (1a–d) are explained after an introduction to some basic information about Mandarin Chinese. Mandarin Chinese is a tonal language with four lexical tones. Every stressed syllable has a designated tone that differentiates meaning. Using numbers 1–5 to represent pitch values from low to high (Chao 1968: 25–26;[5] Li and Thompson 1981: 8), the following illustrates the tones:

(2) First tone:　　 55　　 High level
　　 Second tone:　 35　　 High rising
　　 Third tone:　　 214　　 Dipping/falling-rising
　　 Fourth tone:　 51　　 High falling

The pitch register remains high at level 5 for the first tone *55*. For the second tone *35* the pitch starts mid, and then rises high to level 5. In the third tone *214*, the pitch begins at a relatively low level 2 and moves down to level 1, then it rises to level 4. For the fourth tone *51*, the pitch starts high and falls sharply down to level 1. In addition, when a syllable is unstressed it takes a short pitch register without contour fluctuation. This has been considered the neutral tone. In the data presented in Section 4, if a syllable is followed by one digit instead of two, it indicates that the duration of the vowel and the tone in this syllable are shortened without pitch movements. A neutral tone is indicated if a syllable does not take a numerical digit.

　　 The four tones are not always produced in the citation forms in connected speech. There are tone sandhi rules (Chao 1968: 25–30; Li and Thompson 1981: 8–9) as part of Mandarin phonotactics. In natural speech, tone sandhi rules are strictly followed. A wrong sandhi may disrupt the flow of intonation, often

---

4. The numerals following each syllable indicate Mandarin Chinese tones. See Appendix for transcription notations.

5. Chao (1968) based all his discussions on spoken Beijing Mandarin. But the tone descriptions in his study represent standard Mandarin Chinese as well (Lin 2011).

causing a recycling repair (Tao, Fox, and Gomez de Garcia 1999; Tao 2018). The sandhi rules that are directly related to the present study are on the negative particle 不 *bu51* below (It also applies to the numeral 一 *yi55* 'one', Tao 2002).

(3) a. The negative particle takes the fourth tone *bu51* when followed by the first three tones *55, 35,* and *214.*
           E.g.,   *bu51 ting55* 'not listen'
                      bu51 lai35 'not come'
                      bu51 hao214 'not good'
       b. It takes the second tone *bu35* when followed by the fourth tone *51.*
           E.g.,   *bu35 shi51* 'not be / is not'.

The minimal word (词 *ci35*) in Chinese is a single syllable, in the form of (C)(G) V(X) (Lin 2011: 430), in which C refers to a consonant, G is a glide, V stands for a vowel or syllabic consonant, and X can be one of the nasals *n/ŋ*. A syllable may contain a minimum of one nucleus (vowel), realized in writing by one character. The majority of commonly used contemporary Chinese words contain two or more syllables/characters, about 94% (Lexicon 2008). But among the top 100 most frequently and widely used words, 82% are single syllable/character (Lexicon 2008; Tao 2018). This skewed distribution is also reflected in spoken lexicon: 82.3% of the active lexicon are single-syllabic words among the top 100 most frequently used words (Tao 2015). The author has noticed that these single-syllable words are mostly grouped into multi-syllable units in conversational speeches, but this observation requires further investigation to confirm. For now, the unit *bu35shi51* is an instance of such a phonological unit.

In informal speeches, Chao (1968: 37) noted that certain consonants in Beijing Mandarin were turned into "voiced continuants" when uttered rapidly. The retroflex /zh, ch, sh/ were produced as /r/, and the palatal /j, q, x/ as /I/. So the voiceless fricative /sh/ is often uttered as /r/, and the onset /ri/ becomes a 'frictionless continuant', which is an 'approximant' (Ladefoged 1964).

2.2 The current study

Now going back to (1), the variations of this construction are on the lexical item 是 *shi51* 'be' and the tone of the negative particle *bu51/35*. Each variation is considered a pattern of this construction. Pattern (1a) shows how the unit 'Neg-be' is pronounced in both Beijing Mandarin and standard Mandarin Chinese. The next one, (1b), illustrates how the verb 'be' 是 *shi51* is actually uttered frequently in Beijing Mandarin. The whole syllable *shi51* has been reduced to an approximant, a syllabic /ɻ/ with vowel-like quality (Reetz and Jongman 2011: 186–9; Ashby and

Maidment 2005), maybe a further development from what Chao observed (1968). Specific acoustic information of /ʐ/ is discussed in Sections 4.

Pattern (1c) may be a variation of (1b) when the approximant /ʐ/ is not fully articulated in casual conversations, but the negative particle 不 *bu35* still maintains its high-rising tone irrespective of its immediately following tone (See (3) for tone sandhi changes on 不). This phenomenon implies that most likely, 是 *shi51* is still in the speaker's mental representation of this two-syllable unit 不是 *bu35shi51* 'Neg-be'.

In (1d) a new construction emerges for the same rhetorical question. Specifically, when 是 *shi51* 'be' is eliminated (1d), the tone of the negative particle 不 *bu* depends on the tone of the lexical item immediately following it. This tonal change has moved *bu51* away from its original environment *bu35shi51* (i.e., construction A) to a new construction B (Bybee 2000, 2003, 2010, 2013; Bybee and Scheibman 1999; Heine 2002; Hopper 1994, 1999). In the meantime, its negation scope has also changed.

One of the functions of *shi51* 'be' is to emphasize or confirm a referent or an action. The negative form 不是 *bu35shi51* 'negative-be' may also serve this emphatic function. Generally speaking, "what follows the negative particle is what is being denied" (Li and Thompson 1981: 629). The negative particle 不 *bu35* negates 是 *shi51* 'be' in the rhetorical question. Without 是 *shi51* 'be' (1d), *bu* has changed its scope of negation.

A written sentence in the pattern (1d) may become ambiguous. Depending on the tone, 不 *bu* and sentence-final question particle 吗 *ma* could form a genuine question (e.g., 他不出国了吗? *Ta55 bu51 chu55 guo35 le ma* 'Is he not going abroad (Did he change his mind)?'). It could also present a rhetorical question when *bu35* takes a rising tone (e.g. 他不出国了吗? *Ta55 bu35 chu55 guo35 le ma* 'Isn't it the case that he's gone abroad?), which expresses the speaker's epistemic stance questioning something based on certain facts or presumptions.[6] This dual interpretation illustrates a change of the scope of the negative particle 不 *bu51*. This part is further illustrated in Section 4.4.

This study proposes that the variations and their original construction (e.g. 1a–d) all function as exemplar representations of the same rhetorical question in the mental grammar of Chinese speakers. The findings, therefore, offer another instance supporting the proposal that language is dynamic in that it constantly develops through usage. What starts as a variation of a construction in a speech community may become a new one; consequently, daily usage is central to grammatical changes and it directly impacts communication.

---

6. The latter is an actual utterance of a TV show, and the written part is the subtitle.

3.   Phono-syntactic conspiracy and grammaticalization

Phono-syntactic conspiracy (Tao 2002, 2006, 2009) describes some synchronic developments of phonological fusion that 'conspire' to cause a syntactic change. Specifically, words that are frequently used together in a construction often become phonological chunks (Bybee and Hopper 2001; Bybee and Scheibman 1999; Haiman 1994; Hopper 1998). Chunking has been proposed to underlie general processes in human memory (Anderson 1993; Newell 1990), which applies to all human activities. A linguistic chunk is often fused or bonded phonologically (Bybee 2000, 2006; Bybee and Hopper 2001; Bybee and Scheibman 1999; Haiman 1994; Hopper 1998). This process describes exactly how the phono-syntactic conspiracy takes place in Beijing Mandarin, starting from consonant lenition to phonological simplification, resulting in syllable reduction and a new construction.

The first instance of these processes in Beijing Mandarin (Tao 2002, 2006) is on the noun phrase (NP) 一个+Noun *yi35ge*+N 'one-classifier-Noun', which now has a co-variant in the spoken language in a new construction 一+N: *yi35*+N 'one+N'. Specifically, the combination of the numeral *yi35* 'one' and the most frequently used classifier *ge51* forms a unit that has been used the most frequently of all *numeral+ classifier* combinations in Mandarin Chinese (Tao 2002, 2006). Data from natural conversations revealed that the unit is often produced with a phonological fusion: *yi35'ə*+N, where the stop /g/ is reduced into a glide, and then it is completely dropped as a process of intervocalic consonant deletion (Ingram 1989: 371–3). Furthermore, the schwa /ə/ is eventually dropped, possibly due to syllable simplification (of two adjacent vowels), resulting in a new construction *yi35*+N. The numeral *yi35* 'one' retains the high rising tone that no longer follows the Mandarin tone sandhi rules (see (3) above), as conditioned by the classifier 个 *ge51*. The newly developed classifier-free NP *yi35*+N has assumed an ambiguous construction in written Chinese such as 一车, which could have two different meanings depending on the tone of the numeral: *yi35 che55* 'a/one car'; or *yi51 che55* 'a carload (of…)'. The former is an NP with a numeral and a noun, the latter is a numeral plus a measure word/classifier (Tao 2006, 2009).

This study presents another instance of language change in Mandarin Chinese as presented briefly in (1). The processes offer another instance of phono-syntactic conspiracy, starting from sound erosion and ending in a syntactic change. This new development differs slightly from the one as discussed above (Tao 2006, 2009). It has involved, in addition to consonant lenition, social factors that may have triggered the complete elimination of the verb 是 *shi51* 'be'. The factors that caused the change may also include rapidly developing technology that enables vast media assistance in spreading the variation; cognitive processes of the pho-

nological lenition; and the influential status of spoken Beijing Mandarin. All may have contributed to the normalization of the new construction (1d).

4.   Usage-caused grammatical change

4.1   The data and the specific construction of this study

This study is based on three data sources: first, audio and video-recorded natural conversations by native Beijing Mandarin speakers, all with IRB approvals and speakers' consent. All recordings were done at different homes in Beijing over the past two decades. The second source is videos of mini TV-series that were posted online (YouTube); third, written documents (CCL Corpus, Zhan et al. 2003). The data from the first source have offered acoustic information concerning consonant lenition and variations of the same construction in Beijing Mandarin. The data from the second and third sources reflect how cognitive and social factors assist the spread of the variations into standard Mandarin, which may have further *boosted* the grammaticalization processes.

The function of the rhetorical question (1) is illustrated in (4) below. In (4a), the speaker came to see an old colleague after a decade. As soon as a young woman opened the door, the visitor made the utterance. By choosing this rhetorical question the speaker expressed her epistemic status – seeing her friend yet doubting it because of the age difference. It turned out that she saw her friend's daughter.

(4) a. 你 不是   那个   XX吗?
        Ni21 bu35shi51 nei51ge XX ma
        2SG NEG-be   that-CL XX Q
        唉 不对   不对!
        ai bu35dui51   bu35dui51[7]
        oh NEG-correct NEG-correct
        'Aren't you that XX? Oh No, not true!'
    b. 那   不(是?) 也是   那
        Na51 bu35'r ye21ri31 nei5
        That NEG-be also-be   that
        医疗   器械厂   占了   吗!
        yi55liao35   qi52xie31chang21 zhan21le   ma
        medical-equipment plant   occupy-ASP Q
        'Wasn't that also occupied by that medical-device plant!'   (Beijing, 2004)

---

7. This is an actual example but was not recorded; so the lexical tones are presented in their citation form.

The function of (4b) differs from (4a) in that mutual stance had already been established. The conversation was about how people gave up their private properties 'voluntarily' in the 1970s. When the utterance was made, all participants already had knowledge of the event; therefore, the speaker was using this construction to make an emphatic statement.

4.2   Usage frequency, consonant lenition and an optional allophone

The grammaticalization processes discussed in this study mainly concern the phonological reduction in 不是 *bu35shi51* 'NEG.-be'. The phonological changes in 不是 happens to have impacted the rhetorical question 不是…吗 *bu35shi51 … ma*. This section discusses the phonological processes that lead to the elimination of 是 *shi51* 'be' in this rhetorical construction.

As mentioned earlier in this report, the verb 是 *shi51* 'be' and the negative particle 不 *bu51* both have very high usage frequencies, ranking at #2 and #5 respectively, and the combination of the two 不是 also has a high usage frequency. Based on the CCL corpus (Zhan et al. 2003), there are about 11,723 instances of the rhetorical question 不是…吗 *bu35shi51…ma* 'NEG-be … Q', all seem to belong to spoken language.[8] Notice that the expression 不是 *bu35shi51* occurs far more frequently (over 200,000 instances in the CCL corpus (Zhan et al. 2003)) than in this rhetorical question. Yet, as is explained later in the next section, the phonological changes in this expression just happen to have impacted this construction.

In conversational speeches, when elements are frequently produced together, they are often produced as a prosodic unit[9] (Bybee and Hopper 2001; Chafe 1994), which may not necessarily form a grammatical phrase. High usage frequency of such units may lead to phonological chunking, a decisive process for language development (Bybee 2000, 2003, 2006, 2013; Bybee and Scheibman 1999; Haiman 1994; Hopper 1998) because within a chunk, elements often tend to be fused or bonded phonologically. Currently, the fricative /sh/ in *bu35shi51* is usually uttered as an approximant /ɻ/ in Beijing Mandarin (possibly in other Northern dialects as well, as Chao (1968) proposed).

The processes from fricatives to approximants follow a natural course: At the intervocalic position, a consonant becomes a weaker phoneme. Consonant

---

8. Although not all 11,723 instances of this rhetorical question have been checked, the author randomly selected 50 instances to check their sources. All turned out to come from either spoken data or report of oral speeches.

9. It is also called an intonation unit. In Mandarin Chinese such units may not be a multisyllable word or a grammatical phrase. They are often marked by an intonation contour with 'fused' phonological features (Bybee 2006; Tao 2006).

weakening from a fricative to an approximant "increases the permeability of the vocal tract to airflow" (Lass 1984: 177). This explanation points to the fact that the essence of this phonological development is language usage, which is through everyday oral communications.

The /ɻ/ in contemporary Beijing Mandarin is often syllabic, standing alone for the syllable *shi* as in 不是 *bu35ɻ* 'NEG-be'. The /ɻ/ carries the acoustic quality similar to that of the alveolar rhotics as described in Boyce, Hamilton and Rivera-Campos (2016). It differs from the syllable onset or coda /r/ in Mandarin Chinese, which has been well documented. The onset /r/ is a phoneme. As a coda it is *er-35hua51yin55* (Chen 1999: 39; Duanmu 2007: 218; Lin 2007: 182), a typical practice in Beijing Mandarin. But the rhotic approximant /ɻ/, an optional allophone to the consonants /zh, ch, sh/, has not been discussed much in the previous literature except for Chao (1968), also summarized in Zhang (2005). Norman (1988: 142) proposed that /sh/ is one of the syllabic fricatives in Mandarin Chinese with weak friction that extends to its following vowel. The /ɻ/ in 不是 *bu35ɻ* may have been a further development from Norman's observation. It has become a frictionless continuant, i.e., approximant, with a weaker sonorant than that of a vowel.

In Beijing Mandarin, /sh/ in *bu35shi51* is produced in various patterns (1a–c). These variations are not discrete but are produced in a continuum. They are most likely stored in speakers' memory as exemplar representations of the same consonant/syllable. Instances of (1b–c) are presented in Section 4.3. To present objective information of the variations, some general features of /ɻ/ are first discussed below.

Reetz and Jongman (2011: 186–9) propose that approximants have special acoustic properties similar to those of vowels that are produced at a comparable location in the vocal tract. Their formant patterns are clear but somewhat weaker than for vowels because of their slightly greater constriction; so it results in a shorter steady-state portion and lower acoustic energy. Therefore, approximants are distinct by being shorter and weaker than vowels and by having longer formant transition durations than vowels.

The rhotic approximant is typically uttered in American English with three simultaneous constrictions, all contributing to lowering the third formant F3 (Reetz and Jongman 2011: 188): lip rounding, a narrowing near the alveolar ridge involving the tip of the tongue, and a narrowing in the pharynx due to retraction of the tongue root. Different from the English rhotics, the alveolar rhotic approximant[10] /ɻ/ in Beijing Mandarin does not involve lip rounding. But the lack of lip

---

10. Not all rhotic approximants have the vocal constriction at the alveolar ridge. It is more common (in American English) to show the pharyngeal constriction (Boyce et al. 2016). For this study since /ɻ/ functions as an allophone to /sh/, an alveolar fricative in Chinese, it is assumed this particular /ɻ/ may also have the vocal constriction at the alveolar ridge.

rounding/extrusion does not affect the lowering of the third formant (the maximum amount of F3 lowering due to rounding/protrusion is about 200 Hz, Boyce, personal communication).[11] Therefore, an objective measurement of the utterance of /ɻ/ in the data of this study is the lowered F3.

An important factor in this grammaticalization process is the weakening of the rhotic approximant /ɻ/. Currently, the unit 不是 *bu35shi51* 'NEG-be' is often uttered with two distinctive acoustic features, the high-rising tone in *bu35* and a very weak syllabic /ɻ/ (1b). The /ɻ/ is always unstressed; losing its designated fourth tone as in *shi51* 'be'. Loss of stress and reduction to a neutral tone are early indications of phonological reduction (Bybee et al. 1994: 107).

4.3   Tone sandhi, the rhotic approximant and exemplar representations of /sh/

This section discusses patterns (1b–c) of the rhetorical question that currently occur in Beijing Mandarin. The variations come from how 不是 *bu35shi51* is produced. This dual-syllable unit is often produced with two special features: the high-rising tone of the negative particle 不 *bu35*, and the rhotic approximant /ɻ/ in various clarities. The presence of /ɻ/ is important as the intermediate stage to the syntactic development (See (1)). When /ɻ/ becomes syllabic, it stands alone for 是 *shi51* 'be'. Even when it is not clearly uttered except for maybe an articulatory gesture (1c), to Beijing Mandarin speakers the construction *bu35shi51* is most likely still present in their mental representation because the tone of *bu35* remains high-rising, irrespective of the tone sandhi rules (3). To non-Beijing Mandarin speakers, however, the /ɻ/ is not easily perceived even when it is clearly uttered.[12]

A specific acoustic feature of /ɻ/ is the lowering of the third formant F3 (Boyce et al. 2016; Ladefoged 2003; Reetz and Jongman 2011). The /ɻ/, being an optional allophone to the consonants /zh, ch, sh/ (Chao 1968), is uttered with reduced stress and different intensities depending on pragmatic needs. Some of the recorded data had been shown at a conference.[13] All the audience acknowledged being able to identify the /ɻ/ although not all could be documented in Praat[14] for acoustic analyses. In this report, whenever the /ɻ/ is described as 'audible' it reports the objective

---

11. Suzanne Boyce, December 11, 2016.

12. Li, Aijun, the Institute of Linguistics, Chinese Academy of Social Sciences. April 21, 2016.

13. The Fourth International Symposium on Chinese Language and Discourse, June 11–12, 2016, Bryant University, Rhode Island, USA.

14. Praat is a program for speech analysis and synthesis. http://www.fon.hum.uva.nl/praat/

responses from the audience at that conference. Otherwise, the audibility of /ɻ/ is pointed out by the author, a native Beijing Mandarin speaker.

The following examples all contain 不是 *bu35shi51* from recorded natural conversations. The utterances exemplify patterns (1b–c) in Beijing Mandarin to illustrate the synchronic variations and grammaticalization of the construction 不是…吗 *bu35shi51…Q* 'isn't it the case that…?' The figures are produced from two programs, Praat and BLISS.[15] The waveform and pitch information were produced from BLISS. The correspondence of pitch and word is based on information from Praat. The presence of /ɻ/ is objectively measured by the formant information, calculated from Praat.

Example (5) illustrates how /ɻ/ is uttered in different variations. Speaker B, who was in his 70s, moved into Beijing from the suburb of the city as a young child. There are two instances of 不是 *bu35ɻ*, first a brief answer to a question, then some further elaboration of the answer. The first /ɻ/ was produced with a sudden cut-off plus a slight pause, offering a clear epistemic stance. The abrupt cut-off made it almost impossible to measure the acoustic information of /ɻ/, although audible. The second /ɻ/ is produced with a slight sound extension, which offers clearer acoustic information.

Figure 1a presents the waveform and pitch movement of part of the utterance from speaker B. It shows the rising tone of the negative particles *bu35*. Figure 1b presents some limited formant information of lowered F3, indicating the presence of the approximant /ɻ/.

(5) A: 北京　　　的?
Bei21jing55 DE
Beijing　　person
'Are you a native Beijing?'

→ B: 不是　(0.14) 倒　　不是::: 原籍
Bu35ɻ- (0.14) dao51　bu35ɻ22 yuan35ji35
NEG-be　　　contrary NEG-be　origin
倒　　不是
dao51　bu35ɻ2
contrary NEG-be
'No. Not really. (My) original hometown was not really (Beijing).'

(Beijing 04)[16]

a.
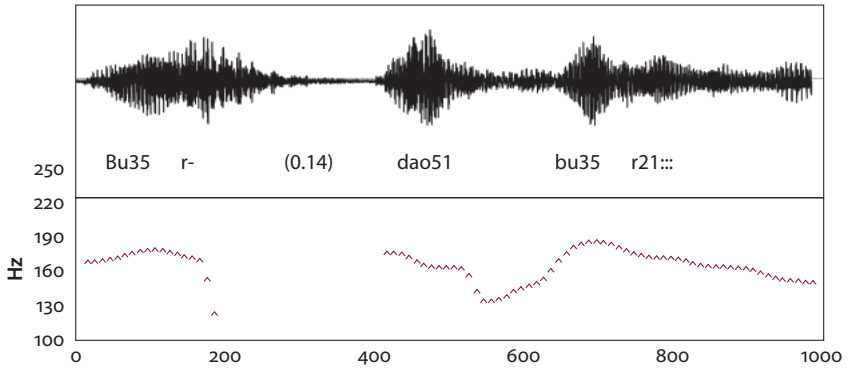
Time_s	F1_Hz	F2_Hz	F3_Hz	F4_Hz
0.752448	619.009187	1279.371814	2612.044541	4299.298604
…		…		…
0.764948	601.003985	1340.523777	1949.088415	3357.980676
0.771198	616.376942	1275.673209	1926.786211	3759.338808
0.777448	790.464795	1067.998417	1826.105571	3748.367995
0.783698	631.181035	872.616766	1772.483905	3291.447388
0.789948	801.769576	1798.197067	2958.997048	4413.506134
…		…		…
0.802448	1039.450925	1702.146453	3129.435911	4741.995357

b. Formant information of /ɻ/ from Figure 1a with lowered F3

Figure 1. Pitch, soundwave and formant information of (5)

Figure 1a shows that both negative particles *bu35* were produced with a rising pitch movement. From 1b we can see lowered F3 for the second /ɻ/ (corresponding time and F3 underlined) compared to the F3 formants produced roughly before and after.

The rest of the examples presented later on all contain the construction 不是…吗 *bu35shi51… ma* 'NEG-be … Q' (1b–c). Pattern (1b) appears in (6–7), where the two NPs after 不是 *bu35shi51* 'NEG-be' carried new information, but were presented in this construction as if the information had already been accessible, probably a means to maintain solidarity.

(6) 他   那   不是   水瓶儿   吗? 老丢。
     Ta55 nei31 <u>bu35ɹ</u>   shui21pingr35 ma lao21 diu55[17]
     3sg   that   neg-be water bottle   Q   often lose
     'Isn't it the case that he has a water bottle? (It) often gets lost'.

<div align="right">(Beijing Dec.04)</div>

a.
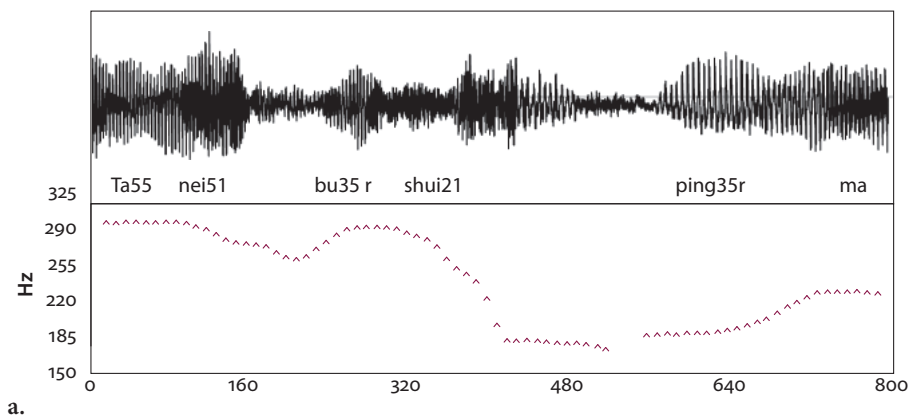
Time_s	F1_Hz	F2_Hz	F3_Hz	F4_Hz
0.215000	1040.66306	2007.974036	3169.909848	4345.663316
0.221250	747.159274	1766.296545	2631.132465	4160.443348
…		…		…
<u>0.252500</u>	502.560355	1072.709825	<u>2499.485306</u>	3951.619071
<u>0.258750</u>	478.602880	1186.369541	<u>2479.633695</u>	3865.785135
<u>0.265000</u>	454.875099	1244.802514	<u>2566.113489</u>	3843.896438
<u>0.271250</u>	544.303208	1279.520326	<u>2613.547658</u>	3823.159594
…		…		…
0.290000	755.904798	1683.504515	2885.146150	4141.221522
0.296250	716.769064	1734.348936	3024.140739	4299.838077
0.315000	1586.31975	2620.115069	3487.656740	4760.698373

b. Formant information of Figure 2a. Note the F3 lowering

Figure 2. Pitch, soundwave and formant information of (6)

In (6) the rhetorical question introduced a new referent, a water bottle, at the start of a turn. The speaker is a teenage girl with relatively high pitch (F0) and fast speech rate. The NP 不是 *bu35ɹ* is uttered in pattern (1b). Figure 2a shows an upward pitch movement of the negative particle *bu35*, indicating the rising tone.

---

17. The word *lao214 'old'* describes repeated events in informal Beijing Mandarin.

Figure 2b shows lowered F3 where the /ɻ/ is produced. The F3 in this example is not drastically low as documented, which is usually below 2000 hz (Boyce et al. 2016; Ladefoged 2003; Reetz and Jongman 2011); yet the audible /ɻ/ and the lowered F3 together may indicate the presence of /ɻ/.

Figure 3 shows the pitch, sound and F3 information of the next Example (7). Figure 3a shows a rising pitch/tone for 'bu35r', and 3b shows a lowered F3, indicating the presence of /ɻ/.

Example (7) also shows Pattern (1b). It is produced by an older woman, whose utterance contained *bu35ɻ* and an audible /ɻ/.

a.
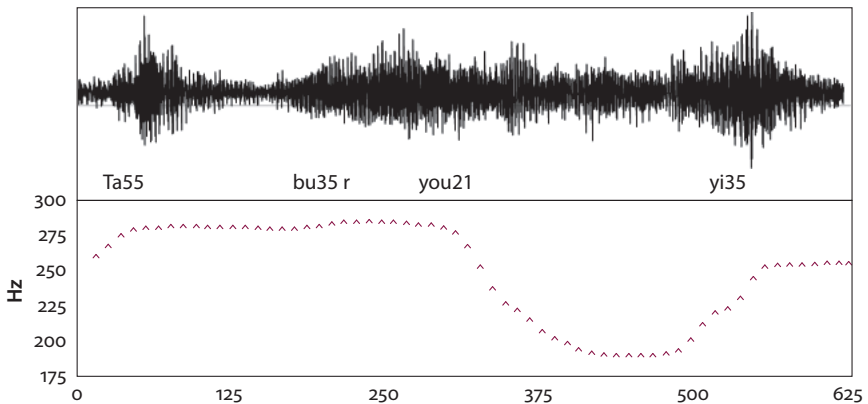
Time_s	F1_Hz	F2_Hz	F3_Hz	F4_Hz
0.155000	1085.279133	2058.427955	3303.205325	4133.976928
…	…	…	…	…
0.195000	739.461669	1393.142384	2021.113889	4143.829008
0.197500	750.104653	1385.174539	1968.392386	3519.303816
0.200000	731.812478	1387.691307	1966.539991	3431.387166
0.202500	665.481785	1417.113522	1970.589183	3539.591694
…	…		…	…
0.212500	690.602967	1394.116400	1978.000422	3456.655829
0.215000	668.887303	1413.509812	1962.201990	3449.303345
0.217500	639.391427	1394.617835	2006.745012	3463.534785
…		…		…
0.227500	590.488117	1456.309431	2150.009262	3440.226687

b. Formant information of Figure 3a. Note the F3 lowering

Figure 3. Pitch, soundwave and formant information of (7)

(7) 他 不 是   有 一   北关   环岛   吗?
     Ta55 <u>bu35ɻ</u> you21 yi35 bei21guan55 huan35dao21 ma
     3sg   neg-be exist one   Northern Gate Round-Island Q
     'Isn't it the case that there is a Northern Gate Rotunda Island?'

<div align="right">(Beijing June_13)</div>

The speaker in (7) used the rhetorical question to introduce a landmark, new information to the hearer. This example also shows a lowered F3 and a rising tone of the negative particle (Figure 3), indicating the utterance of pattern (1b) *bu35ɻ*.

Example (8) repeats (4b), which presents an emphatic function of this rhetorical question in the pattern (1c). The topic was about the participants' past experiences having to give up their properties for a neighborhood-run manufacturer in the 1970s. The speaker was a middle-aged man.

(8) 那   不(r?) 也是   那 医疗器械   厂   占 了   吗!
     That neg-be also-be that medical-equipment plant take over-asp Q
     'Wasn't it also taken over by that medical-device plant!   (Beijing, 2004)
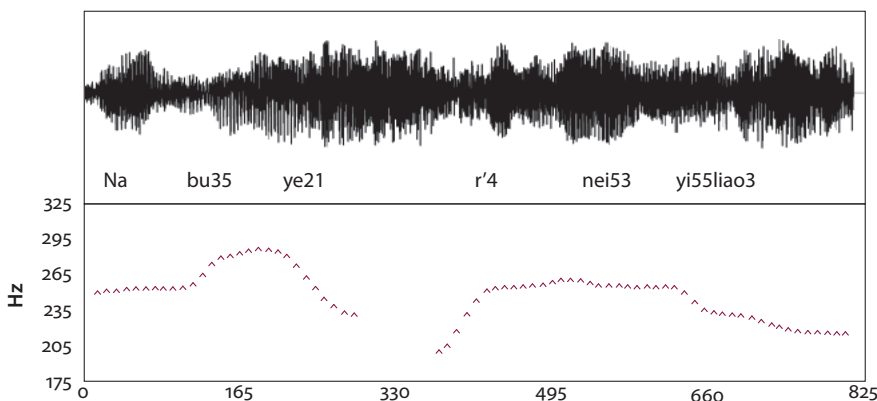
Figure 4. Pitch and soundwave information for (8)

The upward pitch movement indicates the rising tone of the negative particle *bu35* in Figure 4. But there was no lowered F3 calculated from Praat to indicate the presence of /ɻ/. However, the rising tone of 不 *bu35* indicates *shi51* is still in the speaker's mental representation to cause the rising tone. Otherwise, *bu51* should have taken the falling tone, following sandhi rule (3a).

Examples (5)–(8) all came from recordings of natural conversations by native Beijing Mandarin speakers, which illustrate two variations of the unit 不是 *bu35shi51* 'neg-be' (1b–c). The rhotic approximant /ɻ/ is produced in various clarities, from a clearly lowered F3 (5), a slight F3 lowering (6–7) to no trace of F3 lowering at all (8). Despite of the various clarities of /ɻ/, the negative particle 不 *bu35* retains a rising tone as conditioned by the falling tone of 是 *shi51* 'be' (3b).

The next section presents data on the further development of this construction into Pattern (1d).

4.4   Language variation, grammaticalization, and language change

4.4.1   *Misperception and transcription*

This section displays the development of pattern (1d). The acoustically reduced /ɻ/ may not always be clearly captured by non-Beijing Mandarin speakers. So some misperception may have occurred. Some propose that 不是 *bu35shi51* 'ɴᴇɢ-be' has become 不 *bu35* with a rising tone that no longer follows the tone sandhi rules (Wiedenhof 1995). While this study is unable to verify Wiedenhof's data, the fact that the rhotic /ɻ/ is still present in current Beijing Mandarin indicates that most likely, the /ɻ/ might have been present in his data but was not captured.

This misperception can also be observed in transcripts of spoken Beijing Mandarin, as exemplified in (9). It came from a two-people comedian talk (相声 *xiangshe* 'appearance and voice – crosstalk') in typical Beijing Mandarin. These two performers (Guo and Yu) basically act instantaneously,[18] so the script was later transcribed. The character version of (9) came from the transcript (CCL Corpus, Zhan et al. 2003), in pattern (1d) of this study. The *pinyin* version was the actual utterance, extracted from the live show (YouTube), in pattern (1c).

(9)   不　　说好　　　让　　我　　唱　　吗
　　　Bu35(r) shuo55hao21 rang51 wo21 chang51 Q
　　　ɴᴇɢ-be say-good　　let　　1sɢ　sing　　Q
　　　'Didn't we reach a verbal agreement to let me sing?'

(CCL Corpus (Zhan et al. 2003); Guo and Yu)
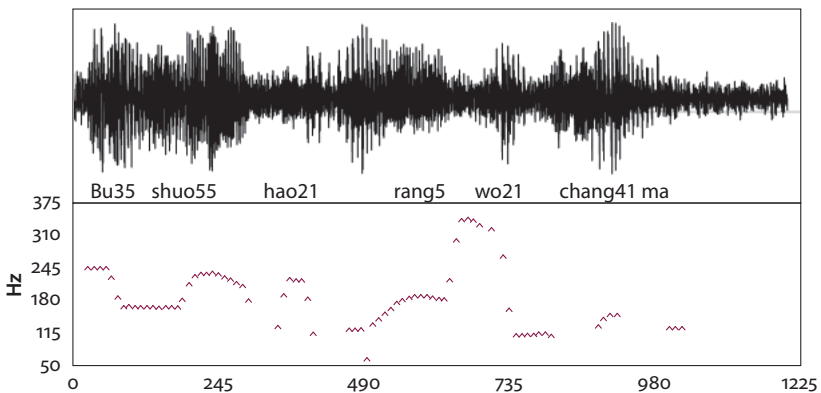
Figure 5. Pitch and Soundwave information for (9)

**18.** Direct report from the performer: https://www.youtube.com/watch?v=-ZyDVofmNHo

The rhetorical question in (9) expressed a challenge, blaming the hearer for break-ing a verbal agreement. This example differs from the previous ones phonologi-cally. It may have involved assimilation. The syllabic /ɻ/ after 不 *bu35* and the onset /sh/ of the next word 说 *shuo55* 'speak' could have been assimilated, leaving only a slight germination of /ɻ/. The negative particle 不 *bu35* has a clear rising tone irrespective of sandhi rule (3a), indicating it is conditioned by *shi51*, following sandhi rule (3b).

The written script in (9) is in pattern (1d) 不…吗 *bu…ma* 'NEG…Q'. Syntac-tically, the resultative construction VP 说好 *shuo55hao214* 'speak-good: reach an agreement' should take the aspectual negative particle 没 *mei35*. Here 不 has as-sumed a new grammatical function to negate the perfective VP 'having reached a verbal agreement', breaking away from its original function to negate its following verb 是 *shi51* in this pattern (Li and Thompson 1981).

Example (9) is not an accidental case. There are about over 260 instances of the construction 这不是…吗[19] in pattern (1d) in the CCL Corpus (Zhan et al. 2003), all are either transcripts of spoken discourse or quotations of speeches in written reports.[20] Examples (10a–b) illustrate two instances from the corpus. Example (10a) is from transcriptions of spoken Beijing Mandarin (a survey of spo-ken Beijing Mandarin in 1982), and (10b) is from a quotation of spoken language from a written report.

> (10) a. 这　不　前　三门儿　　吗?
> 　　　Zhei51 bu　qian35 san55 men35r ma
> 　　　These　NEG front　three gates　　Q
> 　　　'Aren't these the three Front (City) Gates?'[21]
> 　　b. 这　不　来了　　吗!
> 　　　Zhe51 bu　lai35 le　　ma
> 　　　This　NEG come-PERF Q/Interjection
> 　　　'Isn't it the fact that (the car driver) is already here!'

It is unclear which tone would be used on the negative particle 不 *bu* when (10a–b) were uttered, but it is very clear that 不 *bu* does *not* function to negate the NP (10a) or the VP (10b) that follows it. Instead, it is part of the construction 不…吗 as a rhetorical question, and 不 *bu* depends on the question particle 吗 *ma* to in-dicate its function. In (10a) the construction is to indicate something new while

---

19. The first word/grammatical subject of this construction allows a highly diversified nominal, such as 他, 你, 那, etc. The author only searched for one such pattern 这不…吗?

20. Interested readers are recommended to check out the CCL Corpus for more examples.

21. The three city gates at the front side of Tiananmen.

assuming the information is accessible to the hearer. In (10b) the interjection was in response to a question concerning where the speaker's car was. The speaker, a migrant worker, made this claim jokingly that he and his car (a tricycle with a cargo space) have already been there. In this example, 不 clearly does not negate its immediately following verb 来 *lai35* 'come'. Instead, its scope of 'negation' covers the entire VP 'already come'.

It is likely when reading sentences like (9) and (10), the tone of the negative particle is no longer affected by the falling tone of 是 *shi51*. This is the final stage of current grammaticalization processes.

It is important to point out that the sound fusion in spoken Beijing Mandarin is a common practice in conversational speeches, and it is most likely a subconscious practice. Yet when pattern (1d) has been noticed and used in TV shows, then the pattern may have been chosen intentionally to reflect informal colloquial speeches. These two aspects are illustrated in the next section.

4.4.2   *Impact of* 普通话 *Putonghua 'common language'*

This section discusses the impact of Beijing Mandarin to the grammaticalization processes leading to pattern (1d).

The Chinese language includes seven major *dialects*, which in general are not mutually intelligible (Lyovin, Kessler, and Leben 2017: 139).[22] Mandarin Chinese is by far the most widespread and widely spoken of all. To promote communication, the Chinese government has undertaken great efforts to promote 普通话 *Putonghua* 'common language' (Law of the People's Republic of China on the Standard Spoken and Written Chinese Language 2000; Xinhua News Agency 2011), which is 'based upon the Northern dialects with the Beijing dialect as the standard pronunciation' (Chen 1999: 24). There are differences between Beijing Mandarin and *Putonghua*; therefore, technically speaking, *Putonghua* has no native speakers. China's State Language Commission[23] offers an official examination of people's proficiency levels of *Putonghua* for various professions.

The spreading power of *Putonghua* in China has made a great impact on China's linguistic communities. All students receive their education in *Putonghua*, so they are literally 'bilinguals' when they maintain their local 'dialects' outside school. Furthermore, *Putonghua* is the standard for all major media and official communications. Therefore, even if someone does not speak *Putonghua* well, they should be able to comprehend it. The importance of *Putonghua* has inadvertently

---

22. Linguistically these dialects may be considered different languages (Lyovin et al. 2017).

23. There are officially three proficiency levels, each with two scales, to measure the standard that one may achieve in this common language. http://www.moe.edu.cn/publicfiles/business/htmlfiles/moe/moe_66/200408/585.html

promoted Beijing Mandarin to be a 'superstrate' language[24] – informal colloquial conversations on TV or movie shows often try to adopt some Beijing Mandarin. Consequently, many TV shows adopted patterns (1b–d) in their dialogues, and pattern (1d) has been used in written documents too.

Examples (11a–b) present a contrast that is most likely due to this impact. Example (11a) came from an Internet novel, and (11b) was a modified version of the same sentence from a TV mini-series (produced in 2012) based on this novel. The Chinese character version in (11b) is the subtitle to this utterance from the TV show. The *pinyin* version for (11a) is based on standard pronunciation; in (11b) it is based on the actual utterance, extracted from the TV show.

Both (11a–b) function to remind the hearer why the speaker was able to discover an evil plot yet the hearer couldn't. The interjection 嘛 *ma* is a sentence-final particle, mostly used in statements. It collides with the question particle 吗 *ma* at the same syntactic slot so the question particle is eliminated. There is a clear contrast of the same unit 不是 *bu35shi51* 'NEG-be'. The written version (11a) has retained the standard construction 不是 *bu35shi51* 'NEG-be'. But in (11b) the TV adaptation dropped 是 *shi51* completely. Furthermore, the negative particle took a falling tone *bu51*, following sandhi rules (3a).

(11) a.   你   不是   老   跟   我   说
           Ni214 bu35shi51 lao214 gen55 wo214 shuo55
           2SG   NEG-be   always to   1SG   say
           当局   者   迷嘛。
           dang55ju35 zhe214 mi35 ma
           act-in   person unclear INT
           'Didn't you always tell me those inside the game lose their objective
           View!'   (Zhang, W: 女相陆贞传奇 Legend of Lu Zhen, Chapter 39)
      b.   你   不   老   跟   我   说
           Ni21 bu51 lao21   gen55 wo21 shuo55
           2SG   NEG always to   1SG   say
           当局   者   迷   嘛
           dang55ju35 zhe21   mi35   ma
           act-in   person unclear INT
           'Didn't you always tell me those inside the game lose their objective
           View'   (陆贞传奇 Legend of Lu Zhen, Episode 30, 30:33)[25]

---

24. There are other factors that promoted Beijing Mandarin to its current status, since Beijing being China's political, economic and cultural center.

25. https://www.youtube.com/watch?v=94lVqZrb3b8. 30:33 extracted Dec. 31, 2016.
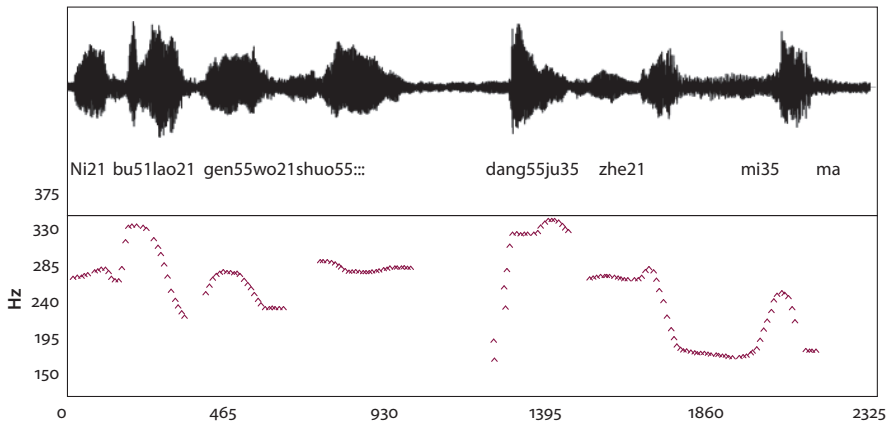
Figure 6. Pitch and Soundwave information for (11b)

The contrast between (11a) and (11b) illustrates the fact that the TV producers must have noticed pattern (1c) in Beijing Mandarin, and tried to adopt that pattern in the show's conversation. However, the high-rising tone in 不 *bu35* in Beijing Mandarin (5–9) might not have been perceived or noticed. Consequently, in the mental representations of the director/producer of this show the pattern involved dropping 是 *shi51* to let the negative particle 不 *bu* follow sandhi rule (3a).

The contrast of (11a–b) shows the impact of Beijing Mandarin. When informal speech is needed in the media, Patterns (1b–c) are likely perceived as a syntactic change into pattern (1d), thus the superstrate Beijing Mandarin helped to complete the grammatical change.

4.5   Interim summary

The newly developed construction 不…吗 *bu…ma* 'NEG … Q', with the negative particle 不 *bu* following the tone sandhi rules (3a–b), does not seem to have been practiced in Beijing Mandarin. At least in the author's recorded conversations, the negative particle *bu35* takes the rising tone while the verb 是 *shi51* 'be' is uttered with varied clarity, either as a fully pronounced syllable 不是 *bu35shi51*, as a rhotic approximant /ɻ/ such as *bu35ɻ* (5–8), or as a fused fricative (possibly) during assimilation or dropped in fast speech (9). But the negative particle 不 *bu35* is always uttered with a high-rising tone.

However, utterances of this rhetorical question in Beijing Mandarin (and possibly in some northern Mandarin dialect)[26] are likely transcribed into written Chinese without 是 (e.g. (9)). When pronouncing these transcriptions, most likely 不 *bu* follows the tone sandhi rules (3a–b).

---

26. This part requires further investigation but the author has a sense that it could be the case.

To be grammaticalized, an item has to be emancipated from its original context (Haiman 1994; Heine 2002). When the negative particle 不 *bu* follows tone sandhi rules (3a–b) in the pattern 不…吗 *bu… ma* (1d), it is 'emancipated' from the high-falling tone of 是 *shi51* 'be'. This step has not been practiced in Beijing Mandarin, but it has been adopted in Chinese media, an indication that the new construction has been acknowledged and accepted.

5.   Conclusion

Language variations are inevitable among any speech communities. Language change may develop out of variations. This study has presented a specific case to illustrate synchronic grammaticalization processes out of phonological variations, which is illustrated with a consonant development pattern: fricative > approximant > syllabic > zero. The first step, fricative > approximant, has been observed in Chao (1968). The present study proposes that the second step, approximant > syllabic, may be a new development under the condition of chunking of a phonological unit, as illustrated in Examples (5)–(8).

Furthermore, the study has proposed two possible causes leading to the last step, syllabic > zero, for grammaticalization to happen in Mandarin Chinese (Examples (9)–(10)). The new construction 不…吗 *bu…ma* 'isn't it the case that…?' differs from the synchronic variations of the original construction 不是…吗 *bu35shi…ma* in three aspects. First, with the elimination of 是 *shi51*, the tone of the negative particle 不 *bu* now follows the tone sandhi rules (a–b). This part differs from how 不 is treated in Beijing Mandarin (Examples (5)–(9)). Secondly, the negative particle no longer negates 是 *shi51* 'be'. Instead, it now depends on the whole construction 不…吗 to negate what follows it, including an NP (10) or VP (9) that used to be negated only by the aspectual marker 没 *mei35*. Thirdly, with the change of its phonological environment, 不 *bu* has assumed a new function, which may have created an ambiguous pattern. Without context, the new construction could have potentially two different interpretations, as a genuine question or a rhetorical question (Section 2).

As mentioned in the Introduction, language variation and change are part of natural human language development out of interaction (Hopper and Traugott 2003). Studies of historical linguistics may offer possible explanations for the connection of constructions A to B (Heine 2002; Hopper and Traugott 2003). But the intermediate stages of the processes can only be inferred. Discourse-functional studies on everyday language use, on the other hand, may document the gradual developmental processes from construction A to B, which is happening

concurrently. This study offers one instance of grammaticalization as it happens in Mandarin Chinese, out of frequent usage.

What is new in this study is the 'outside' force that may have expedited the grammaticalization processes. Highly developed technology and media have made it possible to expose Beijing Mandarin and *Putonghua* to the whole nation. The promotion of *Putonghua* in China has further elevated the influence of Beijing Mandarin. So modern technology and social conditions have contributed to the grammaticalization process.

This study has once again offered evidence to support the usage-based theory and exemplar representations (Bybee 2003, 2013; Pierrehumbert 2001) in language development and grammaticalization. The findings further support the view that language and grammar are fostered and conditioned through everyday conversation of human communication.

Acknowledgements

References

Anderson, John. R. 1993. *Rules of the Mind*. Hillsdale, NJ: Erlbaum.

Ashby, Michael, and John Maidment. 2005. *Introducing Phonetic Science*. Cambridge, UK: Cambridge University Press.

Barlow, Michael, and Suzanne Kemmer. 2000 (eds). *Usage-Based Models of Language*. Sanford, CSLI Publications.

Boyce, Suzanne E., Sarah M. Hamilton, and Ahmed Rivera-Campos. 2016. "Acquiring rhoticity across languages: An ultrasound study of differentiating tongue movements." *Clinical Linguistics & Phonetics* 30 (3–5): 174–201. https://doi.org/10.3109/02699206.2015.1127999

Bybee, Joan. 2000. "The phonology of the lexicon: Evidence from lexical diffusion." In *Usage-Based Models of Language*, ed. by Michael Barlow and Suzanne Kemmer, 65–85. Stanford: CSLI Publications.

Bybee, Joan. 2003. "Mechanisms of change in grammaticization: The role of frequency." In *The handbook of historical linguistics*, ed. by Brian D. Joseph and Richard D. Janda, 602–23. Oxford: Blackwell.

Bybee, Joan. 2006. "From usage to grammar: The mind's response to repetition." *Language* 82 (4): 711–733.

Bybee, Joan. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

Bybee, Joan. 2013. "Usage-based Theory and Exemplar Representations of Constructions." In *The Oxford Handbook of Construction Grammar*, ed. by Thomas Hoffmann and Graeme Trousdale, 49–69. Oxford, UK: Oxford University Press.

Bybee, Joan and Paul Hopper. 2001. "Introduction to frequency and the emergence of linguistic structure." In *Frequency effects and the emergence of linguistic structure*, ed. by Joan Bybee and Paul Hopper, 1–24. Amsterdam: John Benjamins.

Bybee, Joan, William Pagliuca, and Revere D. Perkins. 1991. "Back to the future." In *Approaches to grammaticalization*, ed. by Elizabeth C. Traugott, and Bernd Heine, 19–58. Amsterdam: John Benjamins.

Bybee, Joan, Revere D. Perkins, and William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.

Bybee, Joan, and Joanne Scheibman. 1999. "The effect of usage on degrees of constituency: The reduction of *don't* in English." *Linguistics* 37: 575–596.

Cambridge Dictionary. Accessed December 28, 2017. https://dictionary.cambridge.org/us/dictionary/english/gonna.

Chafe, Wallace. 1994. *Discourse, Consciousness and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: University of Chicago of Press.

Chafe, Wallace. 2002. "Putting Grammaticalization in its place." In *New Reflections of Grammaticalization*, ed. by Ilse Wischer, and Gabriele Dieward, 395–412. Amsterdam: John Benjamins.

Chao, Yuen Ren, 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

Chen, Ping. 1999. *Modern Chinese: History and Sociolinguistics*. Cambridge: Cambridge University Press.

Duanmu, San. 2007. *The Phonology of Standard Chinese* (2nd ed). Oxford: Oxford University Press.

Erbaugh, Mary S., 1986. "The development of Chinese noun classifiers historically and in young children." In *Noun Classes and Categorization*, ed. by Colette G. Craig, 399–436. Amsterdam: John Benjamins.

Fowler, Carol. 1986. "An event approach to the study of speech perception from a direct-realist perspective." *Journal of Phonetics* 14: 3–28.

Goldinger, Stephen D. 1998. "Echoes of echoes? An episodic theory of lexical access." *Psychological Review* 105: 251–279.

Goldinger, Stephen D. 2013. "The cognitive basis of spontaneous imitation: Evidence from the visual world." *Proceedings of Meetings on Acoustics* 19: 1–6.
https://doi.org/10.1121/1.4800039

Goodwin, Charles. 2013. "The co-operative, transformative organization of human action and knowledge." *Journal of Pragmatics* 46 (1): 8–23.

Goodwin, Charles and Marjorie H. Goodwin. 2004 "Participation." In *A Companion to Linguistic Anthropology*, ed. by Alessandro Duranti, 222–244. Malden, MA: Blackwell.

Guo, Degang 郭德纲 and Qian Yu 于谦. 卖吊票 (Mai Diao Piao: Selling Tickets for Hanging Seats. Cross-Talk) https://www.youtube.com/watch?v=x0FKyaZLGtk Accessed February 5, 2017.

Haiman, John. 1994. "Ritualization and the development of language." In *Perspectives on Grammaticalization*, ed. by William Pagliuca. 3–28. Amsterdam: John Benjamins.

Heine, Bernd. 1993. *Auxiliaries: Cognitive forces and grammaticalization*. Oxford: Oxford University Press.

Heine, Bernd. 2002. "On the role of context in grammaticalization." In *New Reflections of Grammaticalization*, ed. by Ilse Wischer, and Gabriele Dieward, 83–101. Amsterdam: John Benjamins.

Heine, Bernd and Tanya Kuteva. 2002. *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press.

Hopper, Paul J. 1994. "Phonogenesis." In *Perspectives on Grammaticalization*, ed. by William Pagliuca, 29–45. Amsterdam: John Benjamins.

Hopper, Paul. J. 1998. "The paradigm at the end of the universe." In *The Limits of Grammaticalization*, ed. by Anna G. Ramat, and Paul Hopper, 147–158. Amsterdam: John Benjamins.

Hopper, Paul. J. 1999. "Conversation and argument structure: The significance of dialogue for grammatical explanation." Paper presented at the International Symposium New Reflections on Grammaticalization. University of Potsdam, Germany.

Hopper, Paul J. and Elizabeth Closs Traugott. 2003 (1993). *Grammaticalization*. Cambridge, UK: Cambridge University Press.

Ingram, David. 1989. *First Language Acquisition: Method, description and explanation*. Cambridge, UK: Cambridge University Press.

Ladefoged, Peter. 1964. *A Phonetic Study of West African Languages*. Cambridge, UK: Cambridge University Press.

Jefferson, Gail. 2004. "Glossary of transcript symbols with an Introduction." In *Conversation Analysis: Studies from the first generation*, ed. by Gene H. Lerner, 13–23. Philadelphia: John Benjamins.

Ladefoged, Peter. 2003. *Phonetic Data Analysis: An introduction to fieldwork and instrumental techniques*. Malden, MA and Oxford: Blackwell.

Langacker, Ronald. 1987. *Foundations of Cognitive Grammar: Theoretical prerequisites*. Stanford: Stanford University Press.

Langacker, Ronald. 1999. *Grammar and Conceptualization*. Berlin: Mouton de Gruyter.

Lass, Roger. 1984. *Phonology: An introduction to basic concepts*. Cambridge, UK: Cambridge University Press.

Law of the People's Republic of China on the Standard Spoken and Written Chinese Language. People's Republic of China Database of Laws and Regulations. Published October 31, 2000. Accessed January 31, 2017. http://www.npc.gov.cn/englishnpc/Law/2007-12/11/content_1383540.htm.

Li, Charles and Sandra A. Thompson. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.

Lin, Yen-Hwei. 2007. *The Sounds of Chinese*. Cambridge, UK: Cambridge University Press.

Lin, Yen-Hwei. 2011. "Loanword adaptation in Standard Mandarin and phonological theory," Current Issues in Chinese Linguistics, edited by Yun Xiao, Liang Tao & Hooi Ling Soh, 426–451. Cambridge: Cambridge Scholars Publishing.

Lindqvist, Cecilia. 1991. *China: Empire of Living Symbols*. New York: Addison-Wesley Publishing Company, Inc.

Lü, Shuxiang 吕叔湘. 1990. 吕叔湘文集(2): 汉语语法论文集 (Lü Shuxiang Wen Ji (2): Hanyu Yufa Lunwen Ji: Collection of Research Articles by Lü Shuxiang (2): Articles on Chinese Grammar). 北京, 商务印书馆 (Beijing: Commercial Press).

Lyovin, Anatole V., Brett Kessler, and William R. Leben. 2017. *An Introduction to the Languages of the World* (2nd edition). Oxford, UK: Oxford University Press.

Newell, Allen. 1990. *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Norman, Jerry. 1988. *Chinese* (Cambridge Linguistic Surveys). Cambridge, UK: Cambridge University Press.

Pierrehumbert, Janet B. 2001. "Exemplar dynamics: Word frequency, lenition and contrast." In *Frequency effects and the emergence of linguistic structure*, ed. by Joan Bybee and Paul Hopper, 137–158. Amsterdam: John Benjamins.

Reetz, Henning and Allard Jongman. 2011. *Phonetics: Transcription, Production, Acoustics, and Perception*. Wiley-Blackwell.

Su, Danjie. 2016. "Grammar emerges through reuse and modification of prior utterances." *Discourse Studies* 18 (3): 330–353. https://doi.org/10.1177/1461445616634551

Tannen, Deborah. 1987. "Repetition in conversation as spontaneous formulaicity." *Text* 7 (3): 215–243.

Tao, Hongyin. 2015. "Profiling the Mandarin Spoken Vocabulary Based on Corpora." In *The Oxford handbook of Chinese linguistics*, ed. by William S-Y. Wang, and Chaofen Sun, 336–347. Oxford: Oxford University Press.

Tao, Liang. 2002. "Phono-syntactic conspiracy and beyond: Grammaticalization in spoken Beijing Mandarin." In *New Reflections of Grammaticalization*, ed. by Ilse Wischer, and Gabriele Dieward, 283–299. Amsterdam: John Benjamins.

Tao, Liang. 2005. "The Importance of Discourse Analysis for Linguistic Theory: A Mandarin Chinese Illustration." In *Linguistic Diversity and Language Theories*, ed. by Zygmunt Frajzyngier, Adam Hodges, and David Rood, 285–317. Amsterdam: John Benjamins.

Tao, Liang. 2006. "Classifier Loss and Frozen Tone in Spoken Beijing Mandarin: The *yi+ge* phono-syntactic Conspiracy." *Linguistics* 44: 91–133.

Tao, Liang. 2009. "Syntactic tone and discourse processing in Beijing Mandarin: A case study." *Journal of Chinese Linguistics* 37 (2): 258–296.

Tao, Liang. 2018. "Self-repair in Mandarin Chinese: The multimodality of conversation." In *Multimodality in Chinese Interaction*, ed. by Xiaoting Li and Tsuyoshi Ono, 255–298. Berlin: De Gruyter Mouton.

Tao, Liang, Barbara Fox, and Jule Gomez de Garcia. 1999. "Tone-choice repair in Conversational Mandarin Chinese". In *Cognition and Function in Language*, ed. by Barbara Fox, Dan Jurafsky, and Laura Michaelis, 268–281. Stanford: CSLI Publications.

Thompson, Sandra A., and Paul Hopper. 2001. "Transitivity, clause structure, and argument structure: Evidence from conversation." In *Frequency and the Emergence of Linguistic Structure*, ed. by Joan Bybee, and Paul J. Hopper, 27–60. Amsterdam: John Benjamins.

Wang, Li. 1958. *Hayu Shigao* (A Draft History of the Chinese Language), Vol. 2. Beijing: Science Press.

Wiedenhof, Jereon. 1995. *Meaning and Syntax in Spoken Mandarin*. Leiden, the Netherlands: Research School CNWS.

Wieger, Léon S. J. 1965 (1927). *Chinese Character, their Origin, Etymology, History, Classification and Signification: A Thorough Study from Chinese Documents*. 2nd ed., English and revision according to the 4th French ed. New York: Paragon Book Reprint Corp.

Xinhua News Agency. 2011. Today in History: On February 6th, the Chinese State Department issued the document '*Directions on promoting Putonghua*'. Central People's Government of the People's Republic of China. Accessed on December 12, 2016. http://www.gov.cn/lssdjt/content_1799191.htm

Zhan, Weidong, Rui Guo, and Yirong Chen. 2003. 北京大学汉语语言学研究中心语料库 The CCL Corpus of Chinese Texts: 700 million Chinese Characters, the 11th Century B.C. – present, Available online at the website of Center for Chinese Linguistics (CCL) of Peking University. http://ccl.pku.edu.cn:8080/ccl_corpus

Zhang, Qing. 2005. "A Chinese yuppie in Beijing: Phonological variation and the construction of a new professional identity." *Language in Society* 34 (3): 431–466.

Zhang, Wei 张巍. *陆贞传奇* (Legend of Lu, Zhen). Accessed March 8, 2017. http://www.99lib.net/book/4225/index.htm

Zipf, George Kingsley. 1968 [1935]. *The Psycho-Biology of Language*. Cambridge, MA: MIT Press.

Appendix

List of transcription notations and glossary abbreviations used in this study (Following Jefferson, 2004; Li and Thompson 1981)

()	Pause time in seconds
::	Sound stretch
'x	Missing or unclearly pronounced sound, often the onset of a syllable.
x'	The onset has turned into a voiced continuant /r/ or /I/ (Chao 1968: 37)
Xx21	Tone sandhi change of the 3rd tone 214
1SG	First person singular (including: 2nd-sg: second person singular, etc.)
ASP	Aspectual particle or oblique marker (no detailed semantic classification)
CL	Classifier
INT	Interjection
NEG	Negative particle
Q	Question particle
x-	Sound cut-off
XX-One digit	Tone cut-off – the tone is not fully pronounced (e.g., *jin5-*)

Contextual variations of internal and external modifications in Chinese requests

Effects of power and imposition

Shuai Li

Georgia State University

This study investigates contextual variations in mitigation production (consisting of internal and external modifications) in Chinese request-making (i.e., what native Chinese speakers consider appropriate to say in hypothetical scenarios). The participants were 22 native Chinese speakers recruited from a university in China. They completed a 20-item Oral Discourse Completion Test (ODCT) tapping two contextual variables: power and imposition. The results show that: (1) both power and imposition exerted significant influence on the frequency of producing internal and external modifications, (2) the various internal and external modifiers were differentially associated with the two contextual variables, and (3) the preferred sequential organization of external modifications differed according to context types.

Keywords: request, modifications, Chinese, contextual variation, power, imposition, sequential organization

1. Introduction

One important consideration in daily speech communication is to convey politeness through the use of linguistic devices that can modify the tone of speech. Such linguistic devices are called mitigations (Caffi 1999, 2007; Fraser 2010; Thaler 2012). In spite of their importance in daily communication, when and how mitigations are used in Chinese remain under-researched topics within the field of linguistic politeness. Specifically in discussing the speech act of request-making, the mitigating effects are realized through the use of internal and external modifications. Yet, only a handful of studies (e.g., Dong 2008; Jia & Huang 2008; Kirkpatrick 1991; Rue & Zhang 2008; Zhang 1995; Zhu 2017) have empirically

investigated internal and external modifications involved in Chinese requests, and few have examined whether and how key contextual factors such as those outlined in Brown and Levinson's (1987) politeness theory affect the use of request modifications (e.g., Hong 2002; Rue & Zhang 2008). Investigating contextual variations of request modifications is a meaningful topic because it is the appropriate match between linguistic forms and contexts that give rise to linguistic politeness. This study aims to contribute to the development of research on Chinese mitigations in general by examining whether and how internal and external modifications used in request-making are influenced by two contextual variables (i.e., power and imposition).

2.   Literature review

2.1   Request modifications and contextual variations

According to Blum-Kulka, House & Kasper (1989), a request sequence can be segmented into three main components: alerter(s), head act(s), and external modification(s), as shown in the following example:

(1)   陈　老师，　　我 没有　　收到　　您的 邮件。您　看
　　　Chén lǎoshī,　　wǒ méiyǒu　shōudào nín de yóujiàn. Nín kàn
　　　Chen Professor, I   have not receive   your e-mail. You see
　　　您　能　再　　给 我
　　　nín néng zài　gěi wǒ
　　　you can   again to me
　　　发　一下　邮件　吗？　　谢谢！
　　　fā　yīxià　yóujiàn ma?　　Xièxie!
　　　send a little e-mail PARTICL? Thanks
　　　'Professor Chen, I have not received your e-mail. Do you think you could resend it to me? Thanks!'

In the above example, 陈老师 (*Chén lǎoshī*, Professor Chen) is the alerter, which functions to catch the interlocutor's attention. 您看您能再给我发一下邮件吗 (*Nín kàn nín néng zài gěi wǒ fā yīxià yóujiàn ma*, Do you think you could resend the e-mail to me) is the head act, because it most explicitly realizes the intended request, and also because it is independent of other components of the request sequence. Outside the boundary of the head act, 我没有收到您的邮件 (*Wǒ méiyǒu shōudào nín de yóujiàn*, I have not received your e-mail) (i.e., providing a reason for the request) and 谢谢 (i.e., *xièxiè*, thanks) are both external modifications that function to moderate the tone of the request. In addition to the external

modifications, the use of 您 (*nín*, an honorific pronoun), 您看 (*nín kàn*, a consultative term), and 一下 (*yīxià*, an understater) within the boundary of the head act also serves the purpose of mitigating the tone of speech. Linguistic devices that provide mitigating effects within a request head act are known as internal modifications.

As the above examples can show, internal modifications in Chinese are typically lexical and phrasal structures with a mitigating function. External modifications, on the other hand, are semantic formulae (i.e., meaning-based strategies) that are not associated with fixed linguistic forms. Both internal and external modifications contain sub-categories. For example, internal modifications include sub-categories such as understater (e.g., 一下, *yīxià*, a little bit), downtoner (e.g., sentence-final particles such as 吧 *ba*, 嘛 *ma*, and 呢 *ne*), consultative terms (e.g., 你看, *nǐ kàn*, in your view), and politeness markers (e.g., 请, *qǐng*, please), etc. External modifications include sub-categories such as grounders (e.g., providing justification/reason for a request), thanking (i.e., thanking one's interlocutor for performing a request), and sweeteners (i.e., flattering one's interlocutor), etc.

Whether and how internal and external modifications are used in making requests is contingent upon the influences of contextual factors. Request modifications are specific types of politeness strategies, which, according to Brown and Levinson (1987), are employed to address people's face needs. Face refers to one's public self-image consisting of positive and negative aspects (Brown & Levinson 1987): the positive face means one's desire to be liked by other members of a society, and the negative face refers to one's desire to be free of imposition. In asking people to do things, a request maker threatens his/her interlocutor's negative face. The degree of the face being threatened by an act (including requests) can be assessed based on three key contextual factors: Power (P), Social distance (D), and Rank of imposition (R). While the three contextual factors can determine together the amount of face being threatened, each factor can also exert its own influence independently. Specifically, power refers to the status difference between interlocutors; social distance, on the other hand, reflects the frequency of interaction between interlocutors. Finally, rank of imposition indicates the extent to which an act affects one's desire for public approval (i.e., positive face) or of self-determination (i.e., negative face). Generally, a greater amount of face being threatened entails more efforts to address face needs. In terms of requests, this means more frequent use of modifications and/or employing various types of modifications. For example, asking for a small favor from one's good friends (e.g., borrowing a pen for temporary use) may not involve any modification, whereas asking for a big favor from one's professor (e.g., extending the deadline of a term paper) will likely lead to the use of multiple internal (e.g., using language-specific honorifics such as 您 *nín* in Chinese) and external modifiers (e.g., apologizing, and explaining the

reason/justification for the extension). Based on the above understanding, contextual variation of request modifications in this study is understood as the variation in frequency and type of internal and external modifications as a result of the influence of contextual variables.

2.2   Internal and external modifications in Chinese requests

To date, research on internal and external modifications in Chinese requests remains limited in that most studies have focused on describing the linguistic forms and/or strategies that can serve mitigating functions (e.g., Dong 2008; Jia & Huang 2008; Jiang 2012; Kirkpatrick 1991; Lu & Wu 2005; Zhan 1992; Zhang 1995); very few studies have investigated contextual variations of using modifications, that is, whether and how the distribution of modifications is influenced by contextual variables such as power and imposition (Hong 1996; Rue & Zhang 2008; Zhang & Wang 1997). In the following, existing studies on internal and external modifications will be reviewed first, followed by a review of the literature regarding the effects of contextual variables on the use of internal and external modifications.

With respect to research on internal modifications in Chinese requests, Zhan's (1992) study was pioneering. While her book mainly focuses on demonstrating how Brown & Levinson's (1987) various politeness strategies (e.g., positive politeness strategies, negative politeness strategies) are realized in Chinese by citing examples selected from modern Chinese novels, Zhan identified several linguistic forms as tone softeners, including verb reduplication (e.g., 开开门, *kāikāi mén*, open the door), 一下 (*yīxià*, a little bit), and sentence-final particles (e.g., 啊 *a*). Zhan did not, however, provide a categorization of internal modifications, nor did she compile a list of internal modifiers.

As another significant step, Zhang (1995) adopted the framework set by the Cross-Cultural Speech Act Realization Project (Blum-Kulka, House & Kasper 1989) to compile a list of Chinese internal modifications with sub-categories and ample examples. She collected data through a Written Discourse Completion Test (WDCT) with 12 request-making scenarios that varied in contextual variables such as power, social distance, and imposition (e.g., a policeman asking a driver to move a car; borrowing a large sum of cash from one's friend). Thirty native Chinese speakers from China completed the WDCT. Based on the data collected, Zhang proposed a coding system for internal modifications, including syntactic downgraders (e.g., conditional clause such as 要是 …, *yàoshì* …, If…) and various lexical downgraders (e.g., downtoners like 吧, *ba*). Zhang did not, however, investigate contextual variations in the use of internal modifications.

Recent research on Chinese internal modifications appears to focus on delineating the pragmatic functions of specific linguistic forms. For example, Lu & Wu

(2005) compared the functions of three mitigators in Chinese requests, namely, *verb, verb* +一下 (verb + *yīxià*), and *verb reduplication*. Relying on several sample request utterances, the researchers argued that these three forms conveyed descending degrees of imposition in terms of how much freedom a requestee enjoys in not complying with a request. In another study, Jiang (2012) discussed the multiple pragmatic functions of the construction *verb* +一下 (verb + *yīxià*), including its role in request making. Jiang's data consisted of a selection of utterances containing targeted construction gleaned from two Chinese TV series. Her analyses showed, among other things, the mitigating effect of 一下 (*yīxià*, a little bit) in Chinese requests.

As the aforementioned review of research on internal modifications in Chinese requests can show, researchers have generally focused on identifying and compiling linguistic forms that can be used as internal modifications (e.g., Zhan 1992; Zhang 1995) or on detailing the pragmatic function of specific internal modifiers (Lu & Wu 2005; Jiang 2012). While it is important to know which forms can function as internal modifiers in Chinese requests, it is equally desirable to examine the contextual distribution of internal modifications, because findings in this regard can reveal the conditions under which one may or may not use internal modifications in order to be appropriate and polite. Empirical research in this regard is clearly needed.

Turning to research on external modifications in Chinese requests, while compiling a list of possible external modifiers has received attention in the field (e.g., Zhan 1992; Zhang 1995), the main scholarly focus has been on investigating the preferred sequential organizations of external modifications in relation to request head acts. In theory, external modifiers can occur before, after, and both before and after request head acts. Several researchers have reported that native Chinese speakers prefer to place external modifications prior to request head acts (e.g., Dong 2008; Kirkpatrick 1991; Jia & Huang 2008; Zhu 2017). For example, Kirkpatrick analyzed 40 letters written by native Chinese speakers to a radio station to request information, service, and products. Thirty-seven of the 40 letters were organized in such a way that a request(s) was put forward following facework (e.g., praising the service of the radio station) and grounder (i.e., justification/ reason for a request). The three letters that neglected facework and placed requests before grounder(s) were generally considered impolite by a group of native speakers, even though the requests involved in the three letters were less imposing in terms of content. Also drawing on authentic data (i.e., e-mails with requests), Zhu (2017) examined, among other things, the cross-cultural difference in rhetorical structure of request-making between Chinese and British graduate students. He found that the inductive strategy (i.e., presenting background information and/ or grounders before request head acts) dominated the Chinese students' e-mail

requests, accounting for 95.40% of the data; by contrast, the inductive and deductive (i.e., presenting background information and/or grounders after request head acts) strategies were more or less equally preferred by the British students. Because facework strategies, providing background information, and grounders are sub-categories of external modifications within the framework mentioned earlier (Blum-Kulka et al. 1989), findings by Kirkpatrick (1991) and Zhu (2017) suggest that, in Chinese requests that are considered to be appropriate, external modifications typically occur before head acts.

Several studies seem to lend further support to the aforementioned conclusions (e.g., Dong 2008; Hong 2002; Jia & Huang 2008). For example, Dong (2008) designed a Written Discourse Completion Test (WDCT) with 14 request-making situations involving various roles (e.g., friends, classmates, professors, and service providers). The WDCT was completed by 25 native Chinese speakers. Out of the 220 request utterances that included external modifications, 133 (or 60.45%) contained external modifications before head acts. In contrast, 56 (or 25.45%) request utterances showed external modifications after head acts, and only 31 (or 14.09%) exhibited external modifications both before and after head acts. Similar patterns were found for the sequential organization of grounders (i.e., the most frequently occurring sub-category of external modification): out of the 164 grounders found in the data set, 122 (or 74.39%) came before head acts, and 42 (or 26.61%) were after head acts. These results corroborated Kirkpatrick's findings in showing that placing external modifications (including grounders) before head acts is a preferred discourse structure of Chinese requests.

As a further support to the above observations from a cross-cultural perspective, Jia & Huang (2008) compared English requests made by two groups of participants: native Chinese speakers and native English speakers who were professionals working in the academia. Qualitative analyses of field notes and e-mail exchanges in professional contexts revealed that native Chinese speakers typically placed external modifiers before request head acts, a discourse structure that contrasted sharply with the head act-first structure preferred by native English speakers (as the results of the study showed). These results showed that the non-native-like discourse structure exhibited in English requests produced by native Chinese speakers was due to the influence of the Chinese way of sequencing requests.

While the aforementioned studies on external modifications all point to a tempting generalization that the preferred sequential organization of Chinese requests is to place external modifications before head acts, a closer review of these studies indicates more refined works are needed to examine the validity of such a generalization. In particular, because the studies discussed above (e.g., Dong 2008; Kirkpatrick 1991; Jia & Huang 2008; Zhu 2017) lumped all request-making situations together for analysis, they tended to overlook potential contextual variations

in preferred sequential organization(s) of Chinese requests (e.g., see Hong's study reviewed below).

In fact, there has been empirical evidence suggesting that native Chinese speakers' preferred structure of organizing requests may be different according to certain contextual variables. In this regard, Hong's (2002) study is the only one that the researcher knows of that explored this issue. Hong designed a Written Discourse Completion Test (WDCT) with three request-making scenarios that varied along the power continuum. One scenario involved a professor asking students to turn in a term paper by a certain date (higher power), one scenario was about borrowing lecture notes from one's classmates (equal power), and one scenario involved a patient asking a doctor to refill a prescription (lower power). The WDCT was administered to 46 native Chinese speakers of various dialects. Of particular relevance to the present study are the findings regarding the sequential organization of grounders (a sub-category of external modification) in relation to request head acts. While the higher power scenarios did not generate grounders, the other two scenarios were associated with different preferred sequential structures of grounders. In the equal power scenario, there were more post-grounders (i.e., providing reasons/justifications after request head acts, 40%) than pre-grounders (i.e., reasons/justifications followed by request head acts, 28%). In the lower power scenarios, however, pre-grounders (i.e., 58%) outnumbered post-grounders (i.e., 24%). Hong's findings suggest that the preferred sequential organization of Chinese requests may be contingent upon certain contextual variables such as power differences between interlocutors.

Hong's findings need to be complemented with additional empirical efforts to gain a more fine-grained understanding of contextual variations in the preference for sequential organizations of external modifications (including grounders). For example, it is legitimate to ask whether and how other contextual variables in addition to power would influence the choice of preferred sequential structure(s) of external modifications. To answer this question would involve a study that incorporates multiple contextual variables into its research design. Moreover, because all previously discussed studies have been either qualitative in nature or have relied only upon descriptive statistics (e.g., raw frequency, percentage), it is not possible to tell whether the observed differences bear any statistical significance. Studies with appropriate inferential statistical procedures can help advance this line of research. Finally, although grounder is typically the most frequently produced sub-category of external modifications in Chinese requests (Lee-Wong 2000; Zhang 1995; Zhu 2017) and therefore merits focused research attention (like in Hong's study), one needs to be cautious in overgeneralizing Hong's findings because her study only concentrated on grounders, rather than on external modifications as a whole category (see Dong 2008; Kirkpatrick 1991; Jia & Huang 2008; Zhu 2017).

In other words, it remains an empirical question whether the observed effects of power on preferred sequential organization of Chinese requests are restricted to grounders only, or are applicable to external modifications as a whole category. Hence, both grounders and external modifications (as a whole category) should be examined in future research to answer this empirical question.

The review of the literature on internal and external modifications in Chinese requests shows scant research on contextual variations in using internal and external modifications. Studies investigating the role of various contextual variables in influencing the production of both categories of modifications are needed. Methodologically, studies with more sophisticated design (e.g., targeting more than one contextual variable) using inferential statistics, as well as studies examining both grounders and external modifications as a whole category, are needed to better evaluate the generalizability of existing research findings. This study aims to address these gaps in the literature. It proposes to examine whether and how oral productions of internal and external modifications are affected by two contextual variables (i.e., power and imposition) in terms of frequency and preferred sequential structure of Chinese requests. The two research questions are:

RQ1:   How does the frequency of producing internal and external modifications vary according to power and imposition?

RQ2:   How does the preferred sequential organization of external modifications and grounders vary according to power and imposition?

3.   Method

3.1   Participants

Participants were 22 Chinese undergraduate students enrolled in a university in Beijing, China. There were 10 males and 12 females. They aged between 19 and 24 years with a mean of 21.75 years ($SD = 1.07$). The students majored in various academic fields, such as English, computer science, information system management, finance, Spanish, and Chinese language and literature, etc.

3.2   Instrument

A 20-item Oral Discourse Completion Test (ODCT) was developed to collect the participants' oral productions of requests. The ODCT scenarios specifically tapped two contextual variables discussed in Brown & Levinson (1987), namely, power (P) and imposition (I). Power (P) was operationalized in terms of role relations

(Spencer-Oatey 2000) and reflected status difference between two interlocutors. Power was a targeted contextual variable. Given the fact that the participants of this study were all university students living on campus, they should presumably have very frequent contact with two groups of people: friends/classmates and professors. As a result, the power variable of this study included two levels: equal power status (P1, interactions between two friends, 10 situations) and hearer higher power status (P2, student-professor interactions, 10 situations).

Imposition was the other targeted contextual variable in this study, because it was identified as a contextual variable that exerted great influence on Chinese requests, particularly on request strategies (Lee-Wong 2000; Yeung 1997), but few empirical results are available regarding whether it has a similar influence on request modifications. Imposition in this study was operationalized as the psychological difficulty of making requests (Takahashi 1998) with two levels: low imposition requests (i.e., small favors) and high imposition requests (i.e., big favors). Following Spencer-Oatey (1993), the researcher developed a metapragmatic assessment questionnaire as a pilot study to verify the two levels of imposition. The questionnaire included 47 request-making scenarios. Each scenario was followed by a six-point scale for measuring the psychological difficulty involved in putting forward the request, with the score of one representing the least difficult and the score of six the most difficult. The questionnaire was administered to 15 native Chinese speakers who did not participate in the main study. To qualify for a low imposition request scenario (R1), at least 80% of the native speakers needed to choose 1, 2, or 3 on the six-point scale; to qualify for a high-imposition scenario (R2), at least 80% of the native speakers needed to choose 4, 5, or 6 on the same scale. Twenty scenarios met these criteria. There were 10 low-imposition and 10 high-imposition scenarios.

As previously mentioned the ODCT included 20 items (i.e., scenarios) and tapped two contextual variables: power and imposition. Because power and imposition each included two levels, there were four context types, namely, equal power status and low imposition (P1R1), equal power status and high imposition (P1R2), hearer higher power status and low imposition (P2R1), and hearer higher power status and high imposition (P2R2). The 20 ODCT items were evenly divided into the four context types (See Appendix A for the 20 scenarios).

Finally, social distance, the third contextual variable included in Brown & Levinson's politeness theory, was included as a controlled variable: participants were told explicitly that the persons involved in each request scenario knew each other very well.

3.3   Procedures

Participants completed the ODCT individually with the researcher in a quiet room on campus. In completing each item, they listened to a scenario description (in Chinese) played by a tape recorder while reading the same description printed in the questionnaire. Afterwards, they responded orally with what they would say in that scenario. Their oral productions were recorded for analysis. Participants generally took 15–20 minutes to complete the ODCT.

3.4   Data analysis

Participants' oral productions, a total of 440 request utterances (20 utterances per person x 22 participants), were transcribed for analysis. To answer the First Research Question (RQ1), each request utterance was coded for internal and external modifications. The coding scheme was developed based on the existing literature (e.g., Lee-Wong 2000; Li 2014; Rue & Zhang 2008; Wen 2014; Zhang 1995) to fit the context of this study. Two separate 2 (power) x 2 (imposition) repeated measures ANOVAs were performed to examine whether and how the frequency of producing internal and external modifications was influenced by the two contextual variables. The alpha level was set as .05.

To answer the Second Research Question (RQ2), external modifications and grounders (a sub-category of external modification) were first categorized in terms of sequential organizations: before (hereafter "prior"), after (hereafter "subsequent"), and both before and after head acts (hereafter "both"). These three types were compared in terms of frequency of occurrence in each of the four context types mentioned earlier (i.e., P1R1, P1R2, P2R1, and P2R2). Due to the violation of the normality assumption required for parametric statistical procedures, non-parametric procedures were used to answer RQ2: *Friedman* tests were performed first to examine whether there was any significant overall difference among the three types of sequential organizations. In the case of statistically significant results, *Wilcoxon* tests were conducted for follow-up pairwise comparisons. The alpha level was set as .05 for the *Friedman* tests and .017 for the *Wilcoxon* tests (for three comparisons).

4.   Results

Research Question One (RQ1) asks about contextual variations in producing internal and external modifications. Table 1 displays the raw frequencies of specific internal modifiers that occurred at least once in the dataset. Some interesting

patterns emerged from the frequency distributions. One internal modifier, the understater 一下 (*yīxià*, a little bit) appeared to be influenced by both power and imposition: it was produced 87 times (46 + 41) in P1 scenarios (equal power scenarios) compared to 50 times (32 + 18) in P2 scenarios (hearer higher power scenarios). On the other hand, this understater occurred 78 times (46 + 32) in R1 (low imposition) scenarios as opposed to 59 times (41 + 18) in R2 (high imposition) scenarios. These results indicate that 一下 (*yīxià*, a little bit) as an understater is more likely to be associated with low imposition and/or equal power request scenarios.

Several other internal modifiers appeared to be mainly influenced by either power or imposition. Concerning the effects of the power variable, the following internal modifiers were produced substantially more frequently in P1 (equal power) scenarios than in P2 (hearer higher power) scenarios: downtoners 吧 *ba*, 呗 *bei*, and *verb reduplication* forms such as 看看 (*kànkan*, take a look); in contrast, politeness markers 麻烦 (*máfan*, to trouble) and 请 (*qǐng*, please), and the honorific pronoun 您 (*nín*) were used exclusively in P2 scenarios. Regarding the influence of the imposition variable, several internal modifiers were mostly produced in R2 (high imposition) scenarios than in R1 (low imposition) scenarios, including the downtoners 先 (*xiān*, first) and 顺便 (*shùnbiàn*, conveniently), the understater 一点儿 (*yī diǎr*), the consultative terms 您/你看 (*nín /nǐ kàn*, in your view), and conditional clauses such as 要是/如果… (*yàoshi /rúguǒ*, if…). The remaining internal modifiers not discussed here did not seem to be affected by either contextual variable, and this was primarily because of their low frequency of production.

Table 1. Raw frequency of internal modifiers across context types

Sub-types of internal modifications	P1R1	P1R2	P2R1	P2R2
Verb reduplication	12	6	1	1
Politeness marker				
麻烦 *máfan* to trouble	0	0	2	5
请 *qǐng* please	0	0	1	2
请问 *qǐngwèn* may I ask	0	0	0	1
Downtoner				
可能 *kěnéng* possible/probable	0	0	0	1
顺便 *shùnbiàn* conveniently	0	1	0	5
稍微 *shāowēi* a little	0	0	0	2
先 *xiān* first	1	6	0	8
吧 *ba* sentence final particle	17	15	2	1
呢 *ne* sentence final particle	0	0	1	4

(*continued*)

Table 1. (*continued*)

Sub-types of internal modifications	P1R1	P1R2	P2R1	P2R2
呀 *ya* sentence final particle	0	2	1	3
呗 *bei* sentence final particle	5	1	0	0
啊 *a* sentence final particle	1	0	2	5
Appealer				
行不行/行吗 *xíng bu xíng /xíng ma* OK?	5	9	3	5
好不好/好吗 *hǎo bu hǎo /hǎo ma* OK?	3	2	0	0
可以不可以/可以吗 *kěyǐ bù shì ba /kěyǐ ma* May I?	1	4	4	2
怎么样 *zěn me yang* How is that?	0	1	0	0
成吗 *chéng ma* OK?	0	2	1	0
Subjectiviser				
(我)想 *(wǒ) xiǎng* I think	0	0	1	3
(我)觉得 *(wǒ) juéde* I feel	0	1	1	0
Understater				
一下 *yīxià* a little bit	46	41	32	18
(一)点(儿) *(yī) diǎn (ér)* a little bit	2	14	0	15
(一)些 *(yī)xiē* some	2	3	0	4
Other	1	2	0	0
Honorific				
您 *nín* honorific pronoun	0	0	65	53
Consultative term				
您/你看 *nín /nǐ kàn* in your view	0	4	1	14
你是不是 *nǐ shì bù shì* how about…	0	0	1	1
Hesitation marker				
(我)不知道 *(wǒ)bù zhīdào* (I) don't know	0	0	0	4
那个 *nà ge* that	1	2	4	7
就是 *jiùshì* just	0	1	1	4
Conditional clause	1	8	0	7

Note. P1: equal status; P2: hearer higher status; R1: low imposition; R2: high imposition

To complement the aforementioned results, Table 2 displays the means and standard deviations of the frequency of occurrence of internal modifiers per request scenario. A repeated measures ANOVA revealed a significant main effect of power on the frequency of producing internal modifications, $F(1, 21) = 13.80$, $p = .001$, $\eta_p^2 = .40$, and a significant main effect of imposition, $F(1, 21) = 17.21$, $p < .001$,

$\eta_p^2 = .45$. The power x imposition interaction effect, however, was not significant, $F(1, 21) = 2.33$, $p = .14$. Due to a lack of significant interaction effect, two separate paired samples t tests were performed to further examine the observed main effects. The results showed that significantly more internal modifications were produced in hearer higher power status (P2) scenarios than in equal power status (P1) scenarios, $t(21) = -3.72$, $p = .001$. Moreover, significantly more internal modifications were used in high imposition (R1) scenarios than in low imposition (R1) scenarios, $t(21) = -4.15$, $p < .001$. The main effects of power and imposition on the frequency of producing internal modifications are summarized in Table 3.

Table 2. Frequency of producing internal and external modifications per scenario

	P1R1		P1R2		P2R1		P2R2	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Internal	0.89	0.46	1.14	0.50	1.12	0.46	1.59	0.70
External	0.47	0.43	1.68	0.93	1.02	0.54	2.15	0.82

Table 3. Main effects of power and imposition on frequency of producing internal and external modifications

	Main effect of power	Main effect of imposition
Internal	P2 > P1	R2 > R1
External	P2 > P1	R2 > R1

Turning to the production of external modifications, Table 4 displays the raw frequencies of the external modifiers that were produced at least once in the data set. Grounder (i.e., providing reasons/justifications of requests) was the most frequently produced external modifier, and its occurrence appeared to be affected by both power and imposition: it was used 97 times in P1 (equal power) scenarios in comparison with 217 times in P2 (hearer higher power) scenarios; moreover, it occurred 115 times in R1 (low imposition) scenarios as opposed to 199 times in R2 (high imposition) scenarios. Most external modifiers were affected by either power or imposition. For example, *promise of reward* and *direct appeal* were exclusively used in P1 scenarios, whereas *thanking* was mainly associated with P2 scenarios. Concerning the role of imposition, eight external modifiers were mostly or even exclusively associated with R2 (high imposition) scenarios. These were *preparator, acknowledging difficulty in carrying out requests, getting a pre-commitment, cost minimizer, promise, apologizing, double checking compliance*, and *offering opting out*. The remaining external modifiers occurred infrequently, which makes it difficult to investigate their contextual variations.

Table 4. Raw frequency of external modifiers across context types

Sub-types of external modifications	P1 R1	P1 R2	P2 R1	P2 R2
Acknowledging a problem e.g., 我知道这是学期论文上交的最后期限… *Wǒ zhīdào zhè shì xuéqī lùnwén shàngjiāo de zuìhòu qīxiàn…* I know this is the deadline for turning in the term paper…	0	0	0	1
Admitting difficulty in performing requests e.g., 我知道你很忙，然后那个地方比较远。*Wǒ zhīdào nǐ hěn máng, ránhòu nà ge dìfang bǐjiào yuǎn.* I know that you are very busy, and that place is a bit far…	0	10	0	7
Apologizing e.g., 真是对不住了啊。*Zhēnshì duìbuzhù le ā.* I am really sorry…	1	8	5	18
Cost minimizer e.g., 打车去吧，钱我报销。*Dǎ chē qù ba, qián wǒ bàoxiāo.* You can get a Taxi, and I will reimburse you.	2	14	0	12
Direct appeal e.g., 拜托了。*Bàituō le.* Please.	1	5	0	0
Disarmer e.g.,(可能数目有点儿多)，你不要害怕啊。*(Kěnéng shùmù yǒu diǎnr duō), nǐ bùyào hàipà ā.* (It might be a big amount of money), don't be scared.	0	1	0	0
Double checking compliance e.g., 没问题吧？*Méiwèntí ba?* No problem, right?	0	5	0	5
Getting a pre-commitment e.g., 能不能帮我一下忙？*Néngbùnéng bāng wǒ yīxià máng?* Can you help me?	0	5	0	1
Grounder e.g., 我实在是没有时间。*Wǒ shízài shì méiyǒu shíjiān.* I really don't have time.	33	64	82	135
Moralizing statement e.g., 哥们之间帮一下忙，没问题吧？*Gēmen zhījiān bāng yīxià máng, méiwèntí ba?* Brothers help each other. No problem, right?	0	1	0	0
Offering alternative e.g., 或者我再给您一个别的邮箱。*Huòzhě wǒ zài gěinínyī gèbié de yóuxiāng.* Or I can give you a different e-mail address.	0	0	1	0
Offering assistance e.g., 这是我的 e-mail 地址。*Zhè shì wǒ de e-mail dìzhǐ.* This is my e-mail address.	0	0	1	2
Offering opting out e.g., 不过如果你忙的话就算了。*Búguò rúguǒ nǐ máng dehuà jiùsuàn le.* But if you are busy, don't worry about it.	0	4	0	4

Table 4. (*continued*)

Sub-types of external modifications	P1 R1	P1 R2	P2 R1	P2 R2
Preparator e.g., 有件事要拜托你。 *Yǒu jiàn shì yào bàituō nǐ.* I have something to ask you for help.	8	24	5	17
Promise e.g., 我明天肯定还你。 *Wǒ míngtiān kěndìng huán nǐ.* I will make sure to return it back to you tomorrow.	1	22	1	13
Promise of reward e.g., 回来请你吃饭。 *Huílai qǐng nǐ chīfàn.* I will treat you a meal after you come back.	1	5	0	0
Self-criticism e.g., 我那个笔记特别不全。 *Wǒ nà ge bǐjì tèbié bù quán.* My notes are really not complete.	0	2	1	0
Sweetener e.g., 您这 ppt 做得太精彩了。 *Nín zhè ppt zuò de tài jīngcǎi le.* Your PowerPoint slides are so wonderful.	0	8	3	5
Thanking e.g., 谢谢你啦。 *Xièxie nǐ la.* Thank you!	5	7	13	17

To supplement the aforementioned observations, Table 2 displays the means and standard deviations of the frequency of occurrence of external modifiers per request scenario. Again, there was a significant main effect of power, $F(1, 21) = 37.68$, $p < .001$, $\eta_p^2 = .64$, a significant main effect of imposition, $F(1, 21) = 81.23$, $p < .001$, $\eta_p^2 = .80$, but the power x imposition interaction effect did not reach a significant level, $F(1, 21) = 0.25$, $p = .62$. Since there was no significant interaction effect, the researcher performed two separate paired samples t tests to examine the effects of power and imposition on the production of external modifications. Significantly more external modifications were used in P2 (hearer higher power) scenarios than in P1 (equal power) scenarios $t(21) = −6.14$, $p < .001$. In addition, more external modifications were produced in R2 (high imposition) scenarios than in R1 (low imposition) scenarios, $t(21) = −9.01$, $p < .001$. The main effects of power and imposition on the frequency of producing internal modifications are summarized in Table 3.

Research Question Two (RQ2) asks whether there is any contextual variation in the preferred sequence organization(s) of external modifications and grounders. Table 5 presents the descriptive statistics showing the frequency of three types of sequential organizations of external modifications: "prior", "subsequent", and "both". In P1R1 scenarios, a *Friedman* test revealed a significant overall difference between the three types, $\chi^2 (2, n = 22) = 12.67$, $p = .001$. Follow-up *Wilcoxon*

tests showed a significant difference between "subsequent" and "both" ($Z = -2.54$, $p = .004$), but not between "prior" and "subsequent" ($Z = -1.27$, $p = .23$) or between "prior" and "both" ($Z = -2.14$, $p = .04$). In P1R2 scenarios, there was no significant difference between the three types, χ^2 (2, $n = 22$) = 1.62, $p = .45$. In P2R1 scenarios, a significant overall difference was found between the three types, χ^2 (2, $n = 22$) = 18.71, $p < .001$. Follow-up analyses showed significant differences between "prior" and "subsequent" ($Z = -3.43$, $p < .001$), between "prior" and "both" ($Z = -3.31$, $p < .001$), but not between "subsequent" and "both" ($Z = -1.29$, $p = .27$). Finally, in P2R2 scenarios, again there was a significant overall difference between the three types, χ^2 (2, $n = 22$) = 14.83, $p < .001$. Follow-up analyses revealed significant differences between "prior" and "subsequent" ($Z = -3.68$, $p < .001$), between "subsequent" and "both" ($Z = -2.75$, $p = .004$), but not between "prior" and "both" ($Z = -1.42$, $p = .16$). The results of the statistical analyses are summarized in the column under the heading "external modifications" in Table 6.

Table 5. Occurrence of external modifications in relation to request head acts

	P1R1		P1R2		P2R1		P2R2	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Prior	0.14	0.19	0.30	0.22	0.48	0.22	0.51	0.28
Subsequent	0.19	0.21	0.25	0.24	0.15	0.13	0.09	0.15
Both	0.04	0.13	0.32	0.29	0.11	0.19	0.35	0.31

Table 6. Summary of contextual variations for external modifications and grounders

	External modifications	Grounders
P1R1	Prior = Subsequent; Prior = Both; Subsequent > Both	Prior = Subsequent > Both
P1R2	Prior = Subsequent = Both	Prior > Subsequent > Both
P2R1	Prior > Subsequent = Both	Prior > Subsequent > Both
P2R2	Prior = Both > Subsequent	Prior > Subsequent > Both

Note. = denotes no statistically significant difference; > denotes statistically significant difference with the category on the left side of the symbol having a larger frequency of production than the category on the right side.

Turning to contextual variations in the occurrence of grounders in relation to request head acts, Table 7 displays the descriptive statistics showing the mean frequency of three types of sequential organizations of grounders: "prior", "subsequent", and "both". Concerning P1R1 scenarios, a significant overall difference in production frequency was found between the three types, χ^2 (2, $n = 22$) = 13.79, $p = .001$. Follow-up tests revealed significant differences between "prior" and "both" ($Z = -3.07$, $p = .001$), between "subsequent" and "both" ($Z = -2.89$,

$p = .002$), but not between "prior" and "subsequent" ($Z = -0.81$, $p = .99$). In P1R2 scenarios, there was a significant overall difference between the three types, χ^2 (2, $n = 22$) = 27.91, $p < .001$. Subsequent tests revealed significant differences between "prior" and "subsequent" ($Z = -2.83$, $p = .003$), between "prior" and "both" ($Z = -3.88$, $p < .001$), and between "subsequent" and "both" ($Z = -2.72$, $p = .004$). In P2R1 scenarios, there was also an overall significant difference between the three types, χ^2 (2, $n = 22$) = 38.00, $p < .001$. Follow-up tests showed significant differences between "prior" and "subsequent" ($Z = -3.94$, $p < .001$), between "prior" and "both" ($Z = -4.18$, $p < .001$), and between "subsequent" and "both" ($Z = -3.49$, $p < .001$). Finally, concerning P2R2 scenarios, there was again an overall significant difference between the three types, χ^2 (2, $n = 22$) = 31.68, $p < .001$. Subsequent tests showed significant differences between "prior" and "subsequent" ($Z = -3.32$, $p < .001$), between "prior" and "both" ($Z = -3.96$, $p < .001$), and between "subsequent" and "both" ($Z = -2.81$, $p = .005$). The column under the heading "grounders" in Table 6 summarizes the findings of these statistical analyses.

Table 7. Occurrence of grounders in relation to request head acts

	P1R1		P1R2		P2R1		P2R2	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Prior	0.14	0.17	0.43	0.25	0.53	0.17	0.68	0.28
Subsequent	0.13	0.16	0.15	0.20	0.15	0.13	0.15	0.20
Both	0.00	0.00	0.01	0.04	0.00	0.00	0.01	0.06

5.   Discussion

RQ1 asks whether and how the frequency of producing internal and external modifications varies according to power and imposition. Statistical analyses revealed that both contextual variables significantly affected the production of both internal and external modifications. Generally, high imposition requests and higher hearer power status led to more frequent use of internal or external modifications. These findings make sense from the perspective of politeness theory proposed by Brown & Levinson (1987). According to them, request-making is a face-threatening act (FTA), and the severity of an FTA is dependent upon contingent contextual variables such as power, social distance, and imposition (Brown & Levinson 1987, p. 76). Hence, high imposition requests and asking for favors from someone of higher power status can lead to a higher degree of the severity of FTA, which requires more extensive mitigating efforts to convey politeness. In comparison, low imposition requests and asking for favors from someone of equal power status

result in a lower degree of the severity of FTA; consequently, fewer modifications are needed to balance out the degree of face threat.

Other than the general patterns of modification production depicted above, perhaps more revealing are the results regarding how power and imposition affected the distribution of specific mitigators – these findings can indicate specific associations between particular mitigators and contextual variables. For both internal and external modifications, there were two types of associations with contextual variables: those primarily or exclusively affected by either power or imposition (e.g., the honorific pronoun 您 (*nín*) occurred exclusively in the P2 scenarios; *promise* as an external modifier was mainly used in R2 scenarios), and those influenced by both power and imposition (e.g., *grounder* as an external mitigator, and the understater 一下, *yīxià*, a little bit). While discussing the reasons underlying each of those associations is beyond the scope of this study, the findings reported here can enrich our understanding of the usage of certain modifiers. For example, although the mitigating function of 一下 (*yīxià*, a little bit) has been noted in the literature (e.g., Jiang 2012; Lu & Wang 2005; Zhang 1995), our findings further show that this function is more likely to be used in low imposition than in high imposition requests, and that it is more frequently produced in equal status scenarios than in hearer higher status scenarios. Among such instances, the linguistic structures containing 一下 (*yīxià*) include *verb* +一下 (verb + *yīxià*) and *verb* +*object* +一下 (verb + *yīxià*), as shown in the following examples: 报纸借我看一下 (*Bàozhǐ jiè wǒ kàn yīxià.* Let me read your newspaper.), and 你能不能帮我一下？ (*Nǐ néngbùnéng bāng wǒ yīxià?* Could you help me?). However, the two contextual variables did not appear to influence which of the two structures would be used in specific requests.

RQ2 investigates contextual variations in the preferred sequential organization(s) of external modifications in general and grounders in particular. The focus of this research question is twofold: the first is to investigate whether there is a predominant sequential organization(s) in each of the four context types (i.e., P1R1, P1R2, P2R1, and P2R2), and the second is to examine whether the predominant sequential organization(s) varies across the four context types. The results showed a complex picture for external modifications. As Table 6 shows, the predominant sequential organization(s) differed across the four context types. In P1R1 scenario, both "prior" and "subsequent" were equally dominating; in P1R2 scenarios, there was no dominating sequential organization; in P2R1 scenarios, "prior" was the only dominating sequential organization; and in P2R2 scenarios, "prior" and "both" were equally dominating. In spite of the variations, "prior" was the only recurring sequential organization across all four context types. In terms of the preferred sequential organization of grounders across the four context types, the patterns are more straightforward. On the one hand, the occurrence of "both"

was almost negligible; on the other hand, between "prior" and "subsequent", except in P1R1 scenarios where "prior" and "subsequent" were equally dominating, "prior" was the only dominating sequential organization in all three remaining context types (P1R2, P2R1, and P2R2). Overall, the only recurring sequential organization across all four context types was "prior".

The aforementioned results partially confirmed previous findings in the literature. Regarding the preferred sequential organization of external modifications, the fact that "prior" was the sole dominating discourse structure or one of the dominating discourse structures across all four context types does seem to reaffirm the prevalent argument that the preferred discourse structure of Chinese requests is to place external modifications before head acts (Dong 2008; Jia & Huang 2008; Kirkpatrick 1991; Zhu 2017). However, our findings clearly suggest a need for qualification of such an argument, because the other two sequential organizations (i.e., "subsequent" and "both") were as dominating as "prior" in three of the four context types. In other words, it may be an overgeneralization to say that native Chinese speakers prefer to put forward their requests after external modifications regardless of contextual constraints.

With respect to the sequential organization of grounders, the results both confirmed and enriched those reported by Hong (2002). As in Hong's study, "prior" was the only dominating sequential organization in hearer higher power scenarios (i.e., combining P2R1 and P2R2 scenarios together). In equal power scenarios, Hong reported more post-grounders than pre-grounders. However, our findings suggest that the situation may be less straightforward depending on the level of imposition involved in a request. Specifically, in equal power and low imposition scenarios, both "prior" and "subsequent" were equally dominating; in equal power and high imposition scenarios, "prior" was the sole dominating sequential organization. The discrepancy between Hong (2002) and this study could have two possible explanations. First, Hong was interested in the effects of power only, and she did not intend to examine the role of other contextual variables. In contrast, this study included both power and imposition variables by design, and therefore was able to obtain a nuanced picture regarding the effects of both contextual variables. Second, Hong's study included only three scenarios, whereas this study contained 20 scenarios. Hence, from a sampling perspective, this study is likely to yield more balanced and reliable results.

6. Conclusions, limitations, and future research

In summary, this study investigated native Chinese speakers' contextual variations in producing modifications in request-making. The overall frequency of using

internal and external modifications was significantly affected by both power and imposition; however, specific internal and external modifiers were found to be mainly associated with one or the other contextual variable. Concerning the preferred sequential organization(s) of external modifications in general and grounders in particular, it was found that the preferred (i.e., dominating) discourse structure(s) differed depending on context type. All of these findings point to the necessity of considering contextual constraints in researching request-making in Chinese. Request-making is a face-threatening act (Brown & Levinson 1987), and its appropriateness in social communication is often contingent upon whether a speaker is able to utilize adequate linguistic resources (e.g., a specific internal modification device, a specific sequential organization of request-making) to address a hearer's negative face needs in specific contexts. In revealing the connections between specific forms of modifications for request-making and specific combinations of contextual variables, this study can contribute to a refined understanding of the complex form-function-context mappings in Chinese pragmatics.

This study is limited in three ways and future research is needed to further contribute to the development of this line of research. The first limitation is the relatively small sample size of 22 participants. The findings reported here are thus tentative and studies with a larger sample size are needed to check the generalizability of the findings. The second limitation is that only two contextual variables were included as independent variables in the design, while additional contextual variables such as social distance should also be explicitly examined for the effects on their use in request-making. Similarly, more levels within each contextual variable can also be included to enable a more fine-grained understanding of the effects of a specific contextual variable on request modifications. For example, in addition to equal power and hearer higher power scenarios, one can add hearer lower power scenarios to better investigate whether and how the three different levels of power difference can influence the use of request modification. Finally, although ODCT as a data collection instrument is widely used in pragmatics research, it is limited in terms of the authenticity of the data collected through it (Taguchi & Roever 2017). Hence, the findings reported in this study should be validated through naturalistic data.

Acknowledgements

References

Blum-Kulka, Shoshana, Julia House, and Gabriele Kasper, G. 1989. *Cross-Cultural pragmatics: requests and apologies*. Norwood, NJ: Ablex.

Brown, Penelope, and Stephen Levinson. 1987. *Politeness*. Cambridge: Cambridge University.

Caffi, Claudia. 1999. "On Mitigation". *Journal of Pragmatics* 31(7): 881–909.

Caffi, Claudia. 2007. *Mitigation*. Amsterdam: Elsevier.

Dong, Xinran. 2008. Chinese requests in academic settings. In *Proceedings of the 20th North American Conference on Chinese Linguistics (NACCL-20), Volume 2*, ed. by Marjorie K. M. Chan, and Hana Kang, 975–988. Retrieved: https://naccl.osu.edu/naccl-20/proceedings.

Fraser, Bruce. 2010. "Pragmatic competence: In the case of hedging." In *New approaches to hedging*, ed. by Gunther Kaltenböck, Wiltrud Mihatsch, and Stefan Schneider, 15–34. Emerald Group Publishing Limited.

Hong, Wei. 1996. "An empirical study of Chinese request strategies." *International Journal of the Sociology of Language* 122: 127–138.

Hong, Wei. 2002. "How does power affect Chinese politeness – An empirical study of Chinese Directives." *Journal of Chinese Language Teachers Association* 37(2): 59–74.

Jia, Xuerui and Furong Huang. 2008. "A Contrastive Study of Requests in Chinese and American Cultures." *Intercultural Communication Studies* XVII(2): 132–148.

Jiang, Xiangping. 2012. "A pragmatic analysis of V+yixia in mandarin Chinese". *Journal of Pragmatics* 44(13): 1888–1901.

Kirkpatrick, Andy. 1991. "Information sequencing in Mandarin letters of request." *Anthropological Linguistics* 33(2): 183–203.

Lee-Wong, Songmei. 2000. *Politeness and face in Chinese culture*. Frankfort am Main: Peter Lang.

Li, Shuai. 2014. "The effects of different levels of linguistic proficiency on the development of L2 Chinese request production during study abroad." *System* 45: 103–116.

Lu, Fubo, and Yimng Wu. 2005. "请求句中"V"、"V一下"与"VV"的语用差异." [Pragmatic differences in between "V", "V一下" and "VV" in requests]. *语言教学与研究* 4: 40–45.

Rue, Yong-Ju, and Grace Qiao Zhang. 2008. *Request strategies: A comparative study in Mandarin Chinese and Korean*. Amsterdam: John Benjamins.

Spencer-Oatey, Helen. 1993. "Conception of social relations and pragmatics research." *Journal of Pragmatics* 20: 27–47.

Spencer-Oatey, Helen. 2000. "Rapport management: A framework for analysis." In *Culturally speaking: managing rapport through talk across cultures*, ed. by Helen Spencer-Oatey, 11–46. London and New York: Continuum.

Taguchi, Naoko, and Carsten Roever. 2017. *Second language pragmatics*. Oxford/New York: Oxford University Press.

Takahashi, Satomi. 1998. "Quantifying requestive imposition: Validation and selection of situations for L2 pragmatic research." *Studies in Language and Culture* 9: 135–159.

Thaler, Verena. 2012. "Mitigation as modification of illocutionary force." *Journal of Pragmatics* 44: 907–919.

Wen, Xiaohong. 2014. "Pragmatic development: an exploratory study of requests by learners of Chinese." In *Studies in second language acquisition of Chinese*, ed. by Zhaohong Han, 30–56. Britol/Buffalo/Toronto: Multilingual Matters.

Yeung, Lorrita. N. T. 1997. "Polite requests in English and Chinese business correspondence in Hong Kong." *Journal of Pragmatics* 27: 505–522.

Zhan, Kaidi. 1992. *The strategies of politeness in the Chinese language*. Berkeley, CA: University of California, Berkeley.

Zhang, Yanying. 1995. "Strategies in Chinese requesting." In *Pragmatics of Chinese as native and target language*, ed. by Gabriele Kasper, 23–68. Honolulu, HI: University of Hawaii, Second Language Teaching & Curriculum Center.

Zhang, Shaojie, and Xiaotong Wang. 1997. "请求"言语行为的对比研究." [A contrastive study on the speech act of request]. 现代外语 3: 63–72.

Zhu, Wuhan. 2017. "A cross-cultural pragmatic study of rapport-management strategies in Chinese and English academic upward request emails." *Language and Intercultural Communication* 17(2): 210–228.

Appendix A. Request scenarios in the Oral Discourse Completion Test

P1R1 Scenarios (equal power status & low imposition)

1. During class break, LI, Xiaochen wants to borrow and read WANG, Ning's newspaper, which is on WANG, Ning's desk. LI, Xiaochen says:
2. There will be a talent show in the evening, and students are decorating the classroom. LI, Xiaochen wants to ask WANG, Ning, who is standing nearby, to get some thumbtacks. The thumbtacks are on a nearby desk. LI, Xiaochen says:
3. During class break, LI, Xiaochen's cell phone runs out of power. So LI, Xiaochen wants to borrow WANG, Ning's cell phone to send a text message. WANG, Ning is sitting next to LI, Xiaochen. LI, Xiaochen says:
4. In the classroom, LI, Xiaochen feels a bit stuffy, so he/she wants to ask WANG, Ning, who is sitting next to the window, to open the window. LI, Xiaochen says:
5. During class break, LI, Xiaochen's pencil does not work. So LI, Xiaochen wants to borrow one from WANG, Ning. LI, Xiaochen says:

P1R2 Scenarios (equal power status & high imposition)

6. There will be a concert on Sunday. LI, Xiaochen wants to go very much but does not have time to buy a ticket. So LI, Xiaochen wants to ask WANG, Ning to buy a ticket for him/her. LI, Xiaochen knows that WANG, Ning is very busy and the ticket box is a bit far from their university. LI, Xiaochen meets WANG, Ning. LI, Xiaochen says:
7. The final examination is coming within two days. LI, Xiaochen wants to borrow WANG, Ning's notes for one day and to give it back the next day. LI, Xiaochen knows that WANG, Ning will also need the notes to prepare for the exam. LI, Xiaochen meets WANG, Ning during class break. LI, Xiaochen says:
8. LI, Xiaochen has an internship next week and needs a laptop. LI, Xiaochen does not have a laptop but knows that WANG, Ning has just purchased a very expensive one a few days ago. So LI, Xiaochen wants to borrow WANG, Ning's laptop for one week. LI, Xiaochen meets WANG, Ning during class break. LI, Xiaochen says:
9. LI, Xiaochen wants to buy a laptop, which is a bit expensive, so he/she wants to borrow a comparatively large sum of money from WANG, Ning. LI, Xiaochen meets WANG, Ning. LI, Xiaochen says:

10. LI, Xiaochen's friend is coming to Beijing to visit him/her today. LI, Xiaochen cannot meet the friend at the airport as he/she has got some things to do today. So LI, Xiaochen wants to ask WANG, Ning to help meet the friend at the airport. The airport is very far from their university and LI, Xiaochen knows that WANG, Ning is very busy. LI, Xiaochen says:

P2R1 Scenarios (hearer higher power status & low imposition)

11. Yesterday, Professor Chen gave out some handouts for his/her class. LI, Xiaochen didn't come to the class due to illness. So LI, Xiaochen wants to get a copy of the handout from Professor Chen. LI, Xiaochen comes to Professor Chen's office. LI, Xiaochen says:

12. LI, Xiaochen didn't quite understand one point during Professor Chen's lecture. So during class break, LI, Xiaochen wants Professor Chen to explain that point for him/her. LI, Xiaochen says:

13. At Professor Chen's office, LI, Xiaochen is discussing some questions with Professor Chen. LI, Xiaochen wants to borrow a pencil from Professor Chen so as to take down what they are talking about. LI, Xiaochen says:

14. Professor Chen used a PowerPoint file in his/her lecture. LI, Xiaochen wants to have a copy of the file. LI, Xiaochen meets Professor Chen after the class. LI, Xiaochen says:

15. Professor Chen said that he/she had sent the individualized assignments to each student's email box. But LI, Xiaochen hasn't got his/her e-mail and therefore wants to ask Professor Chen to send it again. LI, Xiaochen meets Professor Chen during class break. LI, Xiaochen says:

P2R2 Scenarios (Hearer higher power status & high imposition)

16. The final examination will be held the day after tomorrow. LI, Xiaochen cannot attend the exam on that day because he has got something to do, so he/she wants to ask Professor Chen to agree to let him/her take the exam one day after the test date. LI, Xiaochen says:

17. A term paper is due today but LI, Xiaochen hasn't finished it yet. So he/she wants to ask Professor Chen to agree to extend the due date for him/her. LI, Xiaochen comes to Professor Chen's office. LI, Xiaochen says:

18. In the school bookstore, LI, Xiaochen wants to buy a book, which is a bit expensive. But LI, Xiaochen finds that he/she didn't bring money. Just then, LI, Xiaochen sees Professor Chen, so he/she wants to borrow money from Professor Chen. LI, Xiaochen says:

19. Professor Chen is going to attend a conference in America. LI, Xiaochen wants to ask Professor Chen to help buy several books. LI, Xiaochen knows that the conference has a very tight schedule. LI, Xiaochen comes to Professor Chen's office. LI, Xiaochen says:

20. LI, Xiaochen wants to borrow a book from The Beijing Library but doesn't have a library card. LI, Xiaochen knows that Professor Chen has the card. So LI, Xiaochen wants to ask Professor Chen to go to The Beijing Library and help borrow the book. The Beijing Library is a bit far from their university. LI, Xiaochen comes to Professor Chen's office. LI, Xiaochen says:

Some interactional functions of *Yinwei*-clauses in Mandarin Chinese conversation

Xiaoting Li and Jie Luo

University of Alberta, Canada / University of Calgary, Canada

Yinwei 'because' is a causal conjunction or preposition indicating a causal relation between two clauses, NPs and other discourse units in Mandarin Chinese. Building on the previous research, this study examines how *yinwei* is used by conversational participants to organize talk and accomplish interactional tasks in Mandarin conversation. Adopting the methodologies of conversation analysis and interactional linguistics, this study examines 11 hours of everyday Mandarin conversational data, and explores the interactional functions of *yinwei*-clauses. An examination of the data shows that *yinwei*-clauses have a variety of interactional functions in everyday Mandarin conversation. Two particular interactional functions of *yinwei*-clauses are accounts for a speaker's prior action such as disagreement and strong assertion, and parentheticals providing background information related to the ongoing talk.

Keywords: *yinwei*, Mandarin, conversation analysis, interactional linguistics, accounts for prior actions, parentheticals, disagreement, assertion

1. Introduction

In Mandarin Chinese,[1] *yinwei* is a conjunction introducing reason or cause, or a preposition followed by an NP of reason or cause (Lü 1999: 622–623). Excerpts (1) and (2) demonstrate the use of *yinwei* as a conjunction introducing reason or cause, and Excerpt (3) shows the use of *yinwei* as a preposition.

---

1. In this paper, Mandarin Chinese refers to the standard variety of Modern Chinese. For brevity, the term "Mandarin" will be used in the rest of the paper.

(1) 因　为　治　疗　及时所以他的伤　很　快　就　好了。
     yinwei   zhiliao jishi   suoyi ta de   shang hen kuai jiu hao le
     because cure   timely so   3SG POSS   injury very quick just good PFV
     'Because cured in a timely manner, he recovered from injury quickly.'
                                                                    (Lü 1980: 623)

(2) 昨　　天 我没去找　你 因　　为　有　别　的事。
     zuotian   wo mei   qu zhao ni   yinwei   you bie   de   shi
     yesterday I   NEG   go find   you because have other POSS thing
     'I didn't go to find you yesterday, because I have some other things to do.'
                                                                    (Lü 1980: 623)

In Excerpts (1) and (2), *yinwei* is used as a conjunction that introduces the reason
clause. *Yinwei* clause can be positioned either before (see Excerpt (1)) or after (see
Excerpt (2)) its associated main clause.

(3) 小田　因为　这　　件事还　受到　了表扬。
     xiaotian yinwei   zhejian shi   hai shoudao le   biaoyang
     litte tian because this CL event also get   PFV praise
     'Little Tian was praised because of this event.'   (Lü 1980: 622)

In Excerpt (3), *yinwei* is used as a preposition followed by an NP *zhe jian shi* 'this
event'. The descriptions of the uses of *yinwei* so far have been mainly based on
invented sentences and the researcher's own introspection as a native speaker of
Mandarin. The research question of this study is: what are the interactional func-
tions of *yinwei*-clause in unscripted everyday Mandarin conversation? That is,
how is *yinwei*-clause used by conversational participants to build talk and action
in Mandarin conversation? Before proceeding to the discussions of the interac-
tional functions of *yinwei* in Mandarin conversation, we will provide some rel-
evant theoretical and methodological background of the study.

2. Theoretical and methodological background

When being used as a conjunction, *yinwei* introduces a reason clause indicating
causes or reasons for a result (Xing 2001: 59–60; Lü 1999). The *yinwei*-clauses can
be either pre- or post-positioned (Lü 1982). Between the two positioning formats,
pre-positioning of *yinwei*-clause is considered as the default sequence (Chao 1968).
Kirkpatrick (1993) and Young (1982) also argue that pre-positioning of *yinwei*-
clause is preferred by Chinese speakers, because they tend to present supporting
materials for the main point before producing that point. However, based on her
natural Mandarin conversational data, Biq (1995) observes that *yinwei*-clauses are

predominantly (82%) post-positioned in Mandarin conversation. Wang's (1999) findings based on her Mandarin conversational data are consistent with those of Biq (1995) in that majority of *yinwei*-clauses appear after result clauses. Song & Tao (2008, 2009) have examined the positions of *yinwei*-clauses in everyday Mandarin telephone conversation. They have similar findings that post-positioning of *yinwei*-clauses is pervasive in everyday Mandarin conversation. Biq (1995) and Song & Tao's (2008, 2009) studies show that the prevalent position of *yinwei*-clauses in natural spoken discourse may diverge from what has been described in Mandarin grammar (e.g., Chao 1968).

Research on the functions of *yinwei* has also gone through a change of focus from its semantic truth value as a causal conjunction in written discourse (Chao 1968; Xing 2001) to its discourse and interactional use in natural spoken discourse (Wang 1999; Fang 2000; Li 2016). Xing (2001: 49) proposes tripartite clausal relations: causal, contrastive, and coordinate. *Yinwei* is the normative conjunction used to indicate the causal relation between two clauses (Xing 2001: 57). Chao (1968: 115) mentions that when post-positioned, *yinwei*-clauses usually serve as afterthought. Based on observations of her Mandarin conversational data, Wang (1999, 2002) describes that *yinwei* often connects utterances that are not in causal relations. In her data, *yinwei*-clauses are used to indicate elaboration, modification, and/or explanation for the speaker's prior utterance (Wang 1999: 75–77, 2002: 156–157). Although their focus is on the sequencing of *yinwei*-clauses, Song & Tao (2008: 65) briefly discuss two main interactional uses of *yinwei* in Mandarin conversation. One is to further elaborate the speaker's stance to pursue the recipient's affiliative responses; and the other is to account for one's disaffiliative responses. The second interactional function of *yinwei* is also observed in our data. Chang & Su (2012) examines the discourse-pragmatic functions of the causal marker *inui* 'because' in Taiwanese Southern Min (TSM) conversation[2] by comparing them with unmarked causal utterances. They report that *inui* have five types of functions in their data: "pure-cause marking", marking an explanation, marking a justification, displaying understanding, and "information interpolation". Although the present study focuses on the causal conjunction *yinwei* in *Putonghu*a (i.e., the modern standard Chinese in mainland China) conversation, similar uses of *yinwei* as *inui* in TSM, such as "pure-cause marking", marking an explanation or justification and "information interpolation", are also seen in the present data. Li (2016) provides a fine-grained analysis of two particular discourse-interactional functions of *yinwei* in her face-to-face Mandarin conversational data. Specifically, she documents that when occurring after the possible completion of a recipient-initiated sequence that may change the ongoing topic,

---

2. Southern Min is a regional variety of modern Chinese (Chen 1999).

yinwei is deployed by the speaker as a device to continue a pre-prior course of action to pursue his/her original interactional agenda. Further, when appearing after the possible closure of a storytelling, *yinwei* is used to reopen the just closed telling to pursue affiliative response from the recipient.

The previous research has shown that *yinwei* and *yinwei*-clauses have a variety of dynamic interactional functions that are undocumented in the Chinese linguistics literature. This study attempts to contribute to this growing body of research, and to examine the interactional uses of *yinwei*-clauses in everyday Mandarin conversation. Before proceeding to reporting our findings, it is necessary to discuss the data and methodology adopted in this study.

The conversational data used in the previous research are naturalistic everyday Mandarin conversation (Biq 1995; Song & Tao 2008; Li 2016), or a combination of naturalistic conversational interaction and programs in the media (such as interviews and phone-in radio programs) (Wang 1999: 2000; Fang 2000; Chang & Su 2012). The data for this study consist of 11 hours of video recordings of everyday Mandarin face-to-face conversation. There were 26 participants in total, including 13 females and 13 males. The conversations were recorded in Beijing, China, Germany, and Canada from 2008 to 2014. All participants were native speakers of Mandarin. Their ages ranged from 23 to 60. Among the 26 participants, 6 were undergraduate students studying in Germany for 1–3 years; 12 were graduate students studying in China; 2 were retired persons in China; and 6 were undergraduate students studying in Canada for 1–3 years.[3] The constellations of the participants included dyads (three) and multi-party conversations (including four triads, and two quartets). Different settings, including lunch conversations, tea conversations, cooking, chit-chatting, playing games and project discussion, were selected to ensure that the result is not context-specific.

The data were manually transcribed and coded by two native speakers of Mandarin through repeated listening of the data. When divergence in the coding occurred, a third native speaker of Mandarin was asked to listen to the recording and check the coding. After the verification and cross-check process, the two coders agreed on all instances of *yinwei*s identified in the data. The data yield a total of 265 occurrences of *yinwei*-clause. Among them, 89.8% ($N = 238$) are used as accounts for prior actions; 5.7% ($N = 15$) serve as parenthetical insertions; and 4.5% ($N = 12$) are used to accomplish a variety of other interactional tasks such as turn management.

The data were analyzed using the methodology of conversation analysis and interactional linguistics which focus on the linguistic resources and interlocutors' orientation to them in talk-in-interaction (see Heritage 1989; Couper- Kuhlen &

---

3. The potential limitations of the data are discussed in the conclusion section of the paper.

Selting 2001). Conversation analysis (CA) examines the organization of talk and action from the participants' perspective through displays of their understanding of 'what is going on' in interaction (Hutchby & Wooffitt 2008: 12; Heritage 1989). Particularly, interactants display their understanding of the prior talk in sequentially next turns. This is described as the next-turn proof procedure that is a fundamental method used in CA and interactional linguistics. This method ensures that analytical categories and claims made are oriented to as such by participants themselves, rather than subjective interpretations of the analyst.

An examination of the data shows that *yinwei*-clauses have two main interactional functions: accounts for a prior action, and parentheticals introducing background information. These two functions are discussed in Sections 3 and 4.

3.   *Yinwei*-clauses as accounts for speakers' prior actions

Interlocutors perform a variety of actions in and through conversation: greetings, farewells, requests, offers, invitations, complaints, praising, etc. Some actions are performed without the need for an account such as greetings, farewells, and offers, whereas others, such as disagreement and strong assertion, require an explanation or account (Couper-Kuhlen 2011; Sacks 1992). In our data, *yinwei*-clauses can be used as accounts for two particular types of actions: disagreement and assertion.

3.1   Accounts for disagreements

It is not uncommon that interlocutors may disagree with one another in daily conversation. After expressing a disagreement with another interlocutor, a speaker tends to explain the reason or motivation for that disagreement. In the data, *yinwei*-clauses can serve as accounts for such disagreements against a claim made by a prior speaker in an immediately preceding turn. Excerpt (4) provides a case in point.

Excerpt (4) is taken from a conversation between three friends Bai, Hei, and Lu. They are discussing why some companies' employees will fly first class and stay in luxury hotels on business trips.

(4)   First class and luxury hotels
01   Bai: 有些 公司　　 啊;
             youxie gongsi   a;
             some   company PRT
                'Some companies,'

02   他们 那些 出差   的 人员 每 坐 飞机 头等
            tamen naxie chuchai   de renyuan mei zuo feiji toudeng
            they those official business ASSC staff   every take plane first-class
            舱;
            cang;
            cabins
            'their employees take first class flights when they go on business trips,'

03   然后 什么 几 星 级 的 宾馆   就 这 种   嗯
            ranhou shenme ji   xing ji de bingguann jiu zhe zhong en
            then what several star level ASSC hotel   just this kind um
            [这 是.
            [*zhe shi.*
            [this is
            'and stay in luxury hotels.'

04 Hei: [嗯..
            [*en.*
            [mm
            'Mm.'

05 Bai: 不   能   如果 住 住 得 低 了   或是 飞机 坐 的   舱
            buneng ruguo zhu zhu de di le huoshi feiji zuo de cang cang
            NEG can if   live live CSC low PRT or   plane take ASSC cabin
            舱 等 (.) 低 了;
            deng (.) di le;
            cabin   (.) low PRT
            'If employees stay in standard hotels or take economy flights,'

06   这 是 影响   公司   形象   的 嘛.
            zhe shi yingxiang gongsi   xingxiang de ma.
            this be influence company image   PRT PRT
            'it will affect the company's image.'

07 Hei: 那 不 是 公司   形象;
            na bu shi gongsi   xingxiang;
            that NEG be company image
            'That's not about the company's image,'

08 →   因为 我们 公司   是 两 个 小时 以内 都 是
            yinwei women gongsi   shi liang ge xiaoshi yinei dou shi
            because our   company be two CL hour   within all be
            经济   舱;
            jingji   cang;
            economy cabin
            'because our company (requires employees to) take economy class for
            the flights that are less than two hours.'

09   然后 超过　两 个 小 小时 就 是: [商务　[或者 什么
ranhou chaoguo liang ge　xiao xiaoshi jiu　shi: [shangwu [huozhe shenme
then　more　than two CL　hour　just be: [business [or　what
的 嗯.
de　en.
PRT PRT
'And if it's more than two hours it'll be business (class), or something like that.'

10 Bai: [头等　舱　嗯
                                           [*toudeng　cang　en*
                                         [first-class cabin uh
商　嗯.
shang　en.
business uh
                                        'First-class, uh, business
class.'

11 Lu: [商务　舱;
                                    [*shangwu cang;*
                                    [business class
                                    'Business class.'

At the beginning of this interaction, Bai reports on the luxury travel conditions of some companies (lines 1–3), which was receipted by Hei's minimal acknowledgement *en* 'Mm' (line 4). Bai continues her turn that if their employees stay in budget hotels or fly economy class, it will affect the companies' image (lines 5 to 6). The *shi…de* structure in line 6 is used to emphasize and affirm the statement that lower travel and accommodation standards "negatively affect the companies' image" (Li & Thompson 1981: 589). Also, the sentence-final particle *ma* indicates that this statement is self-evident (Lü 1999: 375) and thus further adds to its assertiveness. After this assertion, agreeing responses from recipients are preferred (Schegloff 2007: 68). However, Hei ostensibly disagrees with Bai by negating Bai's statement using the negative particle *bushi* (line 7), which is a dispreferred response (Pomerantz 1984) to Bai's assertion. Immediately after her disagreement, Hei uses *yinwei* to introduce an account for her disagreement by stating that it is not about the company's image but rather the duration of the flight (lines 8 to 9).

This example shows that *yinwei*-clause is used directly after a recipient's disagreement to provide an account for the dispreferred response. Excerpt (5) is another example in which *yinwei*-clause is used to serve the same function.

In this excerpt, Yan, Zha, Luo and Cha are playing a card game called *Shuangsheng*[4] 'Upgrade'. Prior to the excerpt, Zha had just played three cards (a hand)[5] on the table King-King-Ace.[6] Yan, Zha, and Cha are discussing whether the hand that Zha played is legitimate or not.

(5)   Chinese card game

```
01  Cha:    这  是 什么;
            zhe  shi shenme
            this be  what
            'What is this?'
            Gaze        at                          Zha
02  Yan:    主上      边三 轮    是 [A       A  K;
            zhushang  biansanlun  shi [jianjianK
            Zhushang Bian sanlun be [ace        ace king
            'Zhushang Biansanlun[7] is Ace-Ace-King.'
03  Zha:                            [边  三  轮,
                                    [bian san lun,
                                    [bian san lun
                                    'Biansanlun.'[8]
04          A   (.) AAK       KKA       都 可以 呀;
            jian (.) jianjianK    KKAjian    dou keyi  ya;
            ace  (.) ace ace king king king ace all   ok    PRT
            'Both Ace-Ace-King and King-King-Ace are okay. '
05  Yan:    不  行;
            bu  xing;
            NEG OK
            'No.'
06  →       因为    如果 如果 你  出 了  如果 你  出 了  这 个 这 个 A;
            yinwei  ruguo ruguo ni  chu le   ruguo ni  chu le  zhe ge zhe ge jian;
            because if    if    you out PFV if    you out PFV this CL this CL ace
            'Because if you play this Ace,'
```

---

4. *Shuangsheng* 'Upgrade' is a four player Chinese card game. Players are divided into two groups to compete for points. The group that gets the highest number of points in a turn gets upgraded one level. The group with the highest level at the end of the game wins.

5. "A hand" in cards games is used to denote any set of cards being played.

6. King-king-Ace (KKA) is one of multiple card patterns in the game.

7. *Zhushang Biansanlun* is a hand that can be played in the game that denotes Ace-Ace-King.

8. *Biansanlun* is an abbreviation of *Zhushang Biansanlun*.

07   别人　出　个别　　的.
      bieren chu ge bie   de.
      others out CL another ASSC
      'and someone else plays another card'

08   然后　你　这　一对 K　什　　么　的.
      ranhou ni   zhe yidui K   shenme de.
      then   you this one   pair ing   what PRT
      'Then this pair of Kings you have'

09   你　你　这样　　子　　的话 不　能　　　保证 你 是　　最 大 的
      ni   ni   zheyangzi de hua bu   neng baozheng ni   shi zuida de   ya
      you you this way   if   NEG can ensure   you be biggest ASSC PRT
      呀　　可是;
      keshi;
      but
      'cannot ensure that your hand is the best one if you play it like this
      (King-King-Ace).'

In line 1, Cha challenges the hand that Zha played (KKA) with a question 'What is this'. Yan then turns to Zha and asserts that '*Zhushang Biansanlun* is Ace-Ace-King' (line 2) implying that Zha played a wrong hand. Zha explains that *Biansanlun* can be either AAK or KKA (line 4). Yan immediately disagrees with Zha with an explicit disapproval token *buxing* 'not OK' (line 5). Yan's disagreement is a dispreferred response to Zha's claim. After the disagreement, Yan immediately produces a *yinwei*-clasue as an account that *Biansanlun* is AAK, not KKA (lines 6 to 9).

Once again, this example shows that the *yinwei*-clause is used by the recipient as an account for her disagreement against a claim made by the prior speaker.

It can be seen from Excerpts (4) and (5) that *yinwei*-clauses are used to provide an account for an interlocutor's disagreement with the prior speaker.

3.2   Accounts for assertion

Making assertions or claims is a common action that people perform in conversation. Making strong assertions may also be accountable. Our data show that *yinwei*-clauses can be used to account for a speaker's strong assertion.

Excerpt (6) is taken from a conversation among three female participants Min, Per, and Li. They are discussing the layout of Li's apartment where the recording takes place.

(6)   Dorm
01 Per: 我 这 边 整 个 没 插座.
          wo zhe bian zheng ge mei chazuo.
          I   this side whole CL NEG socket
          'I have no electric sockets on this side.'
          (18 lines omitted regarding the discussion about the different position of
          the electric sockets, tables and beds in their rooms.)
19 Min: 那 我 那 个 是 不 是 也 应 该 那 样.
          na   wo na   ge shi bu   shi ye   yinggai na   yang.
          then my that CL be   NEG be also should that
          'Whether that (thing) (in my room) is placed like that?'
20   我 不 知道.
          wo bu   zhidao.
          I   NEG konw
          'I don't know.'
21   因为 我 那 边 没 用 过.
          yinwei   wo na   bian mei yong guo.
          because I   that side NEG use   PFV
          'Because I didn't use that side'
22   我 那 边 床-
          wo na   bian chuang-
          I   that side bed-
          'That side in my room has the bed'
23   我 这 是 柜子 [然后 这 是 床.
          wo zhe shi guizi   [ranhou zhe shi chuang.
          I   this be cabinet [then   this be bed
          '(In my room) this is where the cabinet goes, and this is where the bed
          goes.'
24 Per: [哦 那 个 是 那样 的.
                        [*o   na   ge shi nayang   de.*
                        [oh that CL be that way PRT
                        'Oh. That is placed like this.'
25 Li: 你 肯定 也 是 这 边 有.
          ni   kending ye   shi zhe bian you.
          you must   also be this side have
          'Your room definitely has it on this side'
26 →   [因为 咱俩 是 一 个 走 向 的 屋子.
          [*yinwei   zanlia shi yi   ge   zouxiang de   wuzi.*
          [because we   be same trend PRT   room
          'Because our rooms are the same style.'

27 Per: [对   对   对   对   嗯.
                [*dui   dui   dui   dui   en.*
                [right right right right en
                'Right, right, right, right. Mm.'
28 Min: 嗯 有 可能;
                en   you   keneng;
                mm have maybe
                'Mm, it might be.'

In line 01, Per points out that the electric sockets in her room are in a different location from those in Li's room. Min is uncertain whether the electric sockets in her room are placed on the same wall as the ones in Li's room (lines 19 to 21). After Min's report on the position of the furniture in her room (lines 22 to 23), Li asserts that the electric sockets in Min's room are definitely on the same wall as those in her room (line 25). The adverb *kending* 'definitely' in Li's turn (line 25) conveys a highly assertive tone (Xiandai hanyu cidian 2005: 777). A strong assertion is usually accountable and makes relevant an account. Li provides an account introduced by *yinwei* (line 26) stating that Min's room has the same design as hers.

That *yinwei*-clauses provide accounts for assertion can also be observed in the following example. In Excerpt (7), Nan, Lin, Che, and Shi are talking about why males always find younger females more attractive.

(7)   Males and females
01 Lin: 男   生   为什么 总是 喜欢 比 自己 小
                nansheng weishenme zongshi xihuan bi   ziji xiao   de;
                male   why   always like   more self young PRT 的;
                'Why do males always like younger females?'
                (16 lines omitted regarding the discussion about the human psychology.)
18 Nan: 人类 之前 是 一   种   群居   的   物种 嘛..
                renlei zhiqian shi yizhong qunju de   wuzhong ma
                human before be one   kind sociable ASSC   specie PRT
                'Human beings used to be a sociable species.'
19   群居   物种   他 是 这 样 的.
                qunju   wuzhong ta   shi zhe yang de
                sociable species   3SG be this kind PRT
                'A sociable species is like this,'
20   永   远 会 喜欢 那个 年轻 的.
                yong   yuan hui xihuan nage nianqing de.
                always will like that   CL young   PRT
                '(they) will always like young (females).'

21 →   因为   年轻   的   雌性   生殖力   强,
         yinwei   nianqing de   cixing   shengzhili qiang,
         because younger ASSC femina fertility   strong
         'Because young females have strong fertility.'

22 Lin: 对 [好   好.
         dui [hao   hao.
         right [good good
         'Right, OK, OK.'

23 Nan: [繁殖   力   强;
              [*fanzhi   li   qiang;*
              [fertility strong
         'Good fertility.'

24 Lin: [那   我 我 承   认 你是-
         [*na   wo wo chengren ni   shi-*
         [then I   I   admit   you be-
         'Then I admit you are…'

25 Nan: [请   听   我 请   听   我 说   完;
         [*qing   ting   wo qing   ting   wo shuo wan;*
         [plese listen I   please listen I   say   finish
         'Please let me finish.'

26   那   雌性   为什么   喜欢;
         na   cixing   weishenme xihuan;
         then femina why   like
         'Then why do females like…'

27   就 是女性   为什么   说 喜欢   就是 成熟   的   呢.
         jiu shi nÜxing weishenme shuo xihuan jiushi chengshu de   ne.
         just be female why   say like   just   be   mature ASSC
         'Why do females like mature males?'

To respond to Lin's question in line 1, Nan begins with the common knowledge that human being is a sociable species (line 18). The morphosyntactic structure …*shi zheyang de* '…is like this' (line 19) projects incipient assertive statements about the 'sociable species' (Guo 2012). In line 20, Nan produces the assertion that a sociable species always likes young female. The adverb *yongyuan* 'always' and modal word *hui* 'will' make Nan's utterance in line 20 a rather strong assertion. Immediately after the assertion, Nan provides an account that young females have strong fertility (line 21). Here, the *yinwei*-clause is again used as an account for the speaker's immediately preceding strong assertion.

It should be noted that there are other types of accountable actions, that is, actions that depart from locally relevant expectations and norms (Heritage 1988; Couper-Kuhlen 2009), such as assessment and request for information. However,

yinwei has not been observed to recurrently occur after those types of accountable actions in the current data. Further, *yinwei*-clause is but one morphosyntactic resource that can be used to provide an account in the data. Other morphosyntactic resources such as clauses without *yinwei*, and clauses prefaced by other conjunctions and particles may also be used to perform this interactional function. The discussions of other such morphosyntactic resources used in providing accounts and the use of *yinwei*-clauses as accounts for other types of actions are beyond the scope of this chapter and not supported by the current data. Future research is needed to address those questions.

To exemplify the variety of morphosyntactic resources that can be used to provide accounts in addition to the *yinwei*-clause and particularly the non-use of *yinwei* in clauses providing accounts, as is suggested by one of the anonymous Reviewers, the following excerpt is provided.

In Excerpt (8), Bin (female) and Lei (female) are arguing about the size of a studio apartment that Lei would like to move into, which is referred to as *na fangjian* "that room" in line 1. Bin happens to live in the apartment building where the studio apartment is.

```
    (8)   The room
1   Bin:    那　房间　　跟　你　现在　　房间　　一样　大　啊-
            na  fangjian gen ni  xianzai fangjian yiyang da a-
            that room    with you now     room    same  big PRT
            'That studio apartment is as big as the one you are living in now.'
2           还　没　阳　[台;
            hai  mei yang [tai;
            even NEG      [balcony
            '(It) doesn't even have a balcony.'
3   Lei:                [大　大　大.
                        [da  da  da.
                        [big big big
                        '(It is) big.'
4   Bin:            大　个 [屁.
                    da  ge [pi.
                    big CL [ass
            'Big my ass.'
5   Lei:                [我 去 别人 家　　看　 (.) 绝对　　 大.
                        [wo qu bieren jia   kan (.) juedui    da.
                        [I   go other  home look (.) absolutely big
                        'I went to someone else's home to see it. It is absolutely
            big.'
```

6   Bin: 大 个 屁;
             da ge pi;
             big CL ass
             'Big my ass.'
7   →   它 没 橱:.
             ta   mei chu:.
             3sg NEG cupboard
             'It (has) no cupboard.'
8   (1.0)
9   它 只是 个 柜子;
             ta   zhishi ge guizi;
             3sg just   be CL
             'It's just a cabinet,'
10   所以 你 感觉 大.
             suoyi ni   ganjue da.
             so   you feel   big
             'so you feel (it's) big.'
11   房间   大小 是 一样 的.
             fangjian daxiao shi yiyang de.
             room   size   be same PRT
             'The room size is the same.'
12   你 回来 可以 看 合同   的.
             ni   huilai keyi kan hetong   de.
             you return can   look contract PRT
             'You can check the contract later on.'
13   Lei: 噢.
             ou.
             oh
             Oh.

In this sequence, Bin first presents her opinion that Lei's current apartment is of the same size as the apartment she would like to move into. The other apartment is actually smaller because it does not have a balcony (line 1). But Lei is of different opinion from Bin. Lei claims that the other apartment is big (line 3) and further provides the evidential grounds for her claim; that is, she saw it herself (line 5). Bin expresses her explicit disagreement with Lei through a curse *da ge pi* 'big my ass' (lines 4 and 6). Then Bin provides an account for her disagreement in lines 7–10. That is, the other apartment has less storage space, so one may feel it's bigger. This account is formulated in a complex causal clausal structure marked by *suoyi* 'so' (line 10). That is, because the other apartment does not have a proper cupboard (lines 7 and 9), it feels bigger (line 10). Note that the clause providing account is

not prefaced by *yinwei* or any other conjunctions in line 7. After Lei's account and reiteration of the fact that the two apartments are of the same size (lines 11–12), Bin's response in line 13 registers her acceptance of Lei's opinion as well as the change of her epistemic state (Heritage 1984).

The previous excerpt shows that causal clauses without *yinwei* can also be used to provide accounts for an immediately prior disagreement by the same speaker. The recipient accepts the speaker's opposite opinion after the speaker's accounts. It shows that *yinwei*-clauses, and causal clauses without *yinwei*, among others, are all linguistic resources that can be used to perform interactional tasks such as providing accounts. By focusing on the use of *yinwei*-clauses in providing accounts here, we do not intend to imply that it is an interactional function *unique* to *yinwei*-clauses.

4. *Yinwei*-clause as parentheticals to provide background information

In addition to accounting for a speaker's own prior actions such as disagreement and assertion, *yinwei*-clauses in our data are also used as parentheticals providing background information to the ongoing telling.

Parenthetical insertions are constructions which temporarily suspend the progression of another syntactic construction or wider action sequence (Duvallon & Routarinne 2005). The speaker may halt the current turn-in-progress and insert a *yinwei*-clause as parentheticals. These *yinwei*-clauses usually provide background information for the turn- and telling-in-progress.

Yinwei-clauses as parentheticals occur in two sequential positions: (i) in the middle of an ongoing TCU (turn constructional unit) (Section 4.1); and (ii) between TCUs in a multi-TCU turn (Section 4.2).

4.1 Within an ongoing TCU

Yinwei-clauses as parentheticals may appear in the middle of an ongoing TCU. Excerpt (11) is a case in point. Excerpt (9) is taken from a conversation between two friends Min (male) and Per (female). Min and Per are discussing the famous magician Liu Qian and his magic shows.

(9)   Magic

```
01  Per:   反正    他  我  知道  在北京  卫视  老      看      到他.
            fanzheng ta    wo zhidao zai beijing weishi lao    kandao ta.
            anyway  3SG I   know at Beijing TV     always see     3SG
            'Anyway, I always see him (Liu Qian) on Beijing TV,'

02  Per:   就 中     午 那     段      时间;
            jiu zhongwu nan duan  shijian;
            just noon     that period time
            'during lunch time.'

03  Per:   他 [老      走   在大街上  ] 大街上     给 [给        群众
            ta [lao     zou zai dajie shang] dajie shang gei [gei       qunzhong
            3SG [always walk in  the   street] street give  [give audience show
            变    魔术(.) 啊.
            bian   moshu(.) 啊.
            magic PRT     a.
            'He always shows magic to the audience on the street.'

04  Min:       [北京  卫视           ].              [哦   哦   对
                [beijing weishi.        ]              [o    o    dui
                [beijing TV             ]              [ohoh right right
            对    那些 街头 街头 魔术  那种.
            dui    neixie jietou jietou moshu neizhong.
            those street  street magic that    CL
                'Beijing TV?' 'Oh, right, right, that type of street magic.'

05  Min:   是.
            shi.
            be
            'Right.'

06  Min:   有的 特   神奇;
            youde te     shenqi;
            some  very magic
            'Right, some magic is very miraculous.'

07  Min:   真的   真的  吓   吓   吓   人 那   种.
            zhende zhende xia   xia   xia   ren na   zhong.
            really  really  scary scary scary that kind
            'Really scary.'

08  Min:   ((breathy laughter))

09  Min:   他他在台 湾  那    边 做   得   特    别   多  嘛.
            ta ta zai tai wan na   bian zuo de  te     bie  duo ma.
            he he at Tai wan there do   CSC very much PRT
            'He does a lot of this kind of magic in Taiwan.'
```

10 Min: 然后 在台湾 就 那 个: 就 是(.) 那 个 现场   的
ranhou zai Tai wan jiu na ge: jiu shi(.) na ge xianchang de
then   at Tai wan just that CL just be(.) that CL on-site   ASSC
气   氛.
qifen.
atmosphere
'The atmosphere on site in Taiwan,'

11→Min: 因   为   台 湾   人   比 较 信 鬼 嘛;
yinwei Tai   wan ren   bijiao xin gui ma;
because Taiwaness fairly believe ghost PRT
'because Taiwanese are relatively superstitious.'

12 Per: 嗯.
en.
hm
'Hm.'

13 Min: 现场   气氛   烘托 得 真 真 真   的 跟 就 是
xianchang qifen   hongtuo de zhen zhen zhen de gen jiu shi
spot   atmosphere contrast CSC really really really as just be emerge
出现   灵异   现象   似的.
chuxian   lingyi   xianxiang shide.
supernatural phenomenon
'The atmosphere is like some supernatural phenomenon occurred.'

14 比如   说 他 拿 一 个   可乐瓶 捏 扁 了的.
bi ru   shuo ta na yi ge   kele ping nie bian le de.
for example he   take one CL bottle pinch flat PFV PRT
'For example, he takes a pinched flat cola bottle,'

This interaction begins with Per's telling about the street magic performed by Liu Qian on Beijing TV (lines 1–3). Min makes an assessment of his magic performance in lines 6–7. Min's laughter in line 8 displays that he treats his preceding comment on the magic performance as non-serious (Sacks 1992; Jefferson 1984). Min continues his report that the magician often performs this type of magic in Taiwan in lines 9–10. In line 10, the syntactic structure of Min's turn- and TCU (turn constructional unit)-in-progress is incomplete. Mandarin is topic-prominent language (Li & Thompson 1976, 1981). In line 10, there is only the topic in the sentence: *xianchang de qifen* 'atmosphere on site' with the comment unproduced. In line 11, instead of producing the comment of the sentence, Min halts his TCU-in-progress and produces a *yinwei*-clause 'because Taiwanese are superstitious'. This *yinwei*-clause offers background information related to his report of the magician's performance in Taiwan in the immediately prior talk. Here, the *yinwei*-clause serves as a parenthetical insertion introducing background information for

the ongoing talk. After Per's acknowledgement *en* 'mm' registering the receipt of the parenthetical insertion (line 12), Min returns to and resumes the host TCU (Jefferson 1972: 320) by repeating the topic *xianchang qifen* 'the atmosphere' and continuing with the comment of the sentence in line 13. This excerpt has shown that *yinwei*-clauses can be used as parentheticals inserted in the middle of a TCU.

4.2   Between TCUs

In this section, we demonstrate that *yinwei*-clauses may serve as parentheticals inserted between TCUs in a multi-TCU turn. This sequential position of *yinwei*-clauses can be seen in Excerpt (10). In this interaction, Jia, Wei, and Zhi are members of the Chinese graduate student association at the same university. Prior to this segment, they are talking about how to get funding for an event to be held by their association. In this sequence, Jia is calling a potential sponsor, a moving company, to get financial support to print the instructional pamphlet for new students.

(10)   Instructional pamphlet

15 Jia: 另　　外　一个　　　保证就是　说　　我们只　放　你　的.
        lingwai yi ge baozheng jiu shi shuo women zhi fang ni de.
        another one CL promise just be say we   just put you POSS
        'In addition, I promise that we just put yours,'

16   就　是　搬　　家　　的　我 我们　只　放　你.
        jiu shi ban jia   de wo women zhi fang ni.
        just be move house ASSC I   we   just put you
        'that is, we just put your advertisement.'

17   就 是 不 会 有 任何 的　竞争　　在 我们　这 个 新
        jiu shi bu hui you renhe de jingzheng zai women zhe ge xing
        just be NEG can have any   ASSC acompete at we   this CL new
        生　　　手册　　里.
        sheng shouce   li.
        student pamphlet in
        'That is to say, there is no competition in this Instructional pamphlet.'

18   然后 这 个 手册　　呢(.) 还(.) 还 包　　括 了 很
        ranhou zhe ge shouce   ne(.) hai(.) hai baokuo le hen duo
        then   this CL pamphlet PRT(.) also(.) also include PFV very many
        多　　　方面.
        fangmian.
        aspect
        'This pamphlet also includes many aspects.'

19 →   因为   我 是 (.) 也 是 编辑 之 一   嘛.
          yinwei   wo shi (.) ye   shi bianji zhi yi   ma.
          because I   be (.) also be editor of   one PRT
          'Because I'm also one of the editors (of the pamphlet).'

20   然后 我们 都 是 关于   爱 城 (.)   爱(.)
          ranhou women dou shi guanyu ai cheng(.) ai (.)
          then   we   all be about   ai city (.) ai (.)
          'We are all about,'

          埃德蒙顿   的   吃 喝   玩儿 乐 啊-
          aidemengdun de   chi he   wanr le   a-
          Edmonton   ASSC eat drink play fun PRT
          'Edmonton's food, drinks, and entertainment.'

21   衣   食   住 (.) 住 行   (.) 全部   都 有.
          yi   shi   zhu (.) zhu xing   (.) quanbu dou you.
          clothes food live (.) live travel (.) entire   all   have
          'The basic necessities of life are all included.'

In persuading the moving company to provide funding for the production of their pamphlets, Jia guarantees that they would be the only moving company advertised in the pamphlet (lines 15–17), and introduces the other contents included in the pamphlet (line 18). The morphosyntactic design of Jia's TCU in line 18 *baokuo le henduo fangmian* 'includes many aspects' strongly projects the incipient listing of these "many aspects" in the next TCU. However, instead of elaborating on those aspects, Jia continues her turn with an account of her position as the editor of the pamphlet prefaced with *yinwei* (line 19). This yinwei-clause on her authority in designing the content of the pamphlet provides relevant background information to explain why she has the knowledge of the content of the pamphlet. After inserting this background information, Jia returns to the prior TCU and resumes the just-now projected but suspended turn trajectory by repeating the same turn-initial particle *ranhou* 'then' in line 20 (same as that in line 18) and elaborating on the specific "aspects" included in the pamphlet in lines 20–21. Here, the *yinwei*-clause impedes the projected turn trajectory between two complete TCUs. It also functions as a parenthetical insertion that provides background information relevant to the ongoing turn.

In this section, we have discussed the interactional function of *yinwei*-clauses as parentheticals providing background information. These *yinwei*-clauses occur in two different sequential positions. One is within an ongoing TCU (Excerpt (9)), and the other is inserted between two TCUs in a multi-TCU turn (Excerpt (10)).

5. Conclusions

A number of these studies have examined the use of *yinwei* as a conjunction introducing reason or cause (Chao 1968; Lü 1980; Li&Thompson 1981). Yet few have examined how conversational participants use *yinwei* in the real-time emergence of talk, and how *yinwei*-clauses serve as resources for the construction of turns-at-talk and for the implementation of actions in interaction. Through a detailed examination of the video recordings of real spontaneous everyday Mandarin conversational data, this study has shown how *yinwei* is *actually* used in unscripted everyday Mandarin conversation. Specifically, we have observed that *yinwei*-clauses have two main interactional functions: accounts for a prior action, and parenthetical insertion introducing background information. First, *yinwei*-clauses serve as accounts for a speaker's prior actions such as disagreement and strong assertion. One commonality of disagreement and strong assertion is that they are actions that are accountable. Participants methodically deploy *yinwei* to introduce accounts for those types of actions in our data. Second, *yinwei*-clauses are also used as parentheticals inserted in the middle of a TCU or between two TCUs of a multi-TCU turn. These *yinwei*-prefaced parentheticals usually introduce background information that is related to the ongoing talk. When used in this particular sequential position, *yinwei* is used as a device to justify the inserted utterances being "displaced" *here and now* that impede the progressivity of the TCU- and turn-in-progress. It displays the speaker's orientation to the relevance and salience of the *yinwei*-prefaced utterance in the ongoing talk and interaction.

The interactional functions of *yinwei* reported in this study provide further evidence for the symbiotic relationship between grammar and interaction (Schegloff 1996; Couper-Kuhlen & Selting 2001). That is, grammar is shaped by and itself shapes conversational interaction (Couper-Kuhlen & Selting 2001; Thompson, Fox & Couper-Kuhlen 2015). The two interactional functions of *yinwei* described here emerge from its actual use in Mandarin everyday conversation that is undocumented in traditional Chinese grammar.

Finally, it should be acknowledged that there is a limitation of our data. As is noted earlier, 12 of the 26 participants had studied in Germany and Canada for 1 to 3 years when being recorded. It is possible that their usage of *yinwei* may be influenced by *because* in English and *weil* 'because' in German, though there is no empirical evidence to support or refute this possibility. The analytical claims and categories reported in this paper are all based on our current data. Future research may be conducted on the interactional functions of *yinwei*-clauses based on everyday Mandarin conversation recorded among participants without any foreign language backgrounds in China. In addition, studies comparing the interactional uses of *yinwei* and the uses of other frequently used conjunctions in Mandarin

conversation (such as *suoy*i 'so', *keshi* 'but', *ranho*u 'then', etc.) may yield new insights into the emergent interactional functions of conjunctions in Mandarin conversation in general.

References

Biq, Yung-O. 1995. "Chinese causal sequencing and yinwei in conversation and press reportage. Berkeley Linguist." *The Proceedings of the 21st Annual Meeting of the Berkley Linguistics Society*, 47–59. Berkeley, CA: Berkeley Linguistic Society.

Chao, Yuen-Ren. 1968. *A Grammar of Spoken Chinese*. University of California Press, Berkeley.

Chang, Miao-Hsia and Hsin-Yi Su. 2012. "To mark or not to mark the cause, that is the question: Causal marking in Taiwanese conversation" *Journal of Pragmatics* 44(13): 1743–1763.

Chen, Ping. 1999. *Modern Chinese: History and Sociolinguistics*. Cambridge University Press, Cambridge, UK.

Couper-Kuhlen, Elizabeth. 2009. "On combining clauses and actions in interaction". *Virittäjä*, (Journal of the Society for the Study of Finnish, Kotikielen Seuran aikakauslehti, 113.)

Couper-Kuhlen, Elizabeth. 2011. "When Turns Start with Because: An Exercise in Interactional Syntax". In *Connectives in Synchrony and Diachrony in European Languages, Varieng: Studies in Variation, Contacts and Change in English*, Volume 8, ed. by Anneli Meurman-Solin, and Ursula Lenker.

Couper-Kuhlen, Elizabeth, and Margret Selting. 2001. "Introducing Interactional Linguistics." In *Studies in Interactional Linguistics*, ed. by Margret Selting, and Elizabeth Couper-Kuhlen, 1–22. Amsterdam: John Benjamins.

Duvallon, Outi, and Sara Routarinne. 2005. "Parenthesis as a Resource in the Grammar of Conversation." In *Syntax and Lexis in Conversation: Studies on the Use of Linguistic Resources in Talk-in-Interaction*, ed. by Auli Hakulinen, and Margret Selting, 45–74. Amsterdam: John Benjamins.

Fang, Mei. 2000. "Ziran kouyu zhong ruohua lianci de huayu biaoji gongneng [Reduced conjunctions as discourse markers]." *Zhongguo Yuwen* 5: 459–470.

Guo, Yulian. 2012. "Huayu biaojiyu 'shi zheyang de' ji qi yuyong gongneng [Discourse marker '…is like this' and its pragmatic function]." *Yuyan Yanjiu* 2: 118–123.

Heritage, John. 1984. "A change-of-state token and aspects of its sequential placement". *Structures of social action: Studies in Conversation Analysis*, eds. by Maxwell Atkinson and John Heritage, 299–345. Cambridge, U.K.: Cambridge University Press.

Heritage, John. 1988. "Explanations as accounts: A conversation analytic perspective". *Analysing Everyday Explanation. A Casebook of Methods*, ed. by Charles Antaki, 127–144. London: Sage.

Heritage, John. 1989. "Current Developments in Conversation Analysis." In *Conversation: An Interdisciplinary Perspective*, ed. by Derek Roger, and Peter Bull, 21–47. Clevedon: Multilingual Matters.

Hsieh, Hsin-yun. 1999. "Coherence via Collaboration: A Study of Chinese Causal Connectives." *Working Papers in Linguistics*, 2: 157–186.

Hutchby, Ian and Robbin Wooffitt. 2008. *Conversation Analysis*. Cambridge: Polity.

Jefferson, Gail. 1972. "Side Sequences." In *Studies in Social Interaction*, ed. by David N. Sudnow, 294–233. New York, NY: Free Press.

Jefferson, Gail. 1984. "On the Organization of Laughter in Talk about Troubles." In *Structures of Social Action*, ed. by Maxwell J. Atkinson, and John Heritage, 346–369. Cambridge: Cambridge University Press.

Kirkpatrick, Andy. 1993. "Information Sequencing in Modern Standard Chinese in a Genre of Extended Spoken Discourse." *Text* 13(3): 423–453.

Li, Charles N. and Sandra A. Thompson. 1976. Subject and Topic: A new typology of language. In *Subject and Topic*, eds. by Charles N. Li, and Sandra A. Thompson, 457–489. New York: Academic Press.

Li, Charles N. and Sandra A. Thompson. 1981. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley, California: University of California Press.

Li, Xiaoting. 2016. "Some Discourse-Interactional Uses of Yinwei 'Because' and Its Multimodal Production in Mandarin Conversation." *Language Sciences* 58: 51–78.

Lü, Shuxiang. 1982. *Zhongguo Wenfa Yaolue* [Essentials of Chinese Grammar]. Beijing: Shangwu Yinshuguan.

Lü, Shuxiang. (eds). 1999. *Xiandai Hanyu Babai Ci* [800 Words in Contemporary Chinese]. Beijing: Shangwu Yinshuguan.

Pomerantz, Anita. 1984. "Agreeing and Disagreeing with Assessments: Some Features of Preferred and Dispreferred Turn Shapes." In *Structures of Social Action*, ed. by Maxwell J. Atkinson, and John Heritage, 57–101. Cambridge: Cambridge University Press.

Sacks, Harvey. 1992. *Lectures on Conversation*. Oxford: Blackwell.

Schegloff, Emanuel A. 1996. "Turn Organization: One Intersection of Grammar and Interaction." In *Interaction and Grammar*, ed. by Elinor Ochs, Emaneul A. Schegloff, and Sandra A. Thompson, 52–133. Cambridge: Cambridge University Press.

Schegloff, Emanuel A. 2007. *Sequence Organization in Interaction*. Cambridge: Cambridge University Press.

Selting, Margret, Peter Auer, Dagmar Barth-Weingarten, Jörg Bergmann, Pia Bergmann, Karin Birkner, Elizabeth Couper-Kuhlen, Arnulf Deppermann, Peter Gilles, Susanne Günthner, Martin Hartung, Friederike Kern, Christine Mertzlufft, Christian Meyer, Miriam Morek, Frank Oberzaucher, Jörg Peters, Uta Quasthoff, Wilfried Schütte, Anja Stukenbrock, Susanne Uhmann1. 2009. "Gesprächsanalytisches Transkriptionssystem 2 (GAT 2)." *Gesprächsforschung-Online-Zeitschrift zur verbalen Interaktion* 10: 353–402.

Song, Zuoyan, and Hongyin Tao. 2008. "Hanyu yinguo fuju shunxu de huayu fenxi yu bijiao [Discourse analysis and comparison of Chinese and English causal clauses]." *Hanyu Xuebao* 4: 61–71.

Song, Zuoyan, and Hongyin Tao. 2009. "A Unified Account of Causal Clause Sequences in Mandarin Chinese and Its Implications." *Studies in Language* 33 (1): 69–102.

Thompson, Sandra A., Barbara A. Fox, and Elizabeth Couper-Kuhlen. 2015. *Grammar in Everyday Talk: Building Responsive Actions*. Cambridge: Cambridge University Press.

Wang, Yu-Fang. 2002. "The Preferred Information Sequences of Adverbial Linking in Mandarin Chinese Discourse." *Text* 22(1):141–172.

Wang, Yu-Fang. 1999. "The Information Sequence of Adverbial Clauses in Mandarin Chinese Conversation." *Journal of Chinese Linguistics* 27(2): 45–89.

Xing, Fuyi, 2001. *Hanyu Fuju Yanjiu* [A Study of Chinese Complex Sentences]. Beijing: Shangwu Yinshuguan.

Young, Linda W. 1982. "Inscrutability Revisited." In *Language and Social Identity*, ed. by John J. Gumperz, 72–84. Cambridge: Cambridge University Press.

Appendix

The transcription conventions of GAT-2 (Gesprächsanalytisches Transkriptionssystem 2) (Selting et al. 2009) used in this article:

(.)	micro-pause
(-), (–), (—)	short, middle or long pauses
(1.0)	pauses which are or more than 1.0 second
ACcent	primary or main accent
::	prolongation or stretching of the sound just preceding them
<<coughing> >	accompanying paralinguistic and non-linguistic actions over a stretch of speech
<<f>>	forte, loud
<<p>>	piano, soft
((head nods))	description of an action
,	rising to mid
–	level final pitch movement
;	falling to mid final pitch movement
.	falling to low final pitch movement

Glossing conventions

ASSC	associative	PFV	perfective aspect
CL	classifier	PROS	progressive aspect
CRS	currently relevant state	PRT	particle
EXP	experiential marker	Q	question marker
NEG	negatives		

Preliminaries to delicate matters

Some functions of "I say to you" sequences in Mandarin Chinese conversations

Ni-Eng Lim

Nanyang Technological University, Singapore

The meta-language unit "I say to you" is frequently heard in Mandarin Chinese conversations, and are most commonly expressed as *wo gen ni shuo* 'I say to you', *wo gen ni jiang* 'I talk to you', or *wo gaosu ni* 'I tell you', collectively termed "I-say-to-you" expressions. Quantitative investigations reveal that they are dedicated interactional resources found only in spoken conversation. By using conversation analytic methodology, further examination of their sequential trajectory shows that a core function of "I-say-to-you" expressions is to preface upcoming "delicate" matters, such as dispreferred next action, disagreement or disaffiliative turn, and other actions that may be resistance-implicative for the recipient. As a preface, "I-say-to-you" expressions can be used by the speaker to secure multi-turns space with which to gradually deliver the "delicate" matter and achieve other interactional goals.

Keywords: conversation analysis, Mandarin Chinese, meta-language units, interactional linguistics, preface

1. Introduction

This paper presents an initial report on a particular group of Mandarin Chinese (henceforth 'Chinese') expressions that literally means "I say to you". The most commonly heard forms of these expressions in Chinese conversation are namely *wo gen ni shuo* 'I say to you', *wo gen ni jiang* 'I talk to you', or *wo gaosu ni* 'I tell you'; which I will hence collectively refer to as "I-say-to-you" expressions. Within corpus linguistics, these "I-say-to-you" expressions can be analyzed as lexical bundle(s), as they are sequences of words that occur frequently in chunks/clusters, but are non-idiomatic in nature (Biber & Barbieri 2007; Wray 2005; O'Keeffe et al. 2007). According to Biber (2009) and Conrad & Biber (2004), more than 50% of English

lexical bundles in conversation begin with a personal pronoun + verb phrase (e.g. *I don't know why, I thought that was*), and serve as interpretive 'discourse frames' for expressing stance and/or organizing discourse (pp. 285). Although Biber (2009) functionally classified these bundles as stance expressions or discourse organizers based on "linguistic constructs" (pp. 286); what these "constructs" are, or how they are applied to reveal the workings of lexical bundles, are left unexplained. This study partially addresses this gap by using the methodology of conversation analysis to clearly delineate how the Chinese lexical bundles of *wo gen ni shuo, wo gen ni jiang* and *wo gaosu ni* are used to organize upcoming trajectories of talk in everyday interactions.

Although lexical bundles clearly play a key role in everyday spoken discourse, studies on the discourse-interactional functions of core Chinese lexical bundles are basically non-existent.[1] From these limited studies, Tao (2011), in an investigation of core vocabulary in spoken Chinese, noted the puzzle of how certain lexical categories (such as the speech act verb *shuo* 'say' and preposition *gen* 'to') come to occur at such high frequency, in comparison to other more intuitively interactional elements, such as person pronouns (*wo* 'I', *ni* 'you', or *ta* 'he'). The explanation he proposed for the high frequency of certain less intuitive lexical items is that they arise from how native speakers utilize multi-word units of fixed or formulaic expression (i.e. lexical bundles) that convey specialized discourse-pragmatic meanings in their natural conversation. In an earlier paper on Chinese lexical bundles, I looked at the use of another highly frequent "personal pronoun + verb phrase" unit, namely *wo juede* 'I think', and found that Chinese speakers regularly deploy *wo juede* to frame an initial assessment, especially when such assessments are potentially disaffiliative to the recipient's own opinions. One advantage of such an interactional practice is that the speaker can work to elicit the recipient's assessment in the next turn, while at the same time hedging any possible inapropos elements in his/her own assessment. In this sense, *wo juede* can then function as a "joint-assessment initiator" to 'check the bath-water' (i.e. minimize the cost of proffering an opinion), which in turn has an interactional consequence on the trajectory of talk (Lim 2011). I shall continue with this line of research in the current paper by explicating the interactional practices of Chinese "I-say-to-you" expressions and how they might be used to achieve certain interactional goals.

The following sections are organized as follow: Section 2 briefly describes the database and the method used for analysis. Section 3 provides some statistics on

---

1. While there is obviously a huge range of studies on Chinese discourse markers (Biq 2001, 2007; Feng 2008; Wang et al. 2010; Wu 2014; Li 2016) based on different approaches, these tend to be focused on the lexical-semantics of single lexical items, and not on how clusters of high-frequency words (i.e. lexical bundle) function.

these "I-say-to-you" bundles and establishes their predominant role in conversation. Section 4 is then divided into two sub-sections, where sub-section 4.1 argues for "I-say-to-you" as a preface for delicate actions, and sub-section 4.2 outlines how multiple turns are taken to deliver such actions. Section 5 provides two additional cases where "I-say-to-you" are used to achieve certain interactional goals. The final section (Section 6) summarizes the findings with some comments on commonalities and differences in turn-initial lexical bundles.

2. Data and methodology

The data for this study is from the CallFriend Mandarin Corpus (henceforth CallFriend), distributed by the Language Data Consortium (LDC). This corpus consists of 60 unscripted telephone conversations, with approximately 200,000 characters in spoken data, lasting between 5 and 30 minutes (Canavan & Zipperlen 1996). Participants were recruited by LDC to make a free long-distance telephone call, therefore all calls originated inside the United States and Canada. Once the caller agrees to participate, they were free to call whoever they like, and the call was allowed only if both caller and callee agreed to be recorded. Hence most calls were between family members or close friends, and both the caller and callee in the corpus are native speakers of Mandarin Chinese from Mainland China.

Although the interlocutors were aware that their conversation will be recorded, they were given no instructions on what to talk about. Common topics range from updates on the status of family members, talking about difficult situations, to gossiping about acquaintances. While the calls were originally collected for the purpose of building a spoken database for corpus linguistics, the original audio recordings of these calls were recently released by LDC, thereby also making it a viable source of data for conversation analytic studies.

For comparison with the Chinese written registers, I have also utilized the Lancaster Corpus of Mandarin Chinese (LCMC), with approximately 1 million characters in written data, distributed by the European Language Resources Association and Oxford Text Archive (McEnery & Xiao 2004). The LCMC contains 15 different written genres including but not limited to press reports, newspaper editorials, newspaper reviews, religious texts, fiction, academic prose, skills, and trades & hobbies. Token instances of *wo gen ni shuo, wo gen ni jiang, wo gaosu ni*, and other possible variants within CallFriend and LCMC were first identified using the corpus analysis software Wordsmith (Scott 2016), and their statistical numbers compared. Relevant conversational sequences in which the "I say to you" bundles are embedded (within the CallFriend corpus) were then extracted and transcribed in details for further examination.

As mentioned, the methodological lens through which I qualitatively examine these "I-say-to-you" bundles is Conversation Analysis or 'CA' for short. Developed by Harvey Sacks, Emmanuel Schegloff and Gail Jefferson to study the social organization of naturally occurring talk-in-interaction, CA has proven itself to be a useful tool in addressing various questions from other disciplines, such as anthropology, psychology, education and linguistics. Central to CA's versatility is its premise that talk-in-interaction is the "primordial site for human sociality" (Schegloff 1987), and that the social organization of any institutional enterprise is first and foremost instantiated through the turn-by-turn production of meaningful actions (Heritage 2013b). With this analytic framework that focuses on the description of sequential moves of participants in particularized social contexts (e.g. classroom discourse, archeological excavation, medical consultation etc.), CA was able to cut across traditional boundaries of the social sciences. Heavily influenced by CA, the field of interactional linguistics has also been recently developed, with the "premise that language should not be analyzed in terms of context-free linguistic structures but as a resource for the accomplishment of actions in social interaction" (Kern & Selting 2013). Using this analytic approach, the discourse-interactional functions of Chinese "I-say-to-you" expressions are examined within the sequential context of their occurrence and their resulting actions.

3.   Some preliminary quantitative results

Before turning towards the detailed qualitative exploration of a few "I-say-to-you" sequences, quantitative results from a comparative analysis of both spoken and written corpora is presented here. As the study aims to investigate how and why Chinese speakers might utilize a meta-language unit such as "I-say-to-you", it would be instructive to consider the different permutations of such constructions available to the Chinese speaker. There are many words that are nearly synonymous to "say" in Chinese. Intuitively, the most common would be *shuo* 'say', *jiang* 'talk' and *gaosu* 'tell'. Therefore, as mentioned, "I say to you" can commonly be expressed as *wo gen ni shuo* 'I say to you', *wo gen ni jiang* 'I talk to you', or *wo gaosu ni* 'I tell you' in Chinese conversation. Alternatively, the agentive subject *wo* 'I' can be flipped to *ni* 'you', and using the verb *ting* 'listen', result in a similar meta-language unit that index the recipient's perspective (i.e. *ni ting wo jiang* 'you listen to me talk' and *ni ting wo shuo* 'you listen to me say'). Furthermore, as Chinese is a pro-drop language (Li & Thompson 1989: 657–662), the general meaning of "I say to you" can also be achieved by dropping the initial pronominal subject in the different types of "I say to you" constructions above. For instance, instead of saying

wo gen ni shuo 'I say to you', in cases where the reference of the speaker is clear, *gen ni shuo* 'say to you' would also suffice.

Using the corpus analysis software Wordsmith (Scott 2016), instances of the above different permutations of "I-say-to-you" constructions were each identified within both the spoken (Callfriend) and written (LCMC) corpus, and their tabulated raw frequency counts are shown in Table 1.

Table 1. Raw frequency counts of "I-say-to-you" constructions in spoken and written corpora

Versions of "I say to you"			Spoken corpus (200,000 characters)	Written corpus (1,000,000 characters)
我跟你讲	*wo gen ni jiang*	'I talk to you'	61	0
我跟你说	*wo gen ni shuo*	'I say to you'	54	2*
我告诉你	*wo gaosu ni*	'I tell you'	37	1*
你听我讲	*ni gen wo jiang*	'you listen to me talk'	3	0
你听我说	*ni gen wo shuo*	'you listen to me say'	5	1*
Pro-drop versions of "I say to you"				
Ø 跟你讲	*gen ni jiang*	'talk to you'	1	0
Ø 跟你说	*gen ni shuo*	'say to you'	10	0
Ø 告诉你	*gaosu ni*	'tell you'	3	1*
Ø 听我讲	*ting wo jiang*	'listen to me talk'	0	0
Ø 听我说	*ting wo shuo*	'listen to me say'	0	0
TOTAL			174	5

* Instances appear within quotations in works of fiction

Two evident and significant results can be gleaned from the table above. The first, and perhaps most important, observation is that the various versions of "I-say-to-you" are units exclusively found in spoken interaction. Although the spoken corpus is only about one-fifth the size of the written corpus, the overall frequency of "I-say-to-you" constructions found in CallFriend significantly outweighs instances found in LCMC. If we normalized the results (normalized to 200,000 characters for both corpora), then the ratio of "I-say-to-you" constructions in the spoken corpus to written corpus is about 174 to 1. In fact, when we look closely at the 5 instances of "I-say-to-you" in LCMC, it turns out that they are all found within quotations in fictional works, whose purpose is ostensibly to mimic actual spoken interaction. Hence, there is strong quantitative evidence that these "I-say-to-you" constructions are basically non-existent in written Chinese. On the other hand, they are frequently found in dialogic conversational data as demonstrated

by their high frequency in the spoken corpus. This suggests that "I-say-to-you" is a dedicated interactional resource for addressing conversational issues not found in monologic written language.

A second key observation is that the frequency of different permutations of "I-say-to-you" constructions is significantly disparate. Looking first at the pro-drop versions of the "I-say-to-you" constructions, it is found that neither the spoken nor the written corpus had any instances of *ting wo jiang* 'listen to me talk' or *ting wo shuo* 'listen to me say'. The instances of pro-dropped *gen ni jiang* 'talk to you', *gen ni shuo* 'say to you' and *gaosu ni* 'tell you' in the spoken corpus are also very limited (14 instances in total), relative to their full version with the pronominal subject *wo* 'I' attached (152 instances in total). Within the full versions, the frequency of using *ni* 'you' as the agentive subject (i.e. *ni ting wo jiang* 'you listen to me talk' and *ni ting wo shuo* 'you listen to me say') are also rare, with only 8 instances in the spoken corpus.

In sum, the conveyance of the meta-language expression "I-say-to-you" is distributionally skewed towards three types of constructions which are used exclusively within talk-in-interaction; namely *wo gen ni shuo* 'I say to you' (with 54 instances), *wo gen ni jiang* 'I talk to you' (with 61 instances), and *wo gaosu ni* 'I tell you' (with 37 instances), totalling 152 instances. The clustering of "I-say-to-you" constructions into three specific types further strengthens the argument that these "I-say-to-you" expressions are fixed conversational "chunks" with distinct discourse-interactional functionality. The puzzle here is, of course, why would Chinese speakers frequently use meta-language units such as "I-say-to-you" in conversation to announce beforehand the intention of saying something to another, as opposed to simply saying what one wants to say? What sort of pragmatic functions might they serve in conversation?

4.   Analysis of some 'I-say-to-you' sequences

It is interesting to note that although the translated equivalent of "I say to you" and "I talk to you" are not common discourse expressions found in English conversation, "I tell you" or "Listen up!" on the other hand are intuitively more frequent in sequences where the speaker attempts to negotiate the recipient's opinion to align with his/her own. Drawing on another parallel, translation of the Hebrew Bible (English Standard Version – ESV) are also littered with "I-say-to-you" expressions when it is being ascribed to sayings by Jesus to his disciples, such as "*But **I say to you**, Love your enemies and pray for those who persecute you* (Matthew 5:44 (ESV))" or "*And **I tell you**, ask, and it will be given to you; seek, and you will find; knock, and it will be opened to you* (Luke 11:9 (ESV))". Hence, there seems

to be some cross-linguistic evidence that suggest meta-language expressions of "I-say-to-you" are also common discourse chunks in other languages, and that they occur in situations where speakers are concerned with evoking some sort of hierarchical or authoritative stance.

Indeed, a large majority of the 152 tokens of "I-say-to-you" expressions in the corpus were spoken either by an elder or a parent (i.e. persons of authority) to someone ranked lower along a hierarchy (i.e. own children or other junior members in the family). While there were also a few instances where speakers use "I-say-to-you" towards a peer (such as a friend or colleague) or even towards persons of higher hierarchical status (e.g. their parents), there is still a sense that these speakers were positioning themselves as having certain epistemic or deontic authority over their recipients.

It is difficult, however, to say with certainty if "I-say-to-you" expressions function to evoke a stance of authority for their recipients. For one, while a large majority of the 152 tokens of "I-say-to-you" expressions in the corpus were indeed spoken either by an elder or a parent (i.e. persons of authority) to someone ranked lower along a hierarchy (i.e. own children or other junior members in the family), this could very well be due to a self-selection bias inherent in the collection of the spoken corpus, where enlisted overseas Chinese students were pre-disposed to make long-distance telephone calls back home; thereby extending the opportunity for their concerned parents to issue various directives to them, while at the same time the students would be culturally primed not to express disagreement or issue directives towards their parents. Secondly, the very fact that some "I-say-to-you" expressions in the corpus are also used by peers (such as a friend or colleague) towards each other, or even by children toward their parents, highlights that these expressions are not exclusively used by person(s) of a higher hierarchical status relative to their recipient(s), and therefore an analysis that "I-say-to-you" expressions are simply discourse markers of authority would be inadequate, or an over-generalization at best. Furthermore, if we took the statistical axiom "correlation does not imply causation" seriously, it is *not* reasonable to argue "I-say-to-you" expressions function primarily to denote authority simply based on statistical counts of the analyst's perceived power relations of the interlocutors. The predisposition for persons of higher social status or authority to use "I-say-to-you" could simply stem from a perception of greater entitlement to issue actions that impose on or disaffiliate from their interlocutor (Curl & Drew 2008; Craven & Potter 2010). That is to say, even if there is a strong quantitative correlation between "I-say-to-you" expressions and speakers in position of power over their interlocutor, this may simply be a by-product of the discourse-interactional functions of "I-say-to-you" expressions to preface disaffiliative actions, and not that "I-say-to-you" expressions are markers of authority and status in of themselves. Hence it might be

more fruitful to refrain from positing hypothesized attributes based on statistical correlations, and focused on the empirically deductible sequential actions resulting from "I-say-to-you" expressions in actual contexts.[2]

From the 152 instances of "I-say-to-you" expressions consisting of three types of construction (i.e. *wo gen ni jiang* 'I talk to you', *wo gen ni shuo* 'I say to you', and *wo gaosu ni* 'I tell you'), some common characteristics in their sequential context and trajectory of talk can be observed. A detailed qualitative analysis of these "I say to you" sequences will demonstrate what such sequences are, and how they can be effectively utilised to achieve certain interactional goals.

4.1 Prefacing delicate matters with 'I say to you'

I shall first outline these characteristics using extract Example (1) below, in which *wo gen ni jiang* is being used by the speaker in a prototypical "I-say-to-you" sequence of talk. It shall be shown how "I-say-to-you" expressions in many ways act like "preliminaries" (i.e. prefaces) to some projected "delicate" action. The extract records a telephone conversation between a son and her mother, where the mother first probes the progress of her son's doctoral studies, followed by giving unsolicited advice on how he should manage his time. I shall describe the sequence in two parts. The first part is labeled Example (1a), and documents the sequential context until right before Mom issues her *wo gen ni jiang* in focal line 29. The second part labeled Example (1b) then begins with line 29, and shows the sequence of talk after the focal *wo gen ni jiang* turn is being issued. In the first part of the sequence, Mom probes about a meeting the Son had with his PhD supervisor.

Example (1a)   Wo gen ni jiang [33_3.44]

01   MOM: 诶 你 你 你 那个: 见到:: 导师　　了 吗.
             ei ni ni ni nage: jianda::o daoshi   le   ma.
             SIP 2SG 2SG 2SG that.is meet   supervisor CRS Q
             'Erm… you've… that is… have you seen your supervisor yet?'

---

2. A reviewer pointed out that "I-say-to-you" expressions could be manifested in more specific ways, in which prefacing "delicate" actions and evoking authority are tied together in a non-contradictory manner. For instance, evoking authoritative status with "I-say-to-you" expressions could provide legitimacy for the speaker who intends to carry out the upcoming disaffiliative "delicate" action. My refrain from arguing this position stems from a principled adherence to what I can confidently demonstrate based on conversation analytic proof procedures; in other words, I am unable to rigorously verify that "I-say-to-you" expressions functions to evoke authority, save for the superficial statistical correlation between speakers and their higher hierarchical status. However, I agree that this is the intuitive sense that is deeply held by native speakers of Chinese, and may in the end, be correct.

02   SON: 呃   见着   了.
              e   jianzhao le.
              CFM see.CTP   CRS
              'uh-hmm, I've seen him.'

03   (0.6)

04   SON: 啊-=
              a-=
              SIP
              'erm…'

05   MOM: =啊=
              =a=
              CFM
              'uh huh'

06   SON: >回来   以后< 可能:: .hhh 马上   就   让   我
              >huilai   yihou< kene::ng .hhh mashang   jiu   rang wo
              come.back after   maybe   immediate then let   1SG
              'After I return, maybe… erm… he wants me to immediately…'

07   写   一个   就是::: 毕::   (.) 毕业   之前   的: 时间
              xie   yige   jiushi::: bi::   (.) biye   zhiqian de: shijian (0.3)
              write one.CLF that.is graduate   graduate before GEN time   (0.3)
              'write a… that is… the time remaining before graduation is due…'

08   怎么 安排. 我 写完   了.
              zenme anpai.   wo xiewan   le.
              how   organize 1SG write.complete CRS
              'how I'm going to plan my time. I've finished writing it.'

09   MOM: 喔::.
              o::.
              CFM
              'Oh…'

10   (.)

11   SON: 他 [说-
              ta [shu-
              3SG say
              'He said…'

12   MOM: [计划.
                  [jihua.
                  plan
                  'A schedule.'

13   SON: 哎:. 他 说 你 就 [按]   你 的 计划 (.) 做.=
              ai:. ta   shuo ni   jiu [an]   ni   de jihua (.) zuo.=
              AGR 3SG say 2SG just follow 2SG GEN plan   do
              'Yes. He said you should just follow your planned schedule.'

14   MOM: [哦.]
                                                   [*o.*]
                                                   CFM
                                                   'uh-hm'

15   SON: =他 说   反而:::=
              =*ta shuo faner:::*=
              3SG say   anyway
              'He said anyway…'

16   MOM: =哦.
              =*o.*
              CFM
              'oh'

17   (0.4)

18   SON: 你 应-   你: 就 应该   特别   (.) 努力   的.
              ni yin-   ni: jiu yinggai tebie   (.) *nuli   de.*
              2SG should 2SG just should especially   hardworking GEN
              'You should… you should work extra hard on…'

19   就是 这 段   时间.
              jiushi zhe duan shijian.
              that.is DEM CLF   time
              'that is during this period of time.'

20   (0.5)

21   MOM: 哦::.
              o::.
              CFM
              'oh'

22   SON: 嗯:: [挺   关键   的] 就是 说   你 呢: (.) =
              e::n [*ting   guanjian de*] *jiushi shuo ni   ne:* (.) =
              AGR   quite crucial   GEN that.is say   2SG TP
              'Yes. It's quite crucial, that is (he said) you should'

23   MOM: [呃 呃   ]
                  [*e   e*   ]
                      CFM CFM
                      'uh-hmmm'

24   SON: =努力   [把-
              =*nuli*   [*ba-*
              hardworking BA
              'work hard on…'

25   MOM: [对 对 对.
                             [*dui dui dui.*
                              yes yes yes
                              'That's right.'

26   SON: (这 些) 论文-=
                   (*zhe xie*) *lunwen-*=
                   DEM CLF dissertation
                   'this dissertation'
27   MOM: =对. <是 [关键.
                   =*dui.* <*shi* [*guanjian.*
                   yes   COP   crucial
                   'Yes. It is indeed crucial.'
28   SON: [(>这样<)   拿   下来.
                          [(>*zheyang*<) *na   xialai.*
                              like.this   take down.come
                              'to complete it.'

From the above, it can be seen that Mom is treading gingerly on the issue of what her son's doctoral supervisor had advised him to do. It is not entirely clear why such an issue might be sensitive for Mom to broach, but the design of her turns at talk clearly demonstrates her apprehensive-ness in advancing such a topic. First of all, Mom's question in line 01 is delivered with hearable difficulty, initially with a particle *ei* that index her upcoming talk as being unplanned (Tsai 2008) or in a "by the way" manner; and then with multiple restarts of *ni* 'you' and a repair initiator *nage* 'that is', displaying her trouble with finding an appropriate formulation for the action she wants to pursue. More significant, however, is the manner Mom reacts to her Son's responses. While line 02 is clearly a preferred 'yes' response to Mom's 'yes-no' interrogative at line 01, such a "type-conforming response" (Raymond 2003) is not subsequently treated as being adequate by Mom. This is seen firstly in the gap at line 03 where Mom could have picked up the turn but did not. When her son begins to provide more information on what has transpired in his meeting with the supervisor at lines 06–08, Mom then continually provides continuers (either *o* (哦／喔) or *e* (呃)) at line 09, 14, 16, 21, and 23; thereby prompting further explication from the Son at line 13, lines 18–19, and around lines 22–28. The point here is, Mom's line 01 is, in fact, a sort of "pre-request" (Schegloff 1980) for more details of the meeting. Although line 01 begins with a simple 'yes-no' interrogative, Mom's subsequent treatment of her son's responses demonstrates her intended trajectory of talk. Though such an overt request did not actually materialize, the recipient (i.e. her son) has nonetheless oriented to Mom's treatment of his responses as signalling inadequacy and continues to provide details of the meeting on his own accord.

   Turning now towards the content of this sequential context, we see that the son first recounts his supervisor's instruction for him to plan and stick to a schedule right up until completion of his dissertation; and that he should work extra hard during this period of time. At line 22, the son then declares that such a course

of action is *ting guanjian de* 'quite crucial' for him to successfully complete his doc-
toral studies. Mom has been acknowledging her son's recount of the meeting with
multiple continuers, but without much of her own personal input, until line 25
and 27 where she aligns with her son's opinion by stating that it is 'indeed crucial'.

Before continuing with the analysis, I like to highlight two key takeaways on
the above sequence before the focal *wo gen ni jiang* turn is issued at line 29. The
first observation is Mom's 'pre-request' and her subsequent maneuvers of indirect
inquiry (i.e. providing continuers) amply demonstrate that she orients to enquir-
ing about her son's progress as an action this is not without contingencies, and
possibly one that is dispreferred. The second observation is that Mom has also
responded with a prima facie agreement at line 27 to her son's explicit opinion that
it is of crucial importance to plan and stick to a 'hardworking' schedule. These two
points will become relevant as I continue to explicate the trajectory of talk after *wo
gen ni jiang* is delivered in Example (1b) below.

Example (1b)   Wo gen ni jiang [33_3.44]

29 → MOM: 我 跟　 你 讲　 啊.=
          ***wo gen　 ni　 jiang** a.=*
          1SG follow 2SG say　 SFP
          'Hey, I say to you'

30 　 SON: =嗯.=
          =en=
          CFM
          'mm-hmm'

31 　 MOM: =那 你 计划 做 得　 不要　　 太 紧 了.=
          =na ni jihua zuo de　 bu yao　 tai jin le.=
          DEM 2SG plan do PCM NEG.want too tight CRS
          'Don't be too rigid in your scheduling.'

32 　 SON: =呃 [不　 算　 太-
          =eh [bu　 suan ta-
          AGR NEG count too
          'mmm it's not too…'

33 　 MOM: [要　 留　 点儿 余地.=
              [yao liu　 dianr yudi
              want stay bit　 leeway
              'you have to give some lee-way'

34 　 SON: =[就 这 半年,　] 诶,　　 半年. 任务　 量　 (0.2)
          =[jiu zhe bannian]　 ai　　 bannian renwu liang (0.2)
          just DEM half.year SIP half.year task　 amount
          'it's just for half a year, mmm… half a year, the workload… '

35	MOM:	= [是 不是 啊.]
		= [*shi bushi a.*]
		COP NEG.COP SFP
		'ain't I right?'
36	SON:	工作 量 特别 大.°反正.°
		gongzuo liang tebie da. °fanzheng. °
		work amount especially big anyway
		'it's heavy for half a year anyway'
37		(0.8)
38	MOM:	哦:::.
		o:::.
		CFM
		'oh…'

After the Son has stated the importance of working hard during the upcoming critical period, and Mom's prima facie agreement that such a course of action is indeed 'crucial', Mom then issues the focal *wo gen ni jiang* at line 29, with her son immediately acknowledging such a turn with the continuer *en* (similar to the English 'mm-hmm') at line 30. Though such an exchange may at first appear quite unremarkable, some technical points on this short paired sequence deserve our attention. I first note that the "I-say-to-you" construction constitutes the entire turn of talk (with the exception of the additional utterance-final particle *a* (啊) at its end). In other words, the "I-say-to-you" construction here is not used as a complement-taking predicate (Thompson & Mulac 1991; Thompson 2002) where "I-say-to-you" grammatically frame some other constituent (as in "*wo gen ni jiang* + complement"). Instead, the "I-say-to-you" construction is, in fact, an *independent unit of expression in actual use* (i.e. discourse chunk), constituting an entire turn-constructional unit (TCU) by itself.

Next, the son's subsequent continuer not only evinces the self-sufficiency of the "I-say-to-you" construction but furthermore acknowledges it as being a preamble to some upcoming action. Such a sequence is in many ways analogous to the interactional practice of 'preliminaries' or 'pre's' at the beginning of their sequence (Schegloff 1980). Schegloff similarly begins by examining the logical paradox of a meta-language turn such as "can I ask you a question?", and goes on to show how such 'pre's' are used to project a certain action (e.g. "Can I ask you a question?" projects an upcoming question). Here, from the son's continuer at line 30, it is evident that he also orients to *wo gen ni jiang* at line 29 as a preliminary to something.

But what sort of things are these "I-say-to-you" expressions 'pre's' or a preliminary to? As opposed to pre-questions, since "I-say-to-you" expressions literally just mean that the speaker is going to tell 'something', it is not overtly clear what sort of action or telling an "I-say-to-you" construction is projecting. An

examination of the trajectories of "I-say-to-you" sequences in the collection reveals that one common reason that speakers deploy "I-say-to-you" expressions is to indicate that *the upcoming action is in some way problematic for the recipient, thereby projecting some form of "delicate action"*. What I mean by "delicate action" in this paper include not only actions that are overtly face-threatening in of itself (such as criticisms, admonishments, or directives), but also actions that speakers can reasonably anticipate from the situational context to be misaligned or disaffiliative with what the recipients want. Interestingly, Schegloff (1980) also forward an analysis that raising 'pre-questions' (e.g. "Can I ask you a question?") can indicate to the hearer that the projected actual question is possibly a delicate one,[3] as "projecting the question does the work of displaying that a question hearable as subject to delicate treatment once heard, has been so treated on this occasion, by this speaker, for this recipient" (Schegloff 1980: 134).

In line 31, Mom's directive[4] is indeed one such delicate action. From an action standpoint, issuing a directive on how some others should plan their own schedule is by itself already in violation of negative face (despite Mom's higher hierarchical status relative to her son) (Brown & Levinson 1987). But more importantly, Mom's directive to refrain from planning a tight schedule goes directly against her son's earlier expressed wishes to plan a 'hardworking' schedule, as can be seen in the son's resistance at line 32. That is to say, despite an earlier orientation to interfering with her son's doctoral plans as being "dispreferred", Mom nonetheless prefaces with an "I-say-to-you" expression before issuing an action that she understands to be clearly disaffliative, and possibly resisted by her recipient.

This pattern of use is seen throughout my corpus of "I-say-to-you" expressions with some variations on how the "I-say-to-you" turn is formulated. A possible variation includes instances where there may not be a turn-final particle attached to the "I-say-to-you" expression. Other variations are especially revealing as they serve to further corroborate with my analysis of "I-say-to-you" expression as prefacing "delicate" matters. One such variation is the addition of an address term specifically addressing the recipient, which may come before or after the "I-say-to-you" expression (as illustrated in Example (2) below). Schegloff (1996) examines various ways of doing person reference and shows that 'locally initial reference forms in locally subsequent positions' (e.g. addressing the recipient together with "I-say-to-you") are routinely used to mark 'a sequence boundary and the initiation of a new topical departure' in '*disagreement environments of some sort*' (Schegloff

---

3. He calls them "pre-delicates".

4. Categorizing action is notoriously difficult (Sidnell 2017). Here, Mom's turn could be characterized as being a suggestion, an advice, an instruction or even an order. Regardless, it is some sort of directive.

1996: 453). Another such variation is that some "I-say-to-you" expressions come with specific types of discourse marker (DM) at its beginning (as illustrated in Example (3) below). These discourse markers are those that either mark contrasts (such as *danshi* (但是) 'but', *qishi* (其实) 'actually', or *buguo* (不过) 'however') or new 'topics' (such as *lingwai* (另外) 'additionally', *xianzai* (现在) 'now', or *na* (那) 'then'); both of which indicates that the talk after the "I-say-to-you" expressions is 'something new' and in contrast to what the recipient has just said.

4.2   'Preparing the ground' with 'I say to you'

Another interactional function of "I-say-to-you" expressions stems from their above-argued status as 'pre-s'. A common trajectory of pre-sequences is for the projected action not to be delivered immediately after the 'pre' (though it certainly can be). By having signaled an upcoming projected action and having the recipients acknowledge such 'pre's', speakers can then further issue multiple turns of talk before the projected action is actually produced. These turns can consist of inserted sequences of other 'pre's' (Schegloff 2007: 91–114), or other tellings in preparation for the projected action. Hence, recipients of 'pre's' not only orient to it as projecting more talk, but also orient to talk produced after the 'pre' as being relevant to such projected action. This sequential trajectory is also often seen after "I-say-to-you" expressions.

In Example (1b), after the exchange of the *wo gen ni jiang* construction and the go-ahead, Mom then goes on to instructs (or suggests) to her son not to plan a schedule that is too rigid at line 31. Although the son takes that turn to be complete and begins to protest with *bu suan tai-* 'It's not too…', Mom self-selects to extend her turn pass her first possible completion point to deliver more instruction *yao liu dianr yudi* 'you have to give some lee-way (in your plans)' at line 33; and at this second possible completion, continues again to add a maximally polarised yes-preferred tag question *shi bushi a* 'ain't I right?' at line 35 (Raymond 2003). Hence, despite the son's intrusion at line 32, there is evidence to suggest Mom had intended to issue multiple TCUs, or multiple turns of talk, in support of her own position.

A clearer instance is shown in Example (2) below. In this extract, Dad issues a *wo gen ni shuo* type of "I-say-to-you" expression together with an initial address term '*xiaohong*' (i.e. his daughter's name), before 'beating about the bush' on what the projected action might be. As it turns out, Dad was carefully giving unwelcomed advice to her daughter on getting along with her new mother-in-law, a perennial relationship problem in Chinese family culture. By projecting such a resistance-implicative but possibly as yet unknown action to the recipient with

wo gen ni shuo, Dad is then able to secure multiple turns-of-talk where he can "prepare the ground" before the projected delicate action is actually delivered.

Example (2) Wo gen ni shuo [14_8.25]

```
01  →  DAD:   我  跟    你  说  小虹    啊,
              wo  gen   ni  shuo xiaohong a,
              1SG follow 2SG say   NM       SFP
              'I say to you, Xiaohong'
02     HONG: (什么?)
              (shenme?)
              what
              'what?'
03            (.)
04     DAD:   啊.=
              a.=
              Q
              'uhmm'
05            =[baby_noise]
06     DAD:   小虹    啊.
              xiaohong a.
              NM       SFP
              'Xiaohong'
07     HONG: 啊.
              a.
              CFM
              'mmm'
08            (0.3)
09     DAD:   那 各 人    有 各 人    的 习惯.
              na ge  ren   you ge ren   de xiguan.
              DM each person have each person GEN habit
              'everyone has their habit'
10            (0.5) ((baby noises))
11     DAD:   知道  吗?
              zhidao ma?
              know  Q
              'y'know?'
12     HONG: 呃.
              e.
              CFM
              'ummm'
13            (0.3)
```

14	DAD:	那 <这 <u>肯定</u> 是 这 习惯 不 一<u>样</u> 的.
		ne <zhe <u>kending</u> shi zhe xiguan bu yiya:<u>ng</u> de.
		DM DEM sure COP DEM habit NEG same GEN
		'that's definite, habits are different'
15		(0.2)
16	DAD:	啊?=
		a? =
		Q
		'hmmm?'
17	HONG:	=呃.
		=e.
		CFM
		'ummm'
18	DAD:	就 得: 就 得 慢慢 得 相互 适应.
		jiu de:i jiu dei manman de xianghu shiying.
		DM have.to DM have.to slowly have.to mutual adapt
		'you have to, you have to slowly adapt, mutually adapt'
19		(0.5)
20	DAD:	对 吧?
		dui ba?
		right Q
		'right?'
21	HONG:	嗯:.((stoic tone))
		e:n.
		CFM
		'mmmm'
22	DAD:	啊?
		a?
		Q
		'hmmm?'
23	HONG:	嗯:.((stoic tone))
		e:n.
		CFM
		'mmmm'
24	DAD:	errr errr 他 有 他 的 想法儿 是 吧?
		errr errr ta you ta de xiangfa:r shi ba?
		3SG have 3SG GEN idea COP Q
		'erm… erm… she has her own way of thinking, right?'
25		(0.4)

26	DAD:	诶 你 有 你 的 习惯. 哦?
		ei ni you ni de xiguan. o?
		SIP 2SG have 2SG GEN habit Q
		'you also have your own habits.'
27		(.)
28	DAD:	哦?
		o?
		Q
		'hmmm?'
29		(.)
30	DAD:	这个 事儿 呃 绝绝对 是: 绝绝对 有 区别 的.
		zhege shir e jue juedui shi: jue juedui you qubie de.
		DEM.CLF matter PRT surely COP surely have difference GEN
		<哈?=
		<ha?=
		Q
		'this thing, it's definitely different.'
31	HONG:	=嗯.
		=en.
		CFM
		'mmmm'
32	DAD:	这个 在 一块:儿 一块:儿 慢慢 适应 了
		zhege zai yikuair: yikuair: manman shiying le
		DEM.CLF at together together slowly adapt CRS
		'Just mutually adapt to each other,'
33		就 都- 都 就 好 了 哦?
		jiu dou- dou jiu hao le o?
		DM all all DM okay CRS Q
		'and all will be fine.'
34	HONG:	嗯.
		en.
		CFM
		'mmmm'
35	DAD:	共同 建立 一点儿, (0.5)
		gongtong jianli yidianr, (0.5)
		common develop a.bit
		'develop some common'
36		共同 的 这个 这个 这个 习惯, 就 行 了.
		gongtong de zhege zhege zhege xiguan, jiu xing le.
		common GEN DEM.CLF DEM.CLF DEM.CLF habit DM okay CRS
		'common….uhmm… habits, and all is well.'

37   HONG: [嗯.
            [*en.*
             CFM
            'mmmm'

38   DAD: [哦?
            [*o?*
             Q
            'hmmm?'

39   HONG: 嗯.
            en.
            CFM
            'mmmm'

40   (0.5)

41   DAD: 哦?
            o?
            Q
            'hmmm?'

42   HONG: 嗯.
            en.
            CFM
            'mmmm'

43   DAD: 这个   这个   婆婆   这- 弄 孩子
            zhege   zhege   popo   zhe- nong haizi
            DEM.CLF DEM.CLF mother-in-law DEM do   child
            'this… a mother-in-law… this… dealing with kids… '

44   她 就 很 有 经验   的.
            ta jiu hen you jingyan   de.
            3SG DM very have experience GEN
            'she's very experienced with kids'

45   (0.3)

46   DAD: 跟 我(h) 说 了 一 套(h) 她 有(h) 她 的(h)
            gen w(h)o shuo le   yi ta(h)o ta y(h)ou ta   d(h)e
            with 1SG say CMP one CLF   3SG have   3SG GEN
            'she told me her methods, she has her own…'

47   一   套(h) .hhh 挺 好 的. 我 听 了 挺   新鲜.
            yi   t(h)ao .hhh ting hao de. <wo ting le   ting xinxian.
            one CLF   quite good GEN 1SG hear CRS quite fresh
            'methods. It's good, I understand it to be quite original.'

48   HONG: 嗯:::.
            e:::n.
            CFM
            'mmmmm'

The first two lines of this extract follow the prototypical pattern of an "I-say-to-you" expression and an acknowledgment token by the recipient. However, it is noteworthy that instead of responding with the usual continuer, such as *en* (嗯) or *a* (啊) 'mm-hmm', the daughter (Hong) produces something that sounds like *shenme* (什么) 'what' in line 02. As if unsatisfied with such an acknowledgment, Dad initiates a repair and repeats the summon at line 04 and 06, which then elicits the more appropriate acknowledgment token of *a* (啊) 'mm-hmm' at line 07. Hong's initial token of *shenme* 'what' suggests that the recipient has taken *wo gen ni shuo* to be projecting some disagreeable actions. So not only has the recipient oriented to the prior "I-say-to-you" expression as projecting further talk, but the formulation of this specific continuer also shows that the recipient has projected this further talk to be disaffiliative in some way.

Crucially, rather than directly launching into the delicate matter projected, Dad takes multiple turns of talk after the "I-say-to-you" expression to broach the issue in a step-wise manner. He 'prepares the ground' by first couching his argument in universal principles of harmonious human relations, such as *ge ren you ge ren de xiguan* 'everyone has their habits' at line 09, *zhe kending shi zhe xiguan bu yiyang de* 'that's definite, habits are different' at line 14, and *jiu dei manman de xianghu shiying* '(you) have to slowly adapt, mutually adapt' at line 18. Only at lines 24 and 26 did Dad start to particularize his point by mentioning *ta you ta de xiangfa* 'she has her own way of thinking' and *ni you ni de xiguan* 'you also have your own habits'. But even then, it is telling that he uses a *locally subsequent reference form* (the 3rd person pronoun *ta* 'she') in a *locally initial reference position* (i.e. the first time a person is being referred to), thereby leaving the 'who and what' he is referring to for his recipient to infer (Schegloff 1996). It is noticeable that Dad did not mention the crucial referent of 'mother-in-law' until much later at line 43. In fact, the actual precarious action of advising harmony with her mother-in-law, where Dad suggests mutually adapting to each other through cultivating common habits, occurs earlier at lines 32–36 without the use of any pronouns or reference terms. 'Mother-in-law' was only specifically mentioned in the context of a compliment at lines 43–47.

Another keenly discernible pattern in the extract is Dad's multiple attempts at pursuing positive uptake (with tag questions or question particles) after his turns are met with silence or non-committal token (see line 21), only to have Hong respond unenthusiastically again with *e* (呃) 'umm', *en* (嗯) 'mmm', or even more silent treatment (see lines 27 and 29). This pattern of dogged pursuit and reticence occurred at least five times at lines 9–12, 14–17, 18–23, 24–28 and 35–42.

What these exchanges demonstrate is Dad being alive to the resistance-implicative nature of his actions, and hence the gradual and deliberate work he puts into 'preparing the ground' via gently garnering Hong's agreement to his point of view.

It is consequential then that the multi-turn space necessary to execute this careful interactional work was afforded by his "I-say-to-you" preface at the beginning. In other words, one sort of interactional work an initial "I-say-to-you" expression can do, is to allow speakers multiple turns at talk where they can then carefully construct the projected delicate actions in a step-wise manner. This is certainly the case for many of the "I-say-to-you" sequences in the collection. Besides the type of "ground preparation" we have seen here, other such preparations can include explanations before the disagreement, laying out the rationales of an upcoming advice, or justifications before a difficult request.

To summarize, the deployment of an initial 'I-say-to-you' expression is somewhat akin to story prefaces in how they alert their recipients to the nature of the speakers' upcoming multi-turn 'delicate' actions. This is illustrated using the following schema:

Speaker: (DM) + (Address term), 'I-say-to-you' + (Address term) + (turn-final particle)
Recipient: (Acknowledgement token)
Speaker: Multi-turn sequence of "directives" or other delicate actions

While the term 'delicate' indicates that the projected action by the "I-say-to-you" expression may not necessarily be a disagreement (or even overtly disaffiliative), it is nonetheless something that speakers anticipate recipients might possibly resist, and hence the analysis that "I-say-to-you" expressions project for the recipient that the speaker is about to pursue some sort of contingent action. The next section will now demonstrate how "I-say-to-you" expressions can be used to achieve certain interactional goals through projecting such contingencies.

5. Achieving interactional goals with 'I-say-to-you' sequences

One interesting consequence of using "I-say-to-you" expressions to project upcoming turns as doing something delicate and resistance-implicative is, it prompts the recipient to look out for what that something might possibly be; and in some cases, allow the speaker to circumvent the necessity of actually performing the resistance-implicative act. This is quite similar to the well-known 'Stalled' conversation in conversation analytic literature; where Donny was able to provide various situational cues after "Guess what", allowing Marsha to correctly infer the difficult request without him actually saying it. In the previous example, although Dad had to finally deliver the unwelcomed advice after the "I say to you" sequence, Hong was also analyzably quite cognizant of what Dad has been alluding to, way before the advice itself.

In the next exemplar, we examine a long-distance telephone call between another father (Dad) in China and his daughter (Liu), who resides in Ohio. Dad begins his talk with a contrast marker *lingwai* 'additionally' before issuing *wo gaosu ni*, a third type of "I-say-to-you" expression, to foreshadow the delicate matter at hand. As it turns out, the problem was Dad had offered Liu's assistance to a mutual acquaintance going to Ohio without her prior knowledge or consent, thereby violating Liu's freedom of choice. Now, he has the unenviable tasks of reporting his transgression to Liu, and then persuading her to go along with his offer, actions that are clearly disaffiliative and resistance-implicative. However, through projecting the delicate issue with *wo gaosu ni*, and subsequently laying out certain contextual information, Dad was able to elicit the offer of assistance from Liu herself before his own admission, thereby avoiding potentially adversarial positions.

Example (3)   Wo gaosu ni [47_9.49]

01 → DAD: <另外　　啊 我 告诉 你 啊.
<*lingwai   a   wo gaosu ni   a.*
in.addition SIP **1SG tell   2SG** SFP
'and another thing, I tell you'

02　　LIU: °嗯°.
°*en*°.
CRM
'ummm'

03　　DAD: 那个　　曹刊雨:: .hhh
nage: caokanyu:: .hhh
DEM.CLF NM
'that Cao Kanyu,'

04　　LIU: [°嗯°.
[°*en*°.
　CRM
'ummm'

05　　DAD: [诶:: 前天　　　　　他 爸爸 告诉 我
[*ei:: qiantian   ta baba gaosu wo*
　SIP day.before.yesterday 3SG father tell   1SG
'his dad told me the other day'

06　　　　　正式　　被 俄亥俄 大学　　录取.
zhengshi bei ehaie   daxue   luqu.
formal   BEI NM   university recruit
'has officially been accepted at Ohio'

07　　LIU: 嗯.
en.
CRM
'mmm'

08		(0.4)

09	DAD:	诶 一个 月 的 那个 奖学金

ei yige yue de nage jiangxuejin

SIP one.CLF month GEN DEM.CLF scholarship

'uhmm a months stipend'

10		得: 合:: .hh 合 一 千 左右 吧.

de:i he:: .hh he yi qian zuoyou ba.

have.to gather gather one thousand left.right SFP

'is about a thousand I think'

11	LIU:	嗯.

en.

CRM

'mmm'

12	DAD:	嗯.

en.

CRM

'mmm'

13		(.)

14	DAD:	他 可能 是 <u>八月</u> 底::

ta keneng shi <u>bayue</u> di::

3SG maybe COP August bottom

'He's going to the US end of August'

15		八月 二十八 号 ( ) 上 美国 去.

bayue ershiba hao ( ) shang meiguo qu.

August twenty-eight number up NM go

'August 28th I think'

16	LIU:	嗯.

en.

CRM

'mmm'

17	DAD:	没有 别的 事儿.

meiyou biede shir.

NEG.have other.GEN matter

'that's all I want to say'

18	LIU:	嗯.=

en.=

CRM

'mmmm'

19	DAD:	=嗯.

=en.

CRM

'ermmm'

20	MOM:	俄亥俄州 ((shouting in background))
		ehaie zhou
		NM state
		'Ohio State'

21	DAD:	俄亥俄州立 大学 可能 是.=
		ehaie zhouli daxue keneng shi.=
		NM national university maybe COP
		'Ohio State University I think it is'

22	LIU:	=我 知道 那 他 那个- 你 让 他- 你 让 他
		=wo zhidao na- ta nage- ni rang ta- ni rang ta
		1SG know DEM 3SG DEM.CLF 2SG let 3SG 2SG let 3SG
		'I know… Just tell him'

23		来 了 以后 跟 我 联系 就 行 了.
		lai le yihou gen wo lianxi jiu xing le.
		come CRS after follow 1SG contact DM okay CRS
		'to look me up after he arrive'

24	DAD:	啊. 对 <他 那 他 (>因为<) 我 告诉 他 了. <我 说:,
		a. dui <ta na- ta (>yinwei<) wo gaosu ta le. <wo shu:o,
		SIP right 3SG DEM 3SG because 1SG tell 3SG CRS 1SG say
		'Ahh yes, he… I've already told him, I said'

25		.hhhh 诶::: 有 什么 事 需要 在 那 帮忙 的话 嘞,
		.hhhh e:::i you shenme shi xuyao zai na bangmang dehua le,
		SIP have what matter need at DEM help if SFP
		'ummm, if there's anything you needs help with'

26		.hh 你 就 给 杨琉 去 电话. [嗯 他 说,
		.hh ni jiu gei YangLiu qu dianhua. [en ta shuo,
		2SG DM give NM go phone SIP 2SG say
		'just give Yang Liu a call, and he said…'

27	LIU:	[对.
		[*dui.*
		right
		'right'

After *wo gaosu ni*, the types of news Dad relates to Liu includes the acquaintance (Cao Kanyu) being officially accepted into an 'Ohio University' (lines 3–6), that his monthly stipend is about $1000 (lines 9–10), his specific date of arrival into the U.S. (lines 14–15), and later the specific university (Ohio State University) the acquaintance was accepted into (line 21). Given that these information were precisely selected by Dad to be delivered after the "I-say-to-you" expression, it bears upon Liu, who is already residing in Ohio, to surmise how they may be relevant to the projected "delicate" matter. Note that Liu's continuer at line 16 in response

to the announcement of the acquaintance's specific arrival date displays her orientation that Dad has yet to deliver whatever the "delicate" matter might be; but Dad's *meiyou biede shir* 'that's all I want to say' at line 17 cryptic-ally counters that orientation, which also demonstrates his unwillingness to reveal what the *wo gaosu ni* is about. When Liu continues to be less-than-forthcoming at line 18, Dad then re-issues an earlier announcement (line 06) that the acquaintance is going to an Ohio university at line 21. This time, Liu relents by first saying *wo zhidao na-* 'I know…' at line 22, which suggests that she has already caught on to what Dad is trying to say, before offering her services at line 23. It is only at this point that Dad quickly reveals that he has already communicated that 'offer' at lines 24–26. Interestingly, his formulation of *dui<ta na- ta (>yinwei<) wo gaosu ta le* 'Yes. Because I've already told him' at line 24 may also constitute a Freudian slip that evinces our analysis of his prior hesitancy.

In my final exemplar, I want to demonstrate how seemingly innocent news announcement after an "I-say-to-you" expression is nonetheless oriented to by the recipient as possibly 'delicate' news that requires her attention. Example (4) below is again a telephone extract between a father and her daughter.

Example (4)   Wo gen ni shuo [23_11.32]

```
01 →  DAD:   我  跟    你  说   啊.=
             wo gen    ni  shuo a.=
             1SG follow 2SG say  SFP
             'I say to you.'
02    XIN:   =嗯.
             =mm.
             CRM
             'mmm'
03           (1.0)
04    DAD:   今天 是: (.) 妈妈  的   生日.
             jintian shi: (.) mama de   shengri.
             today  COP Mom      GEN birthday
             'Today is Mom's birthday.'
05           (0.8)
06    XIN:   哇:: 妈:  [   :  ]:. [生日    ] 快乐. =hhhhhh
             wa:: ma:  [   :  ]:. [shengri  ] kuaile. =hhhhhh
             INJ  Mom           birthday   happy
             'Wow! Mom! Happy Birthday!!!'
07    DAD:              [hh hh]   [ hh  hh ]
08    XIN:   =hhh[hh
09    DAD:        [hh hh hh hhh[h .hhhh
```

10	XIN:	[hh hh hhh [hh
11	DAD:	[前天　是　爸爸的　生日.
		[*qiantian shi baba de shengri.*
		front.day COP Dad GEN birthday
		'Day before yesterday was Daddy's
		birthday.'
12	XIN:	.hhh 哦 爸:: 生(h)日　快(h)乐. hhh
		.hhh *o~ ba:: she*(h)*ngri ku*(h)*aile.* hhh
		INJ Dad birthday happy
		'Oh Dad! Happy Birthday!'
13	DAD:	嗯.　.hhhh 那么 你们……
		mm. .hhhh *name nimen*……
		CRM then 2PL
		'Mmm. Well, both of you…… '

After the typical *wo gen ni shuo* and acknowledgment token, a climactic second of silence was allowed to lapse before Dad nonchalantly announces at line 4 that *jintian shi mama de shengri* 'Today is Mom's birthday' with an emphasis on 'today'. Another gap of silence ensues, before the daughter, Xin, seems to realise in what way this news might possibly be 'delicate', and then proceeds to rectifies her misconduct (i.e. failure to exhibit remembrance of her mom's birthday) by loudly wishing her mom 'Happy Birthday' in a celebratory manner.[5] This is closely followed by laughter from both Xin and her father, in what Jefferson, Sacks & Schegloff (1977) characterized as 'Offense-Remedial Laugh Sequence' that comes after the surfacing of an offense. At line 11, Dad makes another similar announcement with a conspicuously similar format in *qiantian shi baba de shengri* 'Day before yesterday was Daddy's birthday'. Note also that Dad did not say *wo de shengri* 'my birthday' to denote his own birthday, but uses the relational address term *baba* 'daddy' to mark his status as relative to his daughter. Again, Xin exuberantly wishes her father 'Happy Birthday', interspersed with awkward laugh tokens at line 12.

This exemplar is particularly revealing in that announcement of birthday dates do not, on their own, seem to implicate resistance or have any relevance as 'delicate' or disaffiliative actions. Yet, through Dad's placement of these statements after *wo gen ni shuo*, Xin oriented to them as her dad pursuing some form of action that is potentially disaffiliative to her. In other words, she correctly infers that the announcements, called to attention by *wo gen ni shuo*, alludes to her misdeed, and the trajectory of talk is possibly heading towards actual disaffiliative

---

5. It is unclear if Mom could actually hear Xin over the phone as it is Dad on the other end of the line. In any case, the performative character of this turn simply strengthens our analysis of Xin's action as being in remedy of her offense.

or resistance-implicative admonishments. With this advanced notice, Xin is then able to take reparative measures. Dad, on the other hand, was also able to achieve the interactional goal of side-stepping direct confrontation with his daughter on her misdeed, by using an "I-say-to-you" expression to indicate his upcoming turns are to be heard as 'delicate' talk.

6.   Concluding remarks and further discussions

Summarily, a quick comparative count between a written and spoken Chinese corpus reveals a telling lack of "I-say-to-you" occurrences in the written genres, suggesting that "I-say-to-you" expressions are dedicated interactional resources found only in spoken conversation. By extracting sequences of talk around these expressions and applying conversation analytic methodology to carefully examine their sequential trajectory, "I-say-to-you" bundles are shown to be relatively independent in actual use. By this account, I mean that "I-say-to-you" can be embedded in a turn-medial position, or issued at the end of a turn. They are, however, most often seen occupying an entire turn; and prefacing a larger sequence of talk at its beginning. One common interactional function of "I-say-to-you" proposed in this paper is that they preface an upcoming action as being *"delicate"*. The term 'delicate' is used here in two senses: one is to indicate the care with which the speaker is approaching his/her upcoming action, and two is how the speaker is projecting the upcoming action as possibly problematic for the recipient. There are many possible scenarios why an upcoming action may be 'delicate'. One obvious possibility is in the sequential context of a responsive next turn, where the speaker decides to issue some dispreferred next action, such as a disagreement or other disaffiliative turn in response to the recipient's prior initial action, as in Example (1). Another possibility is that the upcoming action constitutes one of a range of actions that are inherently resistance-implicative. These are typically face-threatening actions such as issuing criticism, giving some form of directive or unsolicited advice, as in Example (2).

As is characteristic of preface-initiated sequences (or 'pre-s'), it is also observed that speakers of "I-say-to-you" sequences are often allowed multiple turns-of-talk before the projected action is actually delivered (Sacks 1992; Schegloff 1980). While some turns-of-talk immediately after "I-say-to-you" may not seem to be problematic on the surface, they are analyzably 'preparing the ground' for the possibly resistance-implicative action ahead (e.g. speakers can provide accounts before an overt disagreement, lay out rationales before giving advice, or explain situations before a difficult request), as in Example (3). In some of these cases, even before the actual 'delicate' action is broached, recipients will orient and respond to

the "I say to you" sequence as heading towards some type of disaffiliative action, thereby validating the efficacy of the "I say to you" preface, as in Example (4).

In this study on *wo gen ni shuo* 'I say to you', *wo gen ni jiang* 'I talk to you' and *wo gaosu ni* 'I tell you', as well as in my previous paper on the interactional functions of *wo juede* 'I think' (Lim 2009, 2011), a key theme that emerges is how the selection of these lexical bundles, often at 'turn-initial positions' (Heritage 2013a), are recipient-designed to manage important issues of progressivity (Sacks 1987) based on the speaker's intersubjective awareness of their interlocutors. While both Chinese "I-say-to-you" expressions and *wo juede* project some sort of upcoming disaffiliative actions, they are also used in interactionally different and consequential ways. *Wo juede* is used predominantly as a hedging device to frames some initial assessment, and often work to elicit a next assessment from its recipient while "minimizing the cost of proffering an opinion", hence the term "joint-assessment initiator" (Lim 2011: 289–295). In contrast, *wo gen ni shuo, wo gen ni jiang* and *wo gaosu ni* are relatively independent units that usually occupy a turn on their own. Additionally, they don't make a next action by the recipient immediately relevant after their production (as opposed to *wo juede*), but in fact, secure multi-turn spaces with which the speaker can use to better prepare or alert the recipient of the possible disaffiliative action ahead. The complex interactional work that they do in conversation is not immediately apparent, save from a careful analysis of their sequential trajectory.

As a caveat, this paper does not claim a definitive or all-encompassing account of the various "I-say-to-you" constructions. Indeed, if we subscribe to the Principle of No Synonymy (Bolinger 1977; Haiman 1983; Goldberg 1995), then the very fact that there are at least three commonly used variants of "I-say-to-you" formulations (*wo gen ni shuo, wo gen ni jiang* and *wo gaosu ni*) suggests that each variant will have different discourse-interactional functions in different contexts. Teasing out their subtleties is beyond the scope of this paper and calls for more detailed comparison between the variants. This study hopes to work towards that goal by first providing a broad-based description of their common sequential contexts and trajectory, as well as introducing some ways which "I-say-to-you" expressions can be used to further interactional goals in conversation.

References

Bolinger, Dwight. 1977. *Meaning and form*. London: Longman.

Biq, Y.-O. 2001. The grammaticalization of jiushi and jiushishuo in Mandarin Chinese. *Concentric: Studies in English Literature and Linguistics* 27: 53–74.

Biq, Y.-O. 2007. Lexicalization of Phrases Involving the DIstal Demonstrative na in Spoken Mandarin. *The Eighteen North American Conference on Chinese Linguistics*, 24–41.

Biber, Douglas & Barbieri, Federica. 2007. Lexical bundles in university spoken and written registers. *English for specific purposes* 26: 263–286.

Biber, Douglas. 2009. A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International journal of corpus linguistics* 14: 275–311.

Brown, Penelope & Levinson, Stephen C. 1987. *Politeness: Some Universals in Language Usage.* Cambridge: Cambridge University Press.

Canavan, Alexandra & Zipperlen, George. 1996. CALLFRIEND Mandarin Chinese-Mainland Dialect LDC96S55. Web Download. Philadelphia: Linguistic Data Consortium.

Conrad, Susan & Biber, Douglas. 2004. "The frequency and use of lexical bundles in conversation and academic prose." *Lexicographica* 20: 56–71.

Craven, Alexandra & Potter, Jonathan. 2010. "Directives: Entitlement and contingency in action." *Discourse Studies* 12: 419–442.

Curl, Traci S. & Drew, Paul. 2008. "Contingency and action: A comparison of two forms of requesting." *Research on Language and Social Interaction* 41: 129–153.

Feng, G. 2008. Pragmatic markers in Chinese. *Journal of Pragmatics* 40: 1687–1718.

Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure.* Chicago: University of Chicago Press.

Haiman, John. 1983. "Iconic and economic motivation." *Language*: 781–819.

Heritage, John. 2013a. "Turn-initial position and some of its occupants." *Journal of Pragmatics* 57: 331–337.

Heritage, John. 2013b. Language and social institutions: The conversation analytic view. *Journal of Foreign Languages* 36: 2–27.

Jefferson, Gail, Sacks, Harvey & Schegloff, Emmanuel A. 1977. "Preliminary notes on the sequential organization of laughter." *Pragmatics Microfiche*. Cambridge, UK: Cambridge University Department of Linguistics.

Kern, Friederike & Selting, Margret. 2013. Conversation Analysis and Interactional Linguistics, in Carol A. Chapelle (ed.) *The Encyclopedia of Applied Linguistics*: 1012–1016. Malden, MA: Wiley-Blackwell.

Li, Charles N. & Thompson, Sandra A. 1989. *Mandarin Chinese: A functional reference grammar.* Berkeley: University of California Press.

Li, Xiaoting. 2016. Some discourse-interactional uses of yinwei 'because' and its multimodal production in Mandarin conversation. *Language Sciences* 58: 51–78.

Lim, Ni-Eng. 2009. "Stance-taking with Wo Juede in conversational Chinese." *21st North American Conference on Chinese Linguistics (NACCL-21)* 2: 323–340.

Lim, Ni-Eng. 2011. "From Subjectivity to Intersubjectivity: Epistemic Marker Wo Juede in Chinese." In *Current Issues in Chinese Linguistics*, ed. by Yun Xiao, Liang Tao, and Hooi-Ling Soh, 265–300. New Castle: Cambridge Scholars Press.

McEnery, Anthony & Xiao, Richard. 2004. "The Lancaster Corpus of Mandarin Chinese: A corpus for monolingual and contrastive language study." *Religion* 17: 3–4.

O'Keeffe, Anne, McCarthy, Michael & Carter, Ronald. 2007. *From corpus to classroom: Language use and language teaching.* Cambridge: Cambridge University Press.

Raymond, Geoffrey. 2003. "Grammar and Social Organization: Yes/No interrogatives and the structure of responding." *American Sociological Review* 68: 939–967.

Sacks, Harvey. 1987. "On the Preferences for Agreement and Contiguity." In *Talk and Social Organization*, ed. by Graham Button and John R. E. Lee, 54–69. Clevedon: Multilingual Matters.

Sacks, Harvey. 1992. "Lecture 1: Second Stories; "Mm hm;" Story prefaces; 'Local news;' Tellability." In *Lectures on Conversation*, ed. by Gail Jefferson and Emmanuel A. Schegloff, 3–31. Oxford: Blackwell.

Schegloff, Emmanuel A. 1980. "Preliminaries to preliminaries:'Can I ask you a question?'" *Sociological Inquiry* 50: 104–152.

Schegloff, Emmanuel A. 1987. Analyzing single episodes of interaction: An exercise in conversation analysis. *Social Psychology Quarterly*, 50(2): 101–114.

Schegloff, Emmanuel A. 1996. "Some practices for referring to persons in talk-in-interaction: A partial sketch of a systematics." In *Studies in Anaphora*, ed. by Barbara Fox, 437–486. Amsterdam: John Benjamins Publishing Company.

Schegloff, Emmanuel A. 2007. *Sequence Organization in Interaction*. Cambridge: Cambridge University Press.

Scott, Mike. 2016. *WordSmith Tools version 7*. Stroud: Lexical Analysis Software.

Sidnell, Jack. 2017. "Action in interaction is conduct under a description." *Language in Society* 46: 313–337.

Tao, Hongyin. 2011. "Core Vocabulary in Spoken Mandarin and the Integration of Corpus-Based Findings into Language Pedagogy." In *Issues in Chinese Linguistics*, ed. by Yun Xiao, Liang Tao, and Hooi-Ling Soh, 13–27. New Castle: Cambridge Scholars Press.

Thompson, Sandra A. & Mulac, Anthony. 1991. "The discourse conditions for the use of the complementizer that in conversational English." *Journal of pragmatics* 15: 237–251.

Thompson, Sandra A. 2002. "'Object complements' and conversation: Towards a realistic account." *Studies in language* 26: 125–164.

Tsai, I-Ni. 2008. "Projecting the Unanticipatory: The Mandarin Particle Ei and its Projectability in Daily Conversation." *Twentieth North American Conference on Chinese Linguistics (NACCL-20)* 2: 1023–1040.

Wang, Y.-F., Tsai, P.-H. & Yang, Y.-T. 2010. Objectivity, subjectivity and intersubjectivity: Evidence from qishi ('actually') and shishishang ('in fact') in spoken Chinese. *Journal of Pragmatics* 42: 705–727.

Wray, Alison. 2005. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

Wu, R.-J. R. 2014. Managing turn entry: The design of EI-prefaced turns in Mandarin conversation. *Journal of Pragmatics* 66: 139–161.

Appendix I. Transcription conventions

[or [ ]	overlapping or simultaneous talk
=	a "latch" sign, that is, the second speaker follows the first with no discernible silence between them. When the latch sign is between utterances by the same speaker, it indicates that the speaker's talk is continuous even though there is another speaker in between.
(0.5)	length of silence between utterances in tenths of seconds

(.)	micropause
?	rising intonation, not necessarily a question
,	continuing intonation
¿	the inverted question indicates a rise stronger than a comma but weaker than a question mark
!	exclamatory intonation
-	a cut-off or self-interruption
<	the less than symbol indicates that the immediately following talk is "jump-start-ed," i.e., sounds as if it starts with a rush
> <	the combination of "more than" and "less than" symbols indicates that the talk between them is compressed or rushed.
< >	markedly slowed or drawn out, compared to the surrounding talk
°	following talk is markedly quiet or soft
↑↓	mark sharper rises or falls in pitch
:::	indicates prolongation or stretching of the preceding sound (the more colons the longer the stretching)
_:	Inflected falling intonation contour
:_	Inflected rising intonation contour
hhh	exhalation, hearble aspiration, or laughter, the more "h"s, the more aspiration
(hhh)	laughter inside the boundaries of a word
.hhh	inhalation
Wo(h)d	plosive aspiration within a word, which may result from breathiness, crying, or laughter
word	Underlining indicates some form of stress or emphasis
WOrd	especially loud talk
(word)	uncertainty on the transcriber's part
( )	something is said, but unable to do minimal deciphering
(( ))	transcriber's description of event

Appendix II. Abbreviations in Chinese gloss

1/2/3 SG	first/second/third person singular	DEM	Demonstrative
1/2/3 PL	first/second/third person plural	EXP	Experiential aspect marker
AGR	Agreement Particle	GEN	Genitive Case Marker
BA	the *ba* transitivity marker	INJ	Interjection
BEI	the *bei* (and *gei*) 'passive' marker	LOC	Locative Complement
CFM	Confirmation Particle	NEG	Negator
CRS	Current relevant state Particle	NM	Proper Name
COM	Verb complement	PCM	Potential Complementizer
COP	Copula	Q	Question Particle
CMP	Completion aspect marker	SFP	Sentence Final Particle
CTP	Continuous aspect marker	SIP	Sentence Initial Particle
CLF	Classifier	TP	Topic Marker
DM	Discourse marker		

Chinese near-synonyms *jian* (建), *zao* (造), *gai* (蓋) 'to build' revisited

Chenhsuan Huang

National Chung Cheng University, Taiwan

This study reexamines Mandarin Chinese near-synonyms *jian* (建), *zao* (造), *gai* (蓋) 'to build' and their shared patterns in written and spoken genres. Three independent variables – including word length of the object NP, preverbal locative phrase, and building purpose – were tested by a logistic regression analysis (Rbrul) to account for the multiple crosscutting and interacting factors that influence language usage. Multivariate analyses show that word length and building purpose can account for the differences among these verbs in both genres. The analyses suggest that the use of *jian* (建) and *zao* (造) possess more written properties, while *gai* (蓋) favors the spoken genre. The current study contributes to a growing number of studies in Chinese near-synonyms by focusing on genre variation.

Keywords: Mandarin, near-synonym, regression analysis, word length, written, spoken, genre variation

1. Introduction

Near-synonyms are seen as highly difficult for language learners. Due to the limitations of dictionaries, learners cannot simply look up similar words and immediately tell their differences. Also, the meaning of near-synonyms usually depends on various situations; thus, determining how to choose the right word becomes a challenge for both instructors and learners. In modern Chinese, for example, *jian* (建) and *zao* (造) both share the similar concept of the verbal form 'to build', and they can be integrated into a word set such as *jianzao* (建造), meaning 'to build'. Similar structures include *benpao* (奔跑) 'to run', *lingting* (聆聽) 'to listen to', and so on. Characters in these synonym word sets can be interpreted as similar to each other. However, they are not interchangeable in most cases, and have different collocations when appearing in sentences. For instance, both *ling* (聆) and *ting*

(聽) have the same meaning, but *ling* (聆) is used more in written form as *lingyue* (聆樂) 'to listen to music' while *ting* (聽) is a colloquial expression as in *tingyinyue* (聽音樂) 'to listen to music'. For near-synonyms like *wuduan* (武斷) and *guoduan* (果斷), although they both refer to make decisions quickly and confidently, *wuduan* (武斷) is used in a negative way as arbitrary or biased while *guoduan* (果斷) is a positive term. As Cheng and Tian (2002: 192–196) stated, near-synonyms provide options for various situations, including commendation, devaluation, formality, and informality.

Numerous studies (Tsai et al. 1996; Hong 2010; Tsai 2010, 2011; Wang 2012) tried to find the subtle differences between near-synonyms by their syntactic behaviors such as syntactic functions, occurrence frequencies, collocations, and semantic characteristics. For example, Tsai et al. (1996) differentiated the pair of near-synonyms *gaoxing* (高興) and *kuaile* (快樂), which can both mean 'happy' in Chinese. In the case of *gaoxing* and *kuaile*, they found that *gaoxing* can be associated with the sentential-final particle *le*, whereas *kuaile* cannot. This is because *gaoxing*, with the feature [± effect], represents a change of state triggered off by some causes. In the same vein, Liu (1999) investigated the Chinese verbs *jian* (建), *zao* (造), and *gai* (蓋) 'to build', and found that they differ in at least four aspects: (a) their transitive/intransitive use; (b) their co-occurrence with a preverbal locative phrase; (c) the theme of the object NP – *jian* (建) and *zao* (造) may both occur with non-constructional entities (e.g., *jianjun* (建軍) 'build the army'), while *gai* (蓋) tends to stick to the object NPs relating to constructions; (d) and their metaphorical extension, such as *gai* (蓋) never being used metaphorically in non-spatial domains. Furthermore, Liu came to the conclusion that *jian* (建) is more endpoint focused and thus allows a stative, agentless, intransitive usage. On the other hand, *zao* (造) indicates human design and focuses more on the creative role of the agent, thus imposing less restriction on the semantic range of its object NPs. *Gai* (蓋) names the unmarked activity of construction, which is highly spatially oriented and thus often accompanied by a locative specification. Although previous studies prove that the semantic components properly account for the syntactic differences of near-synonyms in most cases, the refined distinctions or common usages among near-synonyms still remain unclear.

In the case of *jian* (建), *zao* (造), and *gai* (蓋) 'to build', the three verbs are also polysemous words, meaning that each has more than one meaning. Hence, treating their meanings under the rubric of near-synonyms would be inappropriate. Based on Liu (1999), *gai* (蓋) is the most restricted and highly specified with its object NP theme relating to buildings, while *jian* (建) is less rigid, allowing non-building object NPs. The object NP of *zao* (造) is the most flexible, since it involves human design or some other complex structures. Accordingly, the theme of the object NP among the three verbs is shown in Table 1.

Table 1. Theme of object NP and metaphorical domains (see Table 11 of Liu 1999: 191)

	Building	Object[a]	Quality/concept	Language[b]
gai	Yes			
jian	Yes	Yes	Yes	
zao	Yes	Yes	Yes	Yes

a. It refers to non-building object NPs.

b. 'Language' here means *ci* (詞) 'phrases' or *ju* (句) 'sentences'. Examples include *zaoci* (造詞) and *zaoju* (造句).

As Table 1 shows, *jian* (建), *zao* (造), and *gai* (蓋) 'to build' share the same thematic object NP, such as *fangzi* (房子) 'house' in the category of 'building'; however, further analysis on the use of mutual types is lacking. In addition, genre is one of the factors that make language use divergent. Variation across spoken and written has been studied intensively (Biber 1988, 1999; Clancy 1982; Miller and Weinert 1998; Iwasaki 2015; Tao 1999). The differences between the two genres may lie in the syntactic patterns (Zou, Smith, and Hoey 2015: 44–45), choice of the lexicon (Wu and Cheng 2011; Lim and Hong 2012; Chang 2015), and even gender (Herring and Paolillo 2006). In order to learn more about the near-synonyms, the present study therefore uses a logistic regression analysis to shed light on how the three verbs *jian* (建), *zao* (造), and *gai* (蓋) 'to build' vary with the same thematic (i.e., building) object in written and spoken forms. The research questions of the study are as follows.

1. Is there any genre difference among the use of *jian* (建), *zao* (造) and *gai* (蓋) 'to build'?
2. If so, which factors have the greatest impact? (i.e. word length of the object NP, preverbal locative phrase or building types)
3. To what extent can those factors account for the variation?

2. Method

2.1 Definition

In order to discover the fine distinction among the three verbs under investigation, I limit our focus structure to [V+NP]. For example, the verbs *jian* (建), *zao* (造), or *gai* (蓋) 'to build' may be followed by an object NP. In addition, the semantic theme of the object is constrained to 'building' only. According to the definition provided by the Ministry of Education Republic of China (Taiwan) (MOE-TW), the so-called *jianzhu* (建築) 'building' in Mandarin should meet these criteria: (a) be a stationary structure with roof, wall, or beams; (b) be located on or beneath

the ground; and (c) provide a personal or public-use purpose, such as dwelling, working, recreation, storage, and so on. Thus, qualified objects such as *fangzi* (房子) 'house', *qiao* (橋) 'bridge', and *shuiku* (水庫) 'reservoir' would be included in the category of *jianzhu* (建築) 'building'. As the structure and the semantic nature of the noun have been fixed, what follows describes the investigated factors of the analysis.

2.2   Data sources

There are two data sources in this analysis: written and spoken. The written data come from *Xinwen Zhishiku* (新聞知識庫) 'News corpus'. *Xinwen Zhishiku* (新聞知識庫) 'News corpus' is an integrated platform of Taiwan local newspapers created by Tudor Tech Systems company. The written corpus included excerpts from eight major Taiwan newspapers, such as the *United Daily News* (聯合報), *United Evening News* (聯合晚報), *United Daily Local News* (聯合報地方版), *Economic Daily News* (經濟日報), Min Sheng *Daily* (民生報), *China Times* (中國時報), *Commercial Times* (工商時報) and *Apple Daily* (蘋果日報). In order to probe the latest language use in Taiwan society, the data was chosen from 2012 to 2015. A keyword search was used to find *jian* (建), *zao* (造), and *gai* (蓋) 'to build' within the texts. In order to return accurate results, the process of data collection was done without any automatic programming software. On the other hand, the spoken data were collected from YouTube videos. The recorded video types included news reports, such as *Huashi Xinwen* (華視新聞), *Dongsen Xinwen* (東森新聞), *Sanli Xinwen* (三立新聞), interior design programs such like *Shejijiazhengtuan* (設計家正團), *Lujianzhudezaidijingyan* (綠建築的在地經驗), interview programs such as *Taiwan yiqianlingyigegushi* (台灣1001個故事), *yibuyijiaoyin* (一步一腳印), and lecture videos with the theme of 'building' within the recent five years (2011–2016). The video data were then transcribed into written form manually. In addition, in order to meet the semantic criteria as mentioned earlier, the whole transcription process was done manually by checking the items one by one. The total length of the videos was about 6.5 hours.

2.3   Data collection

As mentioned earlier, the focus structure of the analysis is [V+NP]. The verb should be monosyllabic morpheme only; disyllabic verbs such as *jianzao* (建造) 'to build' and *jiangou* (建構) 'to build' were excluded from this study. In terms of the noun (object), it must meet the semantic criteria of 'building' by the MOE definition. The excluded tokens can be found in Appendix 1. Since the excluded tokens would not be used in our statistics, the number of the excluded tokens would

then be ignored. In sum, the adopted written data take about 48.7% (202 tokens), and the spoken data are about 51.3% (213 tokens) in this study. Table 2 shows the overall data structure. After the process of data extracting, both written and spoken data would be gathered in Microsoft Excel Spreadsheet for data coding.

Table 2. Data structure

Data types	Adopted tokens
Written	202 (48.7%)
Spoken	213 (51.3%)
Total	**415 (100%)**

2.4   Data coding

As mentioned in previous literature, the differences between spoken and written could be related to the phonological context (García 2011), syntactic patterns, and choice of lexicons. Moreover, to examine the usage variation among *jian* (建), *zao* (造), and *gai* (蓋) 'to build' varying with the same thematic object NP in written and spoken forms, the possible variant in [VO] structure would be then focused on the either the NP part or the position of the structure. For this, the factors investigated in the analysis include phonological length of object NP, preverbal locative phrase, and building types. The detailed explanations regarding the factors are listed below.

2.4.1   *Word length of the object NP*

Modern Chinese, unlike classical Chinese, contains thousands of polysyllabic words (Norman 1988: 154). It is estimated that more than 60% of words in Chinese consist of two characters (Liu 1999). Most of the time, one character is usually counted as one syllable, and thus disyllabic words become the most common and preferred word length in modern Mandarin. Duanmu (2012) used a quantitative approach to prove that word length preferences. He found that length pattern 1+1 of [VO] (e.g., *qing ke* (請客) 'invite guest') dominates others such as 1+2 (e.g., *diu mianzi* (丟面子) 'lose face'), 2+1 (e.g., *diandian tou* (點點頭) 'nod head') of [VO] from a corpus data. In the same vein, I count word length of the object NPs that *jian* (建), *zao* (造), and *gai* (蓋) 'to build' carry. Specifically, the NPs are grouped into three patterns by their phonological length: '1' (e.g., *zaowu* (造屋) 'to build a house'), '2' (e.g., *gai fangzi* (蓋房子) 'to build a house'), and '3' or above, such as *jian tingchechang* (建停車場) 'to build a parking lot' and *gai baihuogongsi* (蓋百貨公司) 'to build a department store'.

2.4.2   *Preverbal locative phrase*

On a syntactic level, Liu (1999) mentioned that *gai* (蓋) has a relatively higher percentage of co-occurrence with a preverbal locative phrase: for instance, *zai Dongju Dao gai le yi zuo dengta* (在東莒島蓋了一座燈塔) '(they) built a light-house in *Dongju* Island'. However, such a statement was based on the general usage of the three verbs under investigation. In this study, I would like to further explore whether the position of locative phrase becomes an influential factor that domi-nates the use of the three verbs.

2.4.3   *Building types*

Many studies have suggested that the semantic domain or collocation determines the choice of the lexicons (Wu and Palmer 1994; Inkpen and Hirst 2002; Lee and Liu 2009). That is, the collocations are not arbitrary; on the contrary, they usually have certain semantic tendencies in that the NPs of the study are set as buildings and one of the criteria for building is its function for personal or public use. Thus, I categorize the buildings by their types, such as residential (e.g., house, apart-ment) or nonresidential (e.g., school, church, bridge, reservoirs, etc.), and to test if the building types may affect the use of the three verbs via quantitative statistical analysis.

2.5   Statistical analysis

In order to answer the research question, I adopted Rbrul (Johnson 2009) for lo-gistic regression analysis for the statistical testing of the data. The advantage of using the Rbrul software is that it can model interactions between independent variables and hierarchical (mixed-effects) models. The basic concept of a variable rule analysis is that it can compute a multivariate statistical model on the basis of observed token counts; each determining factor is assigned a numerical factor weight that describes how it influences the probabilities of choice of either form. In addition, it is also a free statistical program within the software environment R, and it helps analysts find the most salient factor(s) within the variation. To run the program, I needed to divide the variables into two as independent variables and dependent variables. The independent variables were possible factors that deter-mined the variation, including word length of object NP, preverbal locative phrase and building types. The dependent variables were the use of *jian* (建), *zao* (造), and *gai* (蓋) 'to build'. In order to compare the usage differences between spoken and written genre, the Rbrul was thus applied separately.

3. Results

I first present the results of the distribution of *jian* (建), *zao* (造), and *gai* (蓋) 'to build' by cross comparison based on the different genres. Next, the best combinations of factors that can account for the variability of the three verbs in both genres would be showed.

To have a general concept of how these three verbs are used in the data, Table 3 and Table 4 show the token numbers of written and spoken data in detail. As can be seen in the tables, the use of *jian* (建), *zao* (造), and *gai* (蓋) 'to build' has divergent tendencies across each factor group. In word length of the object NP, it seems that the use *zao* (造) has a tendency toward monosyllabic objects, while the spoken *gai* (蓋) seems to have a tendency toward objects with more than one syllable. Moreover, *gai* (蓋) tends to co-occur with preverbal locative phrases in written contexts, but shows no obvious tendency in its spoken form. However, these are the raw counts. A multivariate analysis allows us to take each factor into account in order to determine which factors can best explain the variation.

Table 3. Distribution of *jian* (建), *zao* (造) and *gai* (蓋) 'to build' across factors in written data (raw numbers)

Group factors	*jian* (建)	*zao* (造)	*gai* (蓋)	Total
Word length of the object NP		Tokens		
1	41 (45.05%)	**18** **(19.78%)**	32 (35.16%)	91 (100%)
2	20 (25.64%)	0	**58** **(74.35%)**	78 (100%)
3 or above	10 (30.3%)	0	**23** **(69.7%)**	33 (100%)
Preverbal locative phrase		Tokens		
With	6 31.58%)	0	**13** **(68.42%)**	19 (100%)
Without	65 (35.52%)	18 (9.84%)	100 (54.64%)	183 (100%)
Building types		Tokens		
Residential	26 (27.08%)	17 (17.71%)	53 (55.21%)	96 (100%)
Non-residential	45 (42.45%)	1 (0.94%)	60 (56.6%)	106 (100%)
Total	**71** **(35.15%)**	**18** **(8.91%)**	**113** **(55.94%)**	**202** **(100%)**

Table 4. Distribution of *jian* (建), *zao* (造) and *gai* (蓋) 'to build' across factors in spoken data (raw numbers)

Group factors	*jian* (建)	*zao* (造)	*gai* (蓋)	Total
Word length of the object NP		Tokens		
1	6 (16.22%)	19 (51.35%)	12 (32.43%)	37 (100%)
2	4 (3.1%)	2 (1.55%)	**123 (95.35%)**	129 (100%)
3 or above	5 (10%)	0	**45 (90%)**	50 (100%)
Preverbal locative phrase		Tokens		
With	0	0	**10 (100%)**	10 (100%)
Without	11 (5.42%)	21 (10.34%)	**171 (84.24%)**	203 (100%)
Building types		Tokens		
Residential	7 (4.27%)	21 (12.8%)	**136 (82.93%)**	164 (100%)
Non-residential	4 (8.16%)	0	**45 (91.83%)**	49 (100%)
Total	**11 (5.16%)**	**21 (9.86%)**	**181 (84.98%)**	**213 (100%)**

3.1   The interpretation of jian

Table 5 shows the Rbrul result for *zao* (造) and *gai* (蓋) based on the written data. Since Rbrul can only deal with binominal analysis, the results on *zao* (造) and *gai* (蓋) actually reveal the opposite interpretation for *jian* (建). For instance, if the results of *zao* (造) and *gai* (蓋) have tendency *X*, then it indicates that *jian* (建)

Table 5. *Rbrul* report for *zao+ gai* (*cf. jian*) based on the written data

Factor groups	Factors	Log odds	Tokens	Centered weight
Building types $p < 0.001$	Residential	0.416	96	**0.603**
	Non-residential	−0.416	106	0.397
Word length $p = 0.0109$	'1'	−0.619	91	0.35
	'2'	0.292	78	**0.573**
	'3 or above'	0.327	33	**0.581**

Deviance = 247.63; df = 4; Grand mean = 0.649.
All factors are significant by Bonferroni correction.

would dominate in factors opposite of *X*. Conventionally, weight over 0.5 indicates the associated factor favors the use of certain patterns. Log odds are raw coefficients for the regression model. The larger the number is, the bigger the effect size would be. Here, the factors that distinguish *zao* (造) and *gai* (蓋) from *jian* (建) in written form are building types and word length of the object NP. In building types, the combination of written *zao* (造) and *gai* (蓋) favor the residential buildings as (1) and (2) since its centered weight is 0.603. On the contrary, it also means that *jian* (建) tends toward objects with nonresidential purposes, as with the airport and harbor in (3) and (4), respectively.

(1)　建　　築　師　　謝　　英　　俊　　的　　協力
　　jian　*zhu shi*　*xie*　*ying*　*jun*　*de*　*xieli*　造 屋 法
　　Architect Xie Ying-Jun *de*-MOD cooperate build house method *zao wu fa*
　　'Architect Xie Ying-Jun's co-operative construction method'

(2)　讓　學生　　走入　社區　　　動手　蓋　屋
　　rang　*xuesheng*　*zouru*　*shequ*　*dongshou*　*gai*　*wu*
　　make students walk-in community start build house
　　'Make students to build houses for the community'

(3)　菲律賓　　在 南海　　　　中業　島　上　建　機場
　　Feilubin　*zai Nanhai*　　　*Zhongye dao*　*shang jian*　*ji chang*
　　Philippines at South China Sea Zhongye Island on build airport
　　'Philippines built an airport on *Zhongye* Island in South China Sea'

(4)　基隆　港　建　港　130 周年
　　Jilong　*gang*　*jiang gang*　*130 zhou nian*
　　Keelung harbor build barbor 130 year
　　'*Keelung* Harbor had its 130th anniversary'

In terms of the word length of the object NP, the combination of written *zao* (造) and *gai* (蓋) tends to attract objects with more than one syllable because the centered weight in factors '2' and '3 or above' are 0.573 and 0.581, respectively. Examples include *wuzi* (屋子) 'house' in (5) and *laorenyuan* (老人院) 'retirement home' in (6). In contrast, such result also indicates that *jian* (建) prefers to have a monosyllabic word as its object NP, such as *jianfang* (建房) 'build a house' and *jianchang* (建廠) 'build a factory'.

(5)　義工　　幫 忙　　造　屋
　　yigong　*bangmang*　*zao*　*wuzi*
　　volunteer help build house
　　'Volunteers help to build the house'

(6) 花　三　百　　億　元　蓋　老人　院
     hua san bai   yi   yuan gai laoren yuan
     spent three hundred billion dollar build elder   building
     '(they) spent three hundred billion dollars to build the retirement home'

As for the results in spoken language, Table 6 shows that word length of the object NP is the main factor that distinguishes spoken *zao* (造) and *gai* (蓋) from *jian* (建). The result indicates that the combination of *zao* (造) and *gai* (蓋) tends toward the objects with two syllables or more since the centered weight are 0.616 and 0.702, respectively; on the contrary, it implies that the spoken *jian* (建) prefers the monosyllabic object. It is interesting to note that the number of constraints (factors) for the written *jian* (建) is higher than for its spoken form. Regarding this phenomena, a recent study comparing Mandarin near-synonyms: *zhihou* (之后) and *yihou* (以后) 'later/after' in both written and spoken discourses indicates a similar result that written genre might have more constraints on the usage of a given word (Su 2017).

Table 6. *Rbrul* report for *zao+ gai* (*cf. jian*) based on the spoken data

Factor groups	Factors	Log odds	Tokens	Centered weight
Word length p = 0.0144	'1'	−1.329	37	0.209
	'2'	0.471	129	**0.616**
	'3 or above'	0.858	47	**0.702**

Deviance = 78.142; df = 3; Grand mean = 0.948.
All factors are significant by Bonferroni correction.

3.2   The interpretation of zao

Table 7 displays the Rbrul result for the combination of *jian* (建) and *gai* (蓋) based on the written data. It indicates that the factors that distinguish written *jian* (建) and *gai* (蓋) from *zao* (造) are the word length of NP and building types. As can be seen in the results, the combination of *jian* (建) and *gai* (蓋) prefers the object NP with longer word length because the centered weights reach 0.999 and 0.997 in factors '2' and '3 or above'. Examples can be found in (3) and (6). On the other hand, this also means that *zao* (造) favors shorter word length, especially monosyllabic object NPs with monosyllables, such as *zaofang* (造屋) 'build a house'. Based on the interpretation in Section 3.1, although both *jian* and *zao* (造) favor monosyllabic objects, it is worth noting that the centered weight in *zao* (造) is much higher than that of *jian* (建). To some extent, *zao* (造) has stronger tendency toward monosyllabic object NPs. In terms of the building types, the analysis demonstrates that *jian* (建) and *gai* (蓋) tend to co-occur with objects for

the nonresidential uses (centered weight = 0.824) while *zao* (造) is often used for buildings with residential purpose. Turning to the spoken data, there is no significant effect for the combination of *jian* (建) and *gai* (蓋) versus *zao* (造).

Table 7. *Rbrul* report for *jian + gai* (*cf.zao*) based on the written data

Factor groups	Factors	Log odds	Tokens	Centered weight
Word length *p* = 1.98e-07	'1'	−12.459	91	<0.001
	'2'	6.711	78	**0.999**
	'3 or above'	5.748	33	**0.997**
Building types *p* = 2.45e-05	Residential	−1.541	96	0.176
	Non-residential	1.541	106	**0.824**

Deviance = 72.714; df = 4; Grand mean = 0.911.
All factors are significant by Bonferroni correction.

3.3 The interpretation of gai

Table 8 and Table 9 show the main factor that differentiates *gai* (蓋) from the other two verbs, *jian* (建) and *zao* (造). As is shown in Table 8, the combination of *jian* (建) and *zao* (造) favors monosyllabic NPs as their objects since the centered weight is 0.739. Similary, in the spoken data, the centered weight of subfactor '1' is even higher, which reaches at 0.941 in Table 9. This indicates that the combination of spoken *jian* (建) and *zao* (造) is subject to relatively more strict uses. That is to say, there are fewer chances of finding the object NP with more than one syllable

Table 8. *Rbrul* report for *jian + zao* (*cf. gai*) based on the written data

Factor groups	Factors	Log odds	Tokens	Centered weight
Word Length *p* = 3.28e-07	'1'	1.040	91	**0.739**
	'2'	−0.636	78	0.346
	'3 or above'	−0.404	33	0.4

Deviance = 247.31; df = 3; Grand mean = 0.441.
All factors are significant by Bonferroni correction.

Table 9. *Rbrul* report for *jian + zao* (*cf. gai*) based on the spoken data

Factor groups	Factors	Log odds	Tokens	Centered weight
Word Length *p* = 4.22e-17	'1'	2.772	37	**0.941**
	'2'	−0.983	129	0.272
	'3 or above'	−1.790	47	0.143

Deviance = 104.838; df = 3; Grand mean = 0.15.
All factors are significant by Bonferroni correction.

in spoken *jian* (建) and *zao* (造). On the contrary, such results reveal that *gai* (蓋) prefers multisyllabic NP objects, and this tendency is more common and obvious in the spoken form.

4. Discussion

Three factors were tested for their impact on the use of verbs *jian* (建), *zao* (造), and *gai* (蓋) 'to build'. As the results show, word length of the object NP and building types are the prechosen factors that have the most significant correlations with the verbs. In word length of the object NP, *jian* (建) and *zao* (造) favor monosyllabic morphemes especially for written form , while *gai* (蓋) prefers longer words as its objects in both genres. In our data, the most frequently used objects of *jian* (建) include *wu* (屋) 'house', *chang* (廠) 'factory', *si* (寺) 'temple', *lou* (樓) 'buildings' and *miao* (廟) 'temple'. The most frequently used object of *zao* (造) is *wu* (屋) 'house'. However, the combination of *gai* (蓋) predominates in trisyllabic or longer words, such as *fangzi* (房子) 'house', *bangqiuchang* (棒球場) 'baseball stadium', *bangongdalou* (辦公大樓) 'business buildings', *shehuizhuzhai* (社會住宅) 'social rented housing' and so on. Such preference of word length is more apparent in the spoken form. Written Chinese, or called *shumianyu* (書面語), still retains a large portion of its ancient roots, and thus possess more monosyllabic words than spoken form (Norman 1988; Wilkinson 2000). Also, Lee and Chen (1993) pointed out that word length is one of the factors accounting for the difference in written and spoken Chinese. They argued that Chinese speakers tend to make the word syllables more concise when moving from spoken to written language. For example, the conjunction *erqie* (而且) 'and' becomes *er* (而) or *qie* (且) in written form. Also, the phrase *tade nupengyou* (他的女朋友) 'his girlfriend' would be stated in a concise form, as *qi nuyou* (其女友), in written discourse, in which *tade* (他的) 'his' is replaced by the monosyllabic word *qi* (其), while *nupengyou* (女朋友) 'girlfriend' is shortened to *nuyou* (女友). In other words, syllable size in written discourse is more condensed and shorter than in spoken discourse. This argument can be used to support our findings on the use of *jian* (建), *zao* (造), and *gai* (蓋). To sum up, *jian* (建) and *zao* (造) possess more written properties.

Another possible explanation on the word length could be due to the human cognition. According to Lakoff and Johnson (1980: 126), speaking and writing systems allow us to conceptualize linguistic form in spatial terms, so that we know which words occupy first, or whether a word is relatively long or short. Thus, they proposed several principles in terms of the linguistic forms to their meanings. Based on their principle that MORE OF FORM IS MORE OF CONTENT (Lakoff and Johnson 1980: 127), an English sentence such as 'He ran and ran and ran and ran'

indicates more running than just 'He ran'. Moreover, extending the lengthening of a vowel can also have the same effect; for instance, 'he is bi-i-i-ig' versus 'he is big' (Lakoff and Johnson 1980: 127). Other studies (Lewis and Frank 2016; Mikk et al. 2001) also indicated that word length is highly related to the word's semantic complexity. For the same reason, the object of *gai* (蓋) has a longer word length, and thus provides more content of that kind. On the other hand, shorter forms such as *jian* (建) and *zao* (造) contain less meaning. In terms of the word length preference, it could be associated with lexical hierarchy. As we can see from the examples above, a monosyllabic object only represents the generic term of the kind, such as *wu* (屋) 'house' or *chang* (廠) 'factory' while multisyllablic object NPs give relatively more details about the contents. For instance, *fangzi* (房子) 'house' is endowed with more subjective expression (Kuo 2008) than *wu* (屋) 'house'. Also, *bangongdalou* (辦公大樓) 'business buildings' refers to the specific purpose of the building. To put it simple, *jian* (建) and *zao* (造) are at a higher level of the semantic scope while *gai* (蓋) is at a lower lexical hierarchy. In sum, the choice among *jian* (建), *zao* (造), or *gai* (蓋) 'to build' is not only a matter of word length but, more importantly, is also related to human categorization.

In addition, the results show that the written form is more impacted by the NP's building type, especially for *jian* (建) and *zao* (造). *Jian* (建) tends to collocate with nonresidential buildings, such as temple, school, and so on. *Zao* (造), on the other hand, co-occurs with buildings for its living function and often emphasizes the manual creation. Though it does not show any significant tendency in its building types, by implication *gai* (蓋) has a wider range regarding building type. That is, *gai* (蓋) can go with any type of building. The flexibility in building types would then turn into a higher frequency in our data. As for the spoken data, there is no significant effect on the constraint of building types. It could be that the sample size for spoken data is not large enough to be representative in showing the tendency. Also, the encoded data is written for spoken, not completely spoken. Thus, it may limit the validity of the study. Even so, the outcome again suggests that *jian* (建) and *zao* (造) are more literary, and thus preferred in written language. On the contrary, *gai* (蓋) is considered a spoken use.

5.   Conclusion

This study reanalyzed the Chinese near-synonyms *jian* (建), *zao* (造), and *gai* (蓋) 'to build' at the semantic level and also word length. Differing from previous studies in Chinese near-synonyms, this study points out that there is genre variation among the uses of *jian* (建), *zao* (造), and *gai* (蓋) when they carry the same thematic object NP. The results indicate that word length of the object NP

and building types are the main factors that influence the variation. In terms of method, the study employed the program Rbrul for finding the influential factors of the variation. This technique enabled us to model interactions between independent variables and hierarchical (mixed effects) models, which helped the analyzers to find the most salient factor(s) within the variation. For Chinese instructors, it is suggested that the high frequency word – that is, *gai* (蓋) – should be taught prior to *jian* (建) and *zao* (造) when introducing the verbal concept of building. Moreover, word length is one of the ways by which the written language can be distinguished from the spoken.

References

Biber, Douglas. 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press.

Biber, Douglas. 1999. A register perspective on grammar and discourse: variability in the form and use of English complement clauses. *Discourse Studies* 1(2): 131–150.

Cheng, Xiang-Hui and Tian, Xiao-Ling. 2002. *Xiandai Hanyu 'Modern Chinese'* (現代漢語). Hong Kong: Joint Publishing Co. Ltd. [in Chinese]

Chang, Li-li. 2015. "The Formation of the Temporal Adverbs *Ji* (即), *Bian* (便), and *Jiu* (就)" ([即],[便],[就] 時間副詞功能的形成). *Language and Linguistics* 16(2): 139–168.

Clancy, Patricia. 1982. Written and spoken style in Japanese narratives. *Spoken and Written Language: Exploring Orality and Literacy*, ed. by D. Tannen. Norwood, 55–75. Norwood, N.J.: Ablex Publish Corp.

Duanmu, San. 2012. "Word-length preferences in Chinese: a corpus study". *Journal of East Asian Linguistics* 21.1: 89–114.

García, Christina. 2011. "Distinguishing Two 'Synonyms': A Variationist Analysis of quizá and quizás in Six Spanish Dialects". *Proceedings of the 5th Workshop on Spanish Sociolinguistics*, 103–112. MA: Cascadilla Proceeding Project.

Herring, Susan C., Paolillo, John C. 2006. "Gender and genre variation in weblogs". *Journal of Sociolinguistics* 10(4): 439–459.

Hung, Wei-Ting. 2010. "Analyze the near-synonyms: 'Manufacture' and 'Produce'". *Journal of Applied Chinese* 6: 223–246. [In Chinese]

Inkpen, Diana Zaiu and Hirst, Graeme. 2002. "Acquiring collocations for lexical choice between near-synonyms". *Proceedings of the ACL-02 workshop on Unsupervised lexical acquisition* 9: 67–76. Association for Computational Linguistics.

Iwasaki, Shoichi. 2015. A multiple-grammar model of speakers' linguistic knowledge. *Cognitive Linguistics* 26.2: 161–210. https://doi.org/10.1515/cog-2014-0101

Johnson, Daniel Ezra. 2009. "Getting off the Goldvarb standard: Introducing Rbrul for mixed effects variable rule analysis". *Language and Linguistics Compass* 3(1): 359–383.

Lakoff, George and Johnson, Mark. 1980. *Metaphors we live by*. University of Chicago Press.

Lee, Ching-Ying and Liu, Jyi-Shane. 2009. "Effects of collocation information on learning lexical semantics for near synonym distinction". *Computational Linguistics and Chinese Language Processing* 14 (2): 205–220.

Lee, Shi-Wen and Chen, Chiou-Mei. 1993. "A Comparative Study of Chinese Spoken and Written Language". *Teaching and Research* 15: 63–96.

Lim, Ni-Eng and Hong, Huaqing. 2012. "Intensifiers as Stance-Markers: A Corpus Study on Genre-Variations of Mandarin Chinese". *Chinese Language & Discourse* 3 (2): 129–166.

Liu, In-mao. 1999. "Character and word recognition in Chinese". In *Reading Chinese script: A cognitive analysis*, Jian Wang, Albrecht W. Inhoff and Hsuan-Chih Chen (eds), 173–187. NJ: Lawrence Erlbaum.

Liu, Mei-chun. 1999. "Lexical Meaning and Discourse Patterning: the Three Cases of Mandarin BUILD". In *Cognition and Function in Language*, Barbara Fox, Dan Jurafsky, and Laura A. Michaelis (eds.), 181–199. Stanford: Center for the Study of Language and Information (CSLI Publications).

Lewis, Molly L. and Frank, Michael C. 2016. "The length of words reflects their conceptual complexity". *Cognition* 153: 182–195.

Mikk, Jaan, Heli Uibo, and Jaanus Elts. 2001. "Word length as an indicator of semantic complexity." *Text as a linguistic paradigm: levels, constituents, construct* 187–195.

Norman, Jerry. 1988. *Chinese*. Cambridge University Press.

Su, Danjie. 2017. Semantics and chunking in written and conversational discourses: A corpus study of two near-synonymous words in Mandarin. *Chinese Language and Discourse* 8(1): 51–94. https://doi.org/10.1075/cld.8.1.03su.

Tao, Hongyin. 1999. Discourse taxonomies and their grammatico-theoretical Implications. *Dangdai Yuyanxue* [Contemporary Linguistics], 1.3:15–24.

Tsai, Mei-chih, Huang, Chu-Ren and Chen Keh-Jiamm. 1996. "A corpus based analysis on semantic extraction". *The 9th Computational Linguistics Conference*, 281–293. [in Chinese]

Tsai, Mei-chih. 2010. "*Tongyang* (同樣) and *xiangtong* (相同) are not *yiyang* (一樣): The Referential Differences of 'the Same' in Mandarin Chinese". *Journal of Chinese Language Teaching* 7(1): 51–79. [in Chinese]

Tsai, Mei-chih. 2011. "Convenient' during the process or as a result-event structure of synonymous stative verbs in TCSL". *Journal of Chinese Language Teaching* 8 (3):1–22. [in Chinese]

Wang, Zhi-Yi. 2012. "The lexical semantic features and syntactic functions of relatives of '*yinwei*' and '*youyu*'-A corpus-based approach". *Journal of Applied Chinese* 10: 203–236. [in Chinese]

Wilkinson, Endymion Porter. 2000. *Chinese history: a manual* (Vol. 52). Harvard University Asia Center.

Wu, Pei-Yen and Cheng, Ying. 2011. "Near Synonyms in Teaching Chinese as Second Language: A Case study to '*yi*', '*ju*', and '*yiju*' (從 [依],[據] 與 [依據] 談華語近義詞教學)". *Taiwan Journal of Chinese as a Second Language* 1:23–46.

Wu, Zhibiao and Palmer, Martha. 1994. "Verb semantics and lexical selection". In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics* 133–138.

Zou, Bin, Smith, Simon and Hoey, Michael. 2015. *Corpus Linguistics in Chinese Contexts*. Springer.

Appendix 1. Token exclusions

Contexts		Examples	
Proper names	JIAN	*Jianda gongsi* (建大公司) "*Jianda* company", *Jiantai xiangjiao chang* (建泰橡膠廠) "*Jiantai* rubber factory"	
	GAI	*Yuehan gai* (約翰．蓋) "person' name"	
Idiomatic phrases	ZAO	Shishi zao yingxiong (時勢造英雄) " The times creates heroes"	
	GAI	*gai huoguo* (蓋火鍋) "block shot"(a basketball terms)	
Other semantics	JIAN	Establish	*jian sheng* (建省) "establish the province "
		Suggest	*jianyi* (建議)"sugguest" *jianqing* (建請) "sugguest"(formal)
	ZAO	Manufacture	*zaoche*(造車) "manufacture cars" *zaoci* (造瓷) "manufacture porcelain"
		Invent	*zaozhi* (造紙) "invent paper"
		Commit (negative)	*zaoye* (造業), *zaonie* (造孽)" commit errors"
	GAI	Cover	*bigai* (筆蓋) "pen cover" *gai mianbei* (蓋棉被) "put on the quilt"
		Stamp	*gaizhang* (蓋章) "to stamp" *gaiyouchuo* (蓋郵戳) "stamp the postmark"
		Boasting	*Yuanqican bushi gaide*! (元氣餐不是蓋的) "the yuanqi set (meal) is awesome." *Ting Zhugaixian gai yi gai* (聽朱蓋仙蓋一蓋) "listen to Mr. Zhu 's boasting"
		Non-modern usage	*Gai genben zhi di, wen zhen hexie* (蓋根本之地，未臻和諧) "Therefore, the origin place is not completely steady or in harmony."

Appendix 2. Data coding schemes

Dependent variables	1. *JIAN*; 2. *ZAO*; 3. *GAI*
Independent variables	A. Word Length of Object NP
	'1': e.g. 屋;
	'2': e.g. 房子;
	'3' or above: e.g. 停車場, 商辦大樓.
	B. Preverbal locative phrase
	P=with preverbal locative phrase;
	N=without preverbal locative phrase.
	C. Builing types
	R=Residential
	NR=Non-residential

Constraints on the collocational behaviors of Chinese near-synonyms

A corpus-based analysis

Jia Yang

University of Dayton

This paper reports a corpus-based study to examine how Chinese near-synonyms choose their typical collocates. Near-synonyms commonly misused by English-speaking learners of Chinese were selected for analysis. Results obtained from the corpora (the Chinese Internet Corpus by the University of Leeds and the Lancaster corpus of Mandarin Chinese) indicate that the collocational behaviors of the selected synonyms are constrained by their own semantic, grammatical, prosodic, stylistic and pragmatic features and hence are explainable to second/foreign learners. Findings of this study will contribute to the design of collocation/synonym dictionary as well as the instruction of collocations as a second/foreign language.

Keywords: collocation, corpus-based study, near-synonyms, L2 vocabulary learning

1. Introduction

Collocation has been studied for at least five decades (Xiao & McEnery 2006). While the importance of collocation in language learning has been well recognized during the last three decades, there has still been a taken-for-granted assumption that collocation is arbitrary so that teachers can do little to help learners develop knowledge of L2 collocation. Lewis (1997: 29), one of the influential writers in the field of English language teaching, argued, "Collocation is not determined by logic or frequency, but is arbitrary, decided only by linguistic convention." Under this assumption, a number of researchers believe that teachers in many cases can offer no explanations to students for the particular choices selected by the language community, since this is simply the way the language is (Woolard 2000: 34).

However, findings from corpus linguistic studies have shown that collocational process may not be arbitrary. Walker (2008) demonstrated how English nouns and verbs chose their characteristic collocates that fit their semantics, etymology, semantic prosody, and prototypical literal sense. Xiao and McEnery (2006) conducted a cross-linguistic study on the collocational behavior and semantic prosody of near-synonyms in English and Chinese. Their studies showed that near-synonyms were not interchangeable in either language, as they tended to collocate with words that have different semantic preferences. In other words, the collocational behaviors of near-synonyms were constrained by their semantic preferences. However, in addition to semantic preferences, collocational behaviors of Chinese words might be constrained by other factors. In order to identify those factors, a corpus-based study was conducted on a group of near-synonyms commonly misused by English-speaking learners of Chinese, as shown in my data. Two corpora were used for the analysis: the Chinese Internet Corpus created by the University of Leeds (CICL) and the Lancaster corpus of Mandarin Chinese (LCMC). The Chinese Internet Corpus contained 280 million words compiled from the Internet since February 2005. The Lancaster Corpus has been constructed using written Mandarin Chinese texts published in Mainland China.

2.   Literature review

As one of the important topics in linguistics, "collocation" has been studied under different theoretical frameworks, resulting in various definitions and terminologies. Given these variations, it would be better to trace the developmental history of collocation research and generate a working definition of collocation for the purpose of this study. "Collocation" was first used as a technical term in linguistics by Firth (1957) in his *Modes of Meaning.* His ideas on collocation were summarized as three-fold: **(1)** "Collocations of a given word were statements of the habitual or customary places of that word," for example, certain types of words are more likely to appear with *strong* than *powerful* and vice versa. Castello (2014) examined data collected from the spoken components of the Bank of English (BoE) and suggested that *powerful* is more closely associated with concrete nouns, such as cars, computers, countries, and states, whereas *strong* conveys "much more of a metaphysical connotation than *powerful*," which is usually used to describe "forms of communication" and "emotions and expressions." **(2)** The habitual collocations in which a word appears defined part of that word's meaning, so that "you shall know a word by the company it keeps (Firth 1968: 179)." Accordingly, Firth claimed that one of the meanings of *peer* was defined through its collocation with *school* as in *school peers*. Similarly, one of the meanings of *dark* was defined

by its collocatability with *night,* since *dark* was characteristically used in conjunction with *night.* **(3)** Habitual collocation can also be viewed as a type of "mutual expectancy" between words (Firth 1968: 181), i.e. collocating words can predict one another in the way that if one has been found, we can predict to find the other.

Many linguists, including Halliday (1966), Greenbaum (1974: 82), Sinclair (1991) and Hoey (1991, 2000) have accepted firth's notion of collocation. Their statements of collocation are essentially quantitative and based on frequency. Greenbaum (1974: 82) referred collocation to "a frequent co-occurrence of two lexical items in the language," but he did not specify how frequent the co-occurrence should be. Hoey (1991: 6–7) put one step-forward and defined collocation as "the relationship a lexical item has with items that appear with greater than random probability in its textual context." Influenced by the concept of "mutual expectancy," Jones and Sinclair (1974: 19) claimed that words were collocates of each other, if "they were found together more often than their individual frequencies would predict." Based on the frequency notion and the "mutual expectancy" concept, this study proposed a working definition of collocation as a linguistic phenomenon where words co-occurred frequently in a sample of language so that the presence of one makes the presence of the other more likely than it would otherwise be. Idioms, slangs, and proverbs are not included in the analysis of this study, as their inner components cannot be altered, substituted, deleted from, and inserted. For example, the meaning of the idiom 班门弄斧 *bānmén nòngfǔ* (showing off one's skills at the presence of experts) cannot be inferred from its inner components, and each of its components cannot be altered, substituted, deleted from and inserted. The choice of its components is rather conventional.

3.   Methodology

3.1   Selecting near-synonyms for analysis

Li (2016) investigated the collocational errors in compositions written for the HSK (The Chinese language proficiency test) and reported that the misuse of synonyms was one of the leading causes of the collocational errors. Accordingly, our study selected eight near-synonym pairs which were usually misused by American learners of Chinese to examine how these synonyms choose their typical collocates. The eight synonym pairs were identified from 44 writings collected by the researcher. Twenty-two advanced learners of Chinese produced the collected writings. The learners' average length of Chinese classroom instruction was 4.12 years. The eight selected near-synonym pairs were the most frequently misused near-synonym pairs in the 44 collected writings: 帮忙 *bāngmáng* and 帮助 *bāngzhù* (help), 在

zài and 正在 *zhèngzài* (adverbs to indicate that an action is going on), 保持 *bǎochí* and 维持 *wéichí* (keeping or maintaining an existing state or situation), 家 *jiā* and 家庭 *jiātíng* (home/family), 目光 *mùguāng* and 眼光 *yǎnguāng* (eye light), 眼泪 *yuǎnlèi* and 泪 *lèi* (tears), 乘车 *chéngchē* and 坐车 *zuòchē* (to ride in a car), and 能 *néng,* 会 *huì*, and 可以 *kěyǐ* (modal verbs).

3.2   Identifying typical collocates

A number of statistical tests have been used in corpus linguistics to measure collocation. The most popular ones are T-scores and mutual information scores. Mutual information (MI) is a statistical formula borrowed from information theory. The MI score is a measure of collocational strength. The higher the MI score is, the stronger the link between two items. The MI score is computed by dividing the observed frequency of the co-occurring word in the defined span[1] by the expected frequency of the co-occurring word in that span and then taking the logarithm to the base two of the result. MI score is not dependent on the corpus size (Hunston 2002) and tends to overvalue infrequent words (Xiao & McEnery 2006: 105). In this regard, this study adopted T-score to identify meaningful collocations. T-score measures the confidence that the association between two words is really due to the variable and not just due to chance. Any collocate with a T-score of two or above is regarded as significant, since this indicates that its combination with the node word is not simply the result of chance (Walker 2008). Li (2016) pointed out that the vast majority of collocational errors made by American learners of Chinese were "verb + noun" collocations. Therefore, this study used a 0:1 widow span to identify the typical collocates of verbs and modal verbs. A 1:1 widow span was used to identify the typical collocates of nouns and adverbs, so that "verb + noun" "adjective + noun" and "adverb + verb" collocations were examined. Since the purpose of the study was not to generate a comprehensive list of the selected words' collocates, only collocates immediately adjacent to the node word were examined.

3.3   Data for the study

Data for analysis was drawn from two large-scale corpora. The primary one was the Chinese Internet Corpus created by the University of Leeds (CICL). This Corpus contained 280 million words compiled from the Internet since February 2005. During the compiling process, 500 frequent words in Chinese were selected.

---

1. For example, a 2:2 window span is namely two words to the left and two words to the right of the node word for the search string

Following that, 5000–8000 queries were generated. Each query contained four words from the 500 frequent words. Queries were sent to a search engine and the top 10 URLs returned for each query were collected (Sharoff 2006). The language uses on the Internet might be different from daily language uses in the way that new word choices would be made to create certain rhetoric effects. Accordingly, the Lancaster Corpus of Mandarin Chinese (i.e. LCMC) was used as a supplementary corpus to ascertain whether the results obtained from the Internet Corpus were consistent with the word choices in other genres of the Chinese language. The Lancaster Corpus was selected because it has been constructed using written Mandarin Chinese texts published in Mainland China to ensure some degree of textual homogeneity. It contained approximately one million words of samples collected from fifteen different written text categories, such as editorials, reviews, reportage, fictions, academic prose, etc. (Xiao & McEnery 2006). Since the size of the Lancaster Corpus was much smaller than the Internet corpus, the T-scores obtained from the Lancaster Corpus were much smaller than the Internet corpus. Some of them were less than two. However, when the search word appeared frequently in both corpora, the results obtained from the two corpora were consistent. This indicates that collocations extracted from the Internet corpus were reliable.

4. Constraints on the collocational behaviors of synonyms

Results obtained from the data show that collocational behaviors of the selected near-synonyms pairs are constrained by their own grammatical, semantic, prosodic, stylistic and pragmatic characteristics. Word choice in a given context is not a completely arbitrary process and is explainable by the following constraints.

4.1 Grammatical constraints

One of the constraints is the grammatical feature of the word. This constraint was exhibited by the collocational behaviors of the near-synonym pair 帮忙 *bāngmáng* and 帮助 *bāngzhù*. They share the meaning of "to help," but show distinctive syntactic features, due to their different morphological structures. 帮助 *bāngzhù* is a parallel verb compound, whose components "either are synonymous or signal the same type of predicative notions" (Li & Thompson 1989: 68). As a parallel verb compound, 帮助 *bāngzhù* often takes a noun phrase as its direct object, as shown in Example (1). On the other hand, 帮忙 *bāngmáng* is a verb-object compound, which allows its constituents to be separated, such as 帮妈妈的忙 *bāng māma de máng* (help mum). As pointed out by Li and Thompson (1989), the majority of verb-object compounds do not take a direct object. Therefore, when 帮忙

bāngmáng occurs by itself, it is usually followed by the action operated by the helper, e.g., Example (2), rather than the receiver of the help, as in Example (3). A possessive phrase is needed to precede 忙 *máng* to introduce the direct object, e.g., Example (4). When used as a whole, 帮忙 *bāngmáng* does not take a direct object.

(1) 他 帮助　　我 搬家。
　　Tā bāngzhù wǒ bānjiā.
　　He help　　me move house
　　'He helps me move.'

(2) 他 帮忙　　　搬家。
　　Tā bāngmáng bānjiā.
　　He help　　　move house
　　'He helps move.'

(3) **他 常常　　　帮忙　　我。
　　Tā chángcháng bāngmáng wǒ.
　　He often　　　help　　me.
　　'He helps me.'

(4) 他 常常　　　帮 我 的　忙。
　　Tā chángcháng bāng wǒ de　máng.
　　He often　　　help me ᴘᴏss favor
　　'He often helps me.'

Due to this grammatical difference, 帮忙 *bāngmáng* and 帮助 *bāngzhù* collocate with different types of words. 帮忙 *bāngmáng* usually co-occurs with action verbs, while 帮助 *bāngzhù* can collocate with receiver(s) of the help, or actions verbs when the receiver of the help is omitted. For example, in the sentence "请 大家帮助测试 *qǐng dàjiā bāngzhù cèshì* (Everyone, please help (us) test!)," 帮助 *bāngzhù* is followed by the action of the help "测试 *cèshì*," while the receiver of help is not specified.

4.2　Semantic constraints

Typical collocates with a word are usually constrained by the semantic features of the word. For example, 保持 *bǎochí* and 维持 *wéichí* share the meaning of "keeping or maintaining an existing state or situation." However, 保持 *bǎochí* suggests that the good level of the existing state or situation should not disappear or decrease, while 维持 *wéichí* emphasizes the continuous existence of the current state or situation. Typical collocates with 保持 *bǎochí* often imply the meaning of maintaining a good situation or state, such as 保持身材 *bǎochí shēncái* (keep fit), 保持健康 *bǎochí jiànkāng* (stay healthy) and 保持清洁 *bǎochí qīngjié* (keep clean).

Table 1. Typical collocates with 帮助 *bāngzhù* and 帮忙 *bāngmáng**

Typical collocates with 帮助 *bāngzhù*					Typical collocates with 帮忙 *bāngmáng*				
	CICL freq.	CICL T-score	LCMC freq.	LCMC T-score		CICL freq.	CICL T-score	LCMC freq.	LCMC T-score
他们 *tāmén* they	967	33.41	7	2.56	找 *zhǎo* find	130	11.09	设计 *shèjì* design	1	0.99
你 *nǐ* you	960	32.86	3	1.48	做 *zuò* do	115	10.25	测试 *cèshì* test	1	0.99
他 *tā* he	893	26.27	4	1.85	看 *kàn* look, see	103	9.23	–	–	–
我们 *wǒmén* we	882	26.4	3	1.22	解决 *jiějué* resolve	71	8.35	–	–	–
别人 *biérén* others	619	24.95	–	–	推荐 *tuījiàn* recommend	50	6.92	–	–	–

* The t-scores in this table were obtained through within a 0–1 window span (0 word to the left and 1 word to the right of the search word). Due to the low occurrences of 帮忙 *bāngmáng* in the Lancaster corpus, only two collocates with 帮忙 *bāngmáng* were extracted from the Lancaster corpus. Because of the smaller size of the Lancaster corpus, most T-scores obtained from the Lancaster corpus were less than two, and much lower than the T-scores obtained from the CICL corpus.

维持 *wéichí* indicates the maintenance of an existing state or situation regardless of whether it is good or bad. This semantic difference explains the collocations extracted from the CICL corpus. As shown in Table 2, typical words collocated with 保持 *bǎochí* are words implying the good nature of a state/situation, such as advanced (先进 *xiānjìn*), good (良好 *liánghǎo*), stable (稳定 *wěndìng*) or sober (清醒 *qīngxǐng*). On the other hand, typical collocates with 维持 *wéichí* are usually neutral words, with no indication of the nature of the existing state/situation.

Table 2. Typical collocates with 保持 *bǎochí* and 维持 *wéichí*

Typical collocates with 保持 *bǎochí*			Typical collocates with 维持 *wéichí*		
	CICL freq.	CICL T-score		CICL freq.	CICL T-score
先进性 *xiānjìn xìng* advanced	496	28.78	秩序 *zhìxù* order	448	24.22
良好 *liánghǎo* good	347	21.75	生计 *shēngjì* living	345	21.55
稳定 *wěndìng* stable	224	19.42	生活 *shēnghuó* life	346	21.55
清醒 *qīngxǐng* sober	215	17.22	现状 *xiànzhuàng* current situation	194	14.54
平衡 *pínghéng* balance	160	15.73	原判 *yuánpàn* original judgment	140	13.34

* The T-scores in this table were obtained through within a 0–1 window span (0 word to the left and two words to the right of the node word). Due to the low occurrences of 保持 *bǎochí* and 维持 *wéichí* in the Lancaster Corpus, only the results from the Internet Corpus created by the University of Leeds were included.

The collocational behaviors of 在 *zài* and 正在 *zhèngzài* are constrained by their semantic features as well. According to Lü et al. (1999: 672), 在 *zài* and 正在 *zhèngzài* are both adverbs to indicate that an action is going on. Due to the interaction of the meaning of 正 *zhèng* with 在 *zài,* 正在 *zhèngzài* can only be used to express that an action is going on at a specific time; while 在 *zài* can be used during a period of time. Influenced by this semantic difference, 正在 *zhèngzài* is not used to express repetitive-progressive, e.g. Example (5), continuous-progressive, e.g. Example (6) or habitual progressive, e.g. Example (7).

(5) *他 又　正在　　　看　电视。
　　Tā yòu zhèngzài kàn diàn shì.
　　He again ZHENGZAI watch TV
　　'He is watching TV again.'

(6) *我 一直　　正在　　　找　　你。
　　Wǒ yìzhí zhèngzài zhǎo nǐ.
　　I　all along ZHNGZAI look for you.
　　'I am looking for you all along.'

(7) *他 老是 正在　　　　问 这些 问题
　　Tā lǎoshì zhèngzài wèn zhèxiē wèntí.
　　He always ZHENGZAI ask these questions
　　'He is always asking these questions.'

Table 3. Typical collocates with 在 *zài* and 正在 *zhèngzài**

Typical collocates with 在 *zài*				Typical collocates with 正在 *zhèngzài*				
	CICL freq.	CICL T-score	LCMC freq.	LCMC T-score		CICL freq.	CICL T-score	LCMC freq.	LCMC T-score
都 *dōu* all	983	86.92	85	5.8	这时 *zhèshí* at this time	634	26.47	3	1.7
还 *hái* still	969	79.13	87	7.22	目前, *mùqián* cur-rently	873	29.42	3	1.7
一直 *yīzhí* all along	842	58.15	15	3.25	也 *yě* too	791	20.9	2	1.16
又 *yòu* again	172	17.64	8	2.65	现在 *xiànzài* now	975	32.93	1	0.86
常 *cháng* often	41	4.46	6	2.27	当时 *dāngshí* then, at that time	409	19.68	1	0.93

*　The T-scores in this table were obtained through within a 1–1 window span (one word to the left and one word to the right of the node word). Since the size of the Lancaster corpus was much smaller than the Internet corpus, the T-scores obtained from the Lancaster corpus were much smaller than the Internet corpus.

Constrained by the semantic feature of iteration, 在 *zài* frequently co-occurs with iterative adverbials which indicate repetition of an event/action, such as 又 *yòu* (again), 还 *hái* (still), or habitual adverbials, such as 都 *dōu* (all), 一直 *yīzhí* (all

along, continuously)，常 *cháng* (often). 正在 *zhèngzài,* on the other hand, are more likely to collocate with adverbials with a specific time reference, such as 这时 *zhèshí* (at this time), 目前 *mùqián* (currently), 现在 *xiànzài* (now), or 当时 *dāngshí* (then, at that time).
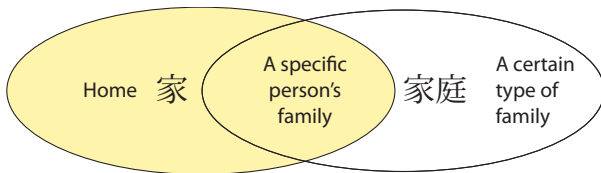
Figure 1. The semantic scope of 家 *jiā* or 家庭 *jiā tíng*

Table 4. Typical collocates with 家 *jiā* and 家庭 *jiātíng*[*]

Typical collocates with 家 *jiā*			Typical collocates with 家庭 *jiātíng*		
	CICL freq.	CICL T-score		CICL freq.	CICL T-score
我 *wǒ* I	1310	113.2	贫困 *pínkùn* poor	267	18.79
我们 *wǒmén* we	1024	78.68	单亲 *dānqīn* single-parent	230	15.55
她 *tā* she	964	58.54	普通 *pǔtōng* normal	198	16.41
邻居 *línjū* neighbour	824	28.77	农民 *nóngmín* farmer	172	13.70
姥姥 *lǎolao* grandma	439	21.39	工薪 *gōngxīn* salary	55	11.31

* The T-scores in this table were obtained through within a 1–1 window span (1 word to the left and 1 word to the right of the node word). Due to the few occurrences of 家庭 *jiātíng* in the Lancaster corpus, Table 4 only included the collocations extracted from the Internet corpus.

Collocational behavior is also constrained by the semantic scope of a word. The semantic scope of a word is "the possible range of interpretation implied by the literal use of a word" (Gerstl 1991: 168). For example, 家 *jiā* and 家庭 *jiātíng* are both translated as "family" in many English-Chinese dictionaries. As a result, learners of Chinese tend to use either 家 *jiā* or 家庭 *jiātíng* interchangeably when they try to express the concept of "family." However, the two words exhibit different collocational behaviors due to their different semantic scopes (see Figure 1).

家 *jiā* normally refers to a specific person's home or family. Consequently, it usually collocates with pronouns, names, or kinship terms, such as 我家 *wǒjiā* (my home/family) 你家 *nǐjiā* (your home/family), 奶奶家 *nǎinǎi jiā* (grandma's home/family), or 小王家 *xiǎowáng jiā* (Little Wang's home/family). 家庭 *jiātíng* refers to either a specific person's family or a certain type of family. Results obtained from the corpora show that typical collocates with 家庭 *jiātíng* are often characteristics of a certain type of family, such as 单亲家庭 *dānqīn jiātíng* (single-parent family), 贫困家庭 *pínkùn jiātíng* (poor family), 农民家庭 *nónmín jiātíng* (farmer's family). Job titles can collocate with either 家 *jiā* or 家庭 *jiātíng*, but the co-occurrence of job title and the two words denote different meanings, for example, 经理家 *jīnglǐ jiā* refers to a specific manager's home, while 经理家庭 *jīnglǐ jiātíng* refers to a type of family, whose head of the household is a manager.

Additionally, collocational behaviors of a word are influenced by its metaphoric meaning. Learners of Chinese sometimes generate erroneous collocations, such as 我看见你的观点 *wǒ kànjiàn nǐ de guāndiǎn* (I see your point). The different metaphoric meaning of "seeing" in English and Chinese causes this erroneous use. "Seeing" in English is usually expressed metaphorically as "understanding." In this sense, "I see" is often used to indicate "I understand." The Chinese equivalent of seeing 看见 *kànjiàn* does not convey such metaphoric meaning. As a result, it cannot collocate with abstract words, such as 观点 *guāndiǎn* (viewpoint), or 想法 *xiǎngfǎ* (thought).

Yu (2004: 666) pointed out that, in Chinese, seeing with one's eyes is often expressed metaphorically in terms of the light extending from the eyes to the target. Based on this conceptual metaphor, 目光 *mùguāng* and 眼光 *yǎnguāng* 'eye-light' tend to collocate with motion verbs. The eye light can be thrown from our eyes to the objects we are seeing, therefore the most frequent collocate with 目光 *mùguāng* and 眼光 *yǎnguāng* is 投向 *tóuxiàng* 'throw at' (see Table 5). The direction of the 'eye light' can be controlled and directed to another target. Thus, 目光 *mùguāng* and 眼光 *yǎnguāng* can collocate with words like 转向 *zhuǎnxiàng* 'change direction' and 移开 *yíkāi* 'to move away.' After the "eye light" extends from the eye, it can be taken back. In this sense, collocations, such as 收回目光/眼光 *shōuhuí mùguāng/yǎnguāng* 'to take back or withdraw eye light,' were observed in the corpora.

When two persons look at each other, their eye lights can bump into each other. Thus, 目光 *mùguāng* and 眼光 *yǎnguāng* can collocate with words like 相遇 *xiānyù* or 碰撞 *pèngzhuàng* 'to bump into each other.' Since the 'eye light' can also be described by the distance or depth it traveled, 目光 *mùguāng* and 眼光 *yǎnguāng* can collocate with words depicting how far the 'eye light" traveled, such as, 目光/眼光短浅 *mùguāng/ yǎnguāng duǎnxiǎn* (short-sighted) and 目光/眼光长远 *mùguāng / yǎnguāng chángyuǎn* (far-sighted).' The 'eye light' extending from

Table 5. Typical collocates with 目光 *mùguāng* and 眼光 *yǎnguāng**

Typical collocates with 目光 *mùguāng*			Typical collocates with 眼光 *yǎnguāng*		
	CICL freq.	CICL T-score		CICL freq.	CICLT-score
投向 *tóuxiàng* throw at	445	21.33	投向 *tóuxiàng* throw at	72	8.66
转向 *zhuǎnxiàng* change direction	268	16.53	长远 *chángyuǎn* far and long	44	6.62
收回 *Shōuhuí* take back or withdraw	115	10.36	独到 *dúdào* special, original	42	6.48
相遇 *xiān yù* bump into	104	10.32	转向 *zhuǎn xiàng* change direction	38	6.21
移开 *yíkāi* move away	87	9.96	收回 *shōuhuí* take back or withdraw	22	4.97
停留 *tíngliú* stay, stopover	86	9.55	短浅 *duǎnxiǎn'* short and shallow	21	4.69
集中 *jízhōng* focus	80	9.15	移开 *yíkāi* move away	21	4.68
锐利 *ruìlì* sharp	56	8.05	敏锐 *mǐnruì* acute	19	4.34

* The T-scores in this table were obtained through within a 1–1 window span (1 word to the left and 1 word to the right of the node word). Since 目光 *mùguāng* and 眼光 *yǎnguāng* appeared very few times in the Lancaster corpus, Table 6 only listed collocates obtained from the CICL corpus.

the eyes has force so that it has a sharp point and can penetrate fast like a dagger. In this regard, 目光 mùguāng and 眼光 *yǎnguāng* can collocate with words like 锐利 *ruìlì* (sharp) and 敏锐 *mǐnruì* (acute). 眼光 *yǎnguāng* also metaphorically indicates one's viewpoint. Accordingly, it can collocate with adjectives describing viewpoints, such as 眼光独到 *yǎnguāng dúdào* (unique viewpoint).

4.3 Prosodic constraints

Prosodic features, particularly those related to rhythm, have important effects on the collocational behaviors of Chinese words. According to Yip and Don (2004: 379), Chinese has gradually developed a preference for disyllabic rhythms due to the disyllabic dominance of the lexical items in the vocabulary. Feng (2009: 149) further explained, "Monosyllabic words in classical Chinese must be used in a disyllabic template in order to be considered grammatical in modern Chinese." Based on this argument, 泪 *lèi* (tears) as a monosyllabic word in classical Chinese must be used in a disyllabic template. As a result, it usually collocated

Table 6. Typical collocates with 泪 *lèi* and 眼泪 *yǎnlèi*[*]

Typical collocates with 泪 *lèi*			Typical collocates with 眼泪 *yǎnlèi*		
	CICL freq.	CICL T-score		CICL freq.	CICL T-score
流 *liú* flow	826	23.71	流 *liú* to flow	632	27.13
滴 *dī* to drop	465	22	掉 *diào* to drop	626	26.4
流下…来 *liúxià… lái* flow down	190	14.17	擦 *cā* to wipe	389	20.43
落下…来 *luòxià… lái* fall	150	12.63	抹 *mǒ* to wipe up	296	17.61
清 *qīng* pure	138	11.74	擦干 *cāgān* to wipe and dry	157	13.82
洒 *sǎ* sprinkle	103	10.66	流下 *liúxià* to flow down	150	12.95
掉下…来 *diàoxià… lái* to drop down	99	10.33	没有 *méiyǒu* no	149	12.64
落 *luò* fall	87	8.44	掉下 *diàoxià* to drop	80	11.73

* The T-scores in this table were obtained through within a 1–1 window span (1 word to the left and 1 word to the right of the node word). Due to the very few occurrences of 泪 *lèi and* 眼泪 *yuǎnlèi* in the Lancaster corpus, Table 6 only included the collocations extracted from the CICL corpus.

with monosyllabic words, such as 流 *liú* (to flow), 滴 *dī* (to drop), 清 *qīng* (pure), 洒 *sǎ* (sprinkle), or 落 *luò* (fall). When 泪 *lèi* (tears) collated with a disyllabic word, it was followed by a complement to position it in a disyllabic template. As shown in Table 6, a complement, such as 来 *lái,* must co-occur with 泪 *lèi* (tears), when 泪 *lèi* collocates with disyllabic words, such as 流下 *liúxià* (flow down), 落下 *luòxià* (fall), and 掉下 *diàoxià* (drop). The disyllabic word 眼泪 *yǎnlèi* (tears) could collocate with either a monosyllabic word or a disyllabic word. As reported by Yip and Don (2004: 385), "monosyllabic verb + disyllabic object" follows the rhythmic structure of "Xx + X." Therefore, when 眼泪 *yǎnlèi* (tears) collocates with a monosyllabic word, 泪 *lèi* is weakened so that the whole collocation still maintained a disyllabic template.

4.4   Stylistic constraints

Stylistic constraints refer to the register preference of a word. In this study, register is analyzed in terms of the type of activities in which the linguistic forms operate (e.g., professional and unprofessional, formal and informal) and the channels of communication (prototypically speech and writing). In general, a word tends to collocate with lexical items that have similar register preference. For example, 坐车 *zuòchē* and 乘车 *chéngchē* both mean, "to ride in a vehicle," but 乘车 *chéngchē* is more favored in written language while 坐车 *zuòchē* is a colloquial word. Due

Table 7. Typical collocates with 坐车 *zuòchē* and 乘车 *chéngchē**

Typical collocates with 坐车 *zuòchē*					Typical collocates with 乘车 *chéngchē*				
	CICL freq.	CICL T-score	LCMC freq.	LCMC T-score		CICL freq.	CICL T-score	LCMC freq.	LCMC T-score
去 *qù* go to	355	19.96	105	5.8	路线 *lùxiàn* itinerary	166	13.41	2	1.7
回家 *huíjiā* go home	101	9.99	124	7.22	前往 *Qiánwǎng* leave for	145	12.07	2	1.7
回 *huí* return	80	8.94	34	3.25	赴 *fù* Go to	61	7.79	1	1.16
回去 *huíqù* go back	41	6.42	16	2.65	返回 *fǎnhuí* come back	46	6.97	1	0.86

* The T-scores in this table were obtained through within a 1–1 window span (1 word to the left and 1 word to the right of the node word).

to this stylistic difference, 乘车 *chéngchē* tends to collocate with formal written words, but typical collocates of 坐车 *zuòchē* are words favored in spoken language (see Table 7).

4.5 Pragmatic constraints

Misuses of "能 *néng*, 会 *huì*, 可以 *kěyǐ*" have often been observed in learners' speech and writings. Previous research of "能 *néng*, 会 *huì*, 可以 *kěyǐ*" mainly focused on their semantic differences and syntactic distributions. For example, Lü et al. (1999) suggested that "能 *néng*" usually indicates ability, while "可以 *kěyǐ*" is preferred for the expression of possibility. Fu and Zhou (1990) investigated the distributions of "能 *néng*, 会 *huì*, 可以 *kěyǐ*" in play scripts and found that when expressing "possibility", "能 *néng*" is preferred in interrogative sentence (64.3%) to affirmative sentences (6%); whereas "会 *huì*" is used more frequently in affirmative sentences (45%), and less frequently in interrogative sentences (20.2%). Lai (2006) reported that both "能 *néng*" and "会 *huì*" can express the meaning of possibility. However, when "能 *néng*" and "会 *huì*" express the meaning of possibility in interrogative sentences, "会 *huì*" is often used with third personal pronouns while "能 *néng*" can be used with any pronouns. Very few studies have attempted to analyze the uses of "能 *néng*, 会 *huì*, 可以 *kěyǐ*" from the pragmatic perspective. One such study is by Xiong (2008). Xiong (2008) suggested that the uses of "能 *néng*, 会 *huì*, 可以 *kěyǐ*" are related to what subjective attitude, prediction and judgment the speaker wants to convey. However, Xiong (2008)'s claim was not based on statistical analysis of the large copra. More evidence drawn from corpora should be provided to analyze the communicative functions conveyed by "能 *néng*, 会 *huì*, 可以 *kěyǐ*." Moving further in this direction, this study conducted a corpus-based analysis to examine how the attitude, prediction and judgment the speaker intends to convey affects the collocational behaviors of "能 *néng*, 会 *huì*, 可以 *kěyǐ*."

Table 8. The pragmatic feature of 能 *néng*, 会 *huì*, 可以 *kěyǐ*

Meaning	Communicative function	Modal verbs	Speakers' perspective
Possibility	Making predictions or hypothesis	能/可以 *néng/kěyǐ*	The speaker wishes the predictions or hypothesis to be possible
		会 *huì*	The speaker has no preferences

"能 *néng*, 会 *huì*, 可以 *kěyǐ*" share the meaning clusters of possibility, permission and ability/skill while each member conveys different perspectives and/or communicative functions. Since the focus of this paper is not a comprehensive description of the use of "能 *néng*, 会 *huì*, 可以 *kěyǐ*," only the meaning cluster of

possibility was discussed in this paper. The meaning cluster of possibility indicates that "能 *néng*, 会 *huì*, 可以 *kěyǐ*" can all be used to make predictions or hypothesis. Predictions are statements or opinions about the future. Hypothesis refers to statements or opinions about a hypothetical situation. This hypothetical situation can be contrary to the fact at present. Further analysis of the results reveals that the choices of "能 *néng*, 会 *huì* and 可以 *kěyǐ*" imply speakers' different attitudes toward the prediction. 能 *néng* and 可以 *kěyǐ* indicate that the subjective and/or objective circumstances allow the prediction or the hypothetical situation to be possible (see Table 8). As shown in Example (8a), 能 *néng* and 可以 *kěyǐ* imply that the speaker wishes the prediction or the hypothesis to be possible. 会 *huì* indicates that the prediction or hypothesis is likely to happen, but the speaker has no preferences to whether the prediction or hypothesis would become true, see Example (8b) and (9a). (9b) is not preferred in the Chinese language, because the speaker does not favor the effect of smoking on health.

(8) a. 经常   运动   能/可以 增进   健康。
         Jīngcháng yùndòng néng/kěyǐ zēngjìn   jiànkāng.
         Often   workout can   improve health
         'Regular workouts can improve health.'
         (The speaker hopes that health will be improved.)

     b. 经常   运动   会 增进   健康。
         Jīngcháng yùndòng huì zēngjìn   jiànkāng.
         Often   workout will improve health
         'Regular workouts will improve health.'
         (The speaker only states the possibility. His/her attitude toward the prediction is neutral.)

(9) a. 吸烟   会 影响   健康。
         xīyān   huì yǐngxiǎng jiànkāng.
         Smoking will affect   health
         'Smoking will affect health.'
         (The speaker states the possible influence of smoking on health.)

     b. *吸烟   能/可以 影响   健康。
         Xīyān   néng/kěyǐ yǐngxiǎng jiànkāng.
         Smoking can   affect   health
         'Smoking can affect health.'

The obtained data from the corpora indicates that the pragmatic preferences conveyed by 能 *néng* and 会 *huì* determine their distinctive behaviors of collocation. As shown in Table 9, 会 *huì* tends to collocate with negative or neutral words, while 能 *néng* tends to collocate with positive words, as 能 *néng* tends to convey the message that the speaker wishes the hypothesis or prediction to be possible.

Table 9. Typical collocates with 能 *néng* and 会 *huì**

Typical collocates with 能 *néng*					Typical collocates with 会 *huì*				
	CICL freq.	CICL T-score		LCMC freq.	LCMC T-score		CICL freq.	CICL T-score		LCMC freq.	LCMC T-score

能 *néng* word	CICL freq.	CICL T-score		LCMC freq.	LCMC T-score	会 *huì*	CICL freq.	CICL T-score		LCMC freq.	LCMC T-score
希望 *xīwàng* wish	1087	69.62	得到 *Dédào* get	13	3.51	出现 *chūxiàn* appear	1064	68.48	出现 *chūxiàn* appear	23	4.76
找到 *zhǎodào* find	977	56.62	理解 *lǐjiě* understand	11	3.23	发现 *fāxiàn* discover	961	63.82	产生 chǎn sheng produce	15	3.75
得到 *dédào* get	958	48.94	发挥 *fāhuī* play	10	3.09	影响 *yǐngxiǎng* influence	960	49.46	影响 *yǐngxiǎng* influence	14	3.55
做到 *zuòdào* achieve	943	45.49	做到 *zuòdào* to live up to	7	2.61	造成 *zàochéng* bring about	955	38.88	发现 *fāxiàn* discover	12	3.33 2.9
理解 *lǐjiě* understand	931	39.66	满足 *mǎnzú* satisfy	6	2.40	导致 *dǎozhì* lead to	972	38.43	引起 *yǐnqǐ* arouse	9	2.92
达到 *dádào* reach	927	33.62	促进 *cùjin* promote	6	2.36	引起 *yǐnqǐ* arouse	969	37.75	造成 *zàochéng* bring about	9	2.91

* The T scores in this table were obtained through within a 0–1 window span (0 word to the left and 1 word to the right of the node word).

5.   Discussion and pedagogical implications

When we acquire our native language, each encounter with a word occurs in a meaningful context. the subsequent encounters reinforce what we have known about that word, and help us narrow down the contexts where the word can be used. In this way, we gradually develop a "feel" for what kinds of linguistic and social contexts the word tends to inhabit. However, second/foreign language learners do not have the luxury of abundant exposures to words over time and in a variety of meaningful contexts. To them, new words are usually introduced through definitions, glosses, and are related to synonyms in the second/foreign language itself. Sometimes, learners tend to equate the new word to an alternative for a word they have already known. As a result, synonyms that are undifferentiated (i.e., translated as the same word) in learners native language are among the most difficult lexical items for second/foreign learners to master. Liu et al. (2014) examined the use of four sets of synonyms by intermediate/advanced Chinese EFL/ESL learners and native English speakers. Their results revealed that the acquisition of synonyms is mainly the learning of their salient collocates. Meanwhile, communicative purpose or context plays an important role in synonym selection. These conclusions suggest that the key for second/foreign language learners to manage synonyms is to help them learn the salient collocates, and expose the learners to synonyms and their salient collocates in meaningful contexts. Those contexts should be presented in a meaningful way that facilitates the learners' understanding of the communicative purposes conveyed by the related synonyms.

The analysis of our study demonstrates that the corpus query procedure and technique used in section three is able to provide teachers a means to examine the typical collocates of synonyms, and help them become prepared for teaching difficult-to-distinguish lexical items. However, "identifying the collocation information that differentiates synonyms is a challenging and sometimes a labor intensive task" (Liu 2013: 32). The five constraints identified in our study provide a framework to analyze the collocation information obtained from corpora, and help teachers generate better pedagogical descriptions of synonyms to second/foreign language learners.

Specifically, in presenting new vocabulary words, teachers can first conduct the corpus query and identify the typical collocates of the target word, and then examine the collocational behaviors of the target word through the five constraints identified in our study. Such examination should scrutinize each of the five constraints individually as well as the possible interactions of the five constraints. For example, 调节心情 *tiáojié xīnqíng* and 调试心情 *tiáoshì xīnqíng* (adjust mood) both comply with the grammatical constraint. However, 调试心情 *tiáoshì xīnqíng* is not preferred due to its violation of the semantic constraint. Only the verb 调节

tiáojié denotes the modulation of emotions or feelings. The search of "调节 *tiáojié**心情 *xīnqíng*" generated 86 results in the Chinese Corpus created by Beijing Language and Culture University (http://bcc.blcu.edu.cn/), while "调试 *tiáoshì** 心情 *xīnqíng*" only had two results. Additionally, both 流 *liú* or 流下 *liúxià* (flow down) can co-occur with 泪 *lèi* (tears), based on the grammatical and semantic constraints. However, when 泪 *lèi* (tears) collates with 流下 *liúxià*, it must be followed by a complement to position it in a disyllabic template. The formation of 流 下泪来 *liúxià lèi lái* is the result of the interaction of grammatical, semantic and prosodic constraints.

If the learners' second/foreign language proficiency is at upper-intermediate/ advanced level, corpus query procedures and techniques can be taught to the learners to help them develop abilities to conduct corpus queries on their own. The five constrains discussed in this study can be employed as a framework to guide their analysis of the query results. Through such a heuristic approach, learners will begin to ask the kinds of questions that will lead them to gradually notice the constraints at work in choosing words for contexts, and eventually help them develop the ability to apply those constraints to their word choices. For students of other levels, teachers can present the analysis of the query to the students, and guide them to pay attention to the specific constraint(s) identified by the teacher. Following the examination of query results, examples can be extracted from the corpus data to develop teaching materials and additional learning exercises. For example, the following exercises can be designed for learners to distinguish the uses of 会 *huì* and 能 *néng*：

Sample exercise 1:   decide whether 会 *huì,* 能 *néng* or either of them can replace each other while keeping the original meaning and tone of the sentence.

1. 有些书架已经歪歪斜斜， 让人禁不住担心 (会；能; 会/能) 塌下来。
   Yǒuxiē diàodǐng de shūjià yǐjīng wāi wāixie xié, ràng rén jīnbuzhù dānxīn (huì; néng; huì/néng) tā xiàlái.
   'Some bookshelves have been slanted. People cannot help but worrying that the shelfs may collapse.'
2. 我成了麦霸，因为我觉得唱歌(会；能; 会/能)减肥。
   Wǒ chéngle mài bà, yīnwèi wǒ juédé chànggē (huì; néng; huì/néng) jiǎnféi.
   'I become a Microphone Hog, because I think singing can (help me) lose weight.'
3. 据说笑(会；能; 会/能)增长脸上的皱纹， 没关系, 那 不是 皱纹， 而 是美和哲理的痕迹。
   Jùshuō xiào (huì; néng; huì/néng) zēngzhǎng liǎnshàng de zhòuwén, méiguānxì, nà bùshì zhòuwén, érshì měi hé zhélǐ de hénji.

'It is said that laughter will grow wrinkles on the face. However, it does not matter. They are not wrinkles, but traces of beauty and philosophy.'

4. 尽量请假回来吧, 我请了很多同学, 到时候一定很热闹, 我好希望你 (会 ; 能 ; 会/能)回来 !

   Jǐnliàng qǐngjià huílái ba, wǒ qǐngle hěnduō tóngxué, dào shíhòu yīdìng hěn rènào, wǒ hǎo xīwàng nǐ (huì; néng; huì/néng) huílái!

   'Try your best to ask for a leave and come back. I invited many classmates. It must be very spirited. I do hope you can come back.'

5. 父母离异(会 ; 能 ; 会/能)影响孩子的心情 , 还(会 ; 能 ; 会/能) 影响性 格 。

   Fùmǔ líyì (huì; néng; huì/néng) yǐngxiǎng háizi de xīnqíng, hái (huì; néng; huì/néng) yǐngxiǎng xìnggé.

   'The divorce of parents may affect the child's mood, but also may affect (the child's) character.'

Sample exercise 2:   Judge whether the uses of 会 *huì* or 能 *néng* are appropriate in the following given situations.

1. An employee is expressing his/her aspiration to the superior.

   我希望能对自己的企业做出贡献 ,   同时也希望能因此而获得必要 的奖励。

   Wǒ xīwàng néng duì zìjǐ de qǐyè zuò chū gòngxiàn, tóngshí yě xīwàng néng yīncǐ ér huòdé bìyào de jiǎnglì.

   'I hope to contribute to my company, and I also hope to receive the necessary rewards because of my contribution.'

2. A college student is suggesting his/her roommate not stay up late.

   你早一点睡觉吧。经常熬夜的话会影响身体健康。

   Nǐ zǎo yīdiǎn shuìjiào ba. Jīngcháng áoyè dehuà huì yǐngxiǎng shēntǐ jiànkāng.

   'Please go to bed earlier. Staying up late often will affect your health.'

6.   Conclusion

Collocational information is a vital component in the learning of new vocabulary words in a second/foreign language. However, there are few suggestions in the learning Chinese as a second/foreign language literature on how to provide the collocational information to students, and impart a sensitivity to the distinctions between closely related lexical items. Based on a corpus-study of eight near-synonym pairs frequently misused by learners of Chinese, this study identified five constraints on the collocational behaviors of the selected near-synonym pairs. The

analysis indicates that the corpus query procedure and technique used in our study can be an effective method for teachers to examine the collocational behaviors of synonyms. The constraints identified in this study can be used as a framework for teachers or learners to analyze the collocation information extracted from corpora. The study also shows how corpus query data can be used to design exercises to expose learners to the synonyms and their collocations in different contexts, and then help learners truly understand the communicative purposes conveyed by the related synonyms.

References

Castello, Dominic. 2014. "A Corpus Study of *Strong* and *Powerful*." Retrieved December 14, 2017 from https://www.birmingham.ac.uk/Documents/college-artslaw/cels/essays/corpuslinguistics/A-Corpus-Study-of-Strong-and-Powerful.pdf

Feng, ShengLi. 2009. "On modern written Chinese." *Journal of Chinese Linguistics* 37(1): 145–162.

Firth, John Rupert. 1957. "Modes of meaning." In *Papers in Linguistics 1934–1951*, ed. By John Rupert Firth, 190–215. Oxford: Oxford U Press.

Firth, John Rupert. 1968. "A synopsis of linguistic theory 1930–1955." In *Selected Papers of J. R. Firth 1952–59*, ed. by Frank. R. Palmer, 1–32. Bloomington: Indiana University Press.

Fu, Yuxian and Xiaobin Zhou. 1990. 口语中的助动词 *Kouyu zhong de zhudongci* [Chinese modal auxiliary verbs in spoken language]. Guang Dong Education Press.

Gerstl, Peter. 1991. "A Model for the Interaction of Lexical and Non-Lexical Knowledge in the Determination of Word Meaning." In *Lexical Semantics and Knowledge Representation*, eds. by James Pustejovsky and Sabine Bergler, New York: Springer-Verlag.

Greenbaum, Sidney. 1974. "Some verb-intensifier collocations in American and British English." *American Speech* 49: 79–89.

Halliday, Michael. 1966. "Lexis as a linguistic level." In *Memory of J.R. Firth*, eds. by Charles Bazell et al., 148–162. London: Longman.

Hoey, Michael. 2000. "A world beyond collocation: New perspectives on vocabulary teaching." In *Teaching Collocations*, ed. by M. Lewis, 224–245. Hove: Language Teaching Publications.

Hoey, Michael. 1991. *Pattern of Lexis in Text*. Oxford: Oxford University Press.

Hunston, Susan. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Jones, Susan and John Sinclair. 1974. "English lexical collocations: a study in computational linguistics." *Cahiers de Lexicologie* 23(2): 15–61.

Lewis, Michael. 1997. *Implementing the lexical approach: putting theory into practice*. Hove, English: Language Teaching Publications.

Lai, Peng. 2006. "汉语能愿动词语际迁移偏误生成原因初探 hanyu nengyuan dongciyujie qianyi pianwu shengcheng yuanyin chutan [A Study on the Causes of Interlingual Transfer Errors in the Acquisition of Chinese Modal Auxiliaries]." 语言教学与研究[Yuyan jiaoxue yu yanjiu: Language Teaching and Linguistic Studies] 2006 (5): 67–74.

Li, Shouji. 2016. "美国留学生词语搭配的偏误分析及其对词语教学的启示 meiguo liuxuesheng ciyu dapei de pianwu fenxi jiqi dui ciyu jiaoxue de qishi [A Corpus-based Analysis of Collocational Errors by American Learners of Chinese and its Implication forthe Teaching of Vocabulary]." *Chinese as a Second Language ― Journal of the Chinese Language Teachers Association* 51(1): 51–69.

Li, Charles N. and Sandra A. Thompson. 1989. *Mandarin Chinese. A Functional Reference Grammar*. London and New York: University of California Press.

Liu, Dilin. 2013. "Using corpora to help teach difficult-to-distinguish English words." *English Teaching*, 68(3), 27–50.

Liu, Dilin and Shouman Zhong. 2014. "L2 vs. L1 use of synonymy: An empirical study of synonym use/acquisition." *Applied Linguistics* 37(2): 239–261.

Lü, Shuxiang, Wenlan Cai, Linding Li, Jian Liu, Qinghai Meng, Jiaxuan Shen, Jianci Wang, and Huaide Zhen. 1999. 现代汉语八百词 Xiandai Hanyu Babai Ci [Eight Hundred Words in Modern Chinese]. Beijing: The Commercial Press.

Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sharoff, Serge. 2006. Creating general-purpose corpora using automated search engine queries. In M. Baroni and S. Bernardini, editors, *WaCky! Working papers on the Web as Corpus*. Gedit, Bologna.

Walker, Crayton. 2008. "Factors which influence the collocation process." In *Cognitive Linguistic Approaches to Teaching Vocabulary and Phraseology*, eds. by F. Boers and S. Lindstromberg, 291–308. Berlin, New York: Mouton de Gruyter.

Woolard, George. 2000. "Collocation ― encouraging learner independence." In *Teaching Collocation*, ed. by Michael Lewis, 28–46. Hove: Language Teaching Publication.

Xiao, Richard and McEnery, Tony. 2006. "Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective." *Applied Linguistics* 27(1): 103–129.

Xiong, Wen. 2008. "The pragmatic functions of the Chinese Modal Auxiliaries: In reference to the Neng Verb Group." *Chinese Studies Review* 3: 30–48.

Yip, Po-Ching, and Don, Rimmington. 2004. *Chinese: a comprehensive grammar*. London: Routledge.

Yu, Ning. 2004. The eyes for sight and mind. *Journal of Pragmatics* 36: 663–686.

CHAPTER 9

Genericity and sentences with an AP state complement in Mandarin Chinese

Fei Ren
Georgetown University

This study investigates the semantic and pragmatic constraints on the generic/episodic interpretation of Chinese sentences containing a state complement (SC) realized by an adjectival phrase (AP). It argues that the generic interpretation of such sentences is a result of the interaction of the semantics of the verb or verb phrase before 得 *de* (V/VP-得), the AP complement after 得 *de*, and pragmatic knowledge. A sentence with an AP state complement will be interpreted as generic when the V/VP-得 in the sentence expresses repeatable or sum events, and when it is determined, given one's pragmatic knowledge and the semantics of the AP complement, that the property or state expressed by the AP can apply to a relevant event or individual in all events expressed by V/VP-得 in general.

Keywords: generic, episodic, state complement, 得 *de*, semantics, pragmatics

1. Introduction

A generic sentence expresses a regularity, which is a principled generalization over groups of individuals or particular episodes or facts (Krifka et al. 1995). It is non-deictic, and so its truth value cannot be determined solely with reference to a particular time (Carlson 1989). By contrast, an episodic sentence expresses a specific event, which holds at a particular time.

(1) a. John smokes a cigar after dinner.
     b. John smoked a cigar after dinner today.

(1a) expresses a regular event, which characterizes a property of the subject based on a set of particular episodes. However, (1b) describes a specific smoking event located in a past time interval. Generic sentences with an eventive predicate like (1a) are also called habitual sentences, expressing generalization over situations

specified by the eventive predicate rather than over individuals. The term "generic sentence" in this study refers to this type of sentences.

This paper explores the genericity of Chinese sentences with one type of state complement (SC) (状态补语/情态补语). SC is an important and highly grammaticalized category in Mandarin Chinese. Structurally, it refers to the sentence-final constituent XP introduced by the obligatory particle 得 *de* that immediately follows a predicate verb or adjective (V/A). SC constitutes part of the main predicate that takes the form of V/A-得-XP in an SC sentence, and can be realized by an AP, VP, CP, NP,[1] or a fixed expression, as illustrated in (2) and (3) below.

(2)   V-得 *de* + AP/VP/CP/NP
大家听到这个消息都跑得飞快 / 不知去向 /
Dajia   tingdao zhe ge xiaoxi dou pao -de feikuai /   bu zhi   quxiang/
People hear   this CL news   all   run -DE very fast / not know whereabout
一个都不剩 /一身汗。
yi   ge dou bu sheng / yi   shen han.
one CL even not left /   one CL   sweat
'When people heard this news, they all ran fast / ran away and disappeared without a trace / ran away and not even one person stayed / ran and became sweaty from running.'

(3)   A-得 *de* + CP/VP/VP/Fixed expression
53岁的藏胞尼玛次旺高兴得两眼发亮/上了天/
53 sui   de zangbao Nima Ciwang gaoxing -de liang yan faliang/shang -le
53 year DE Tibetan Nima Ciwan   happy   -DE two   eyes shine/ go up PERF
tian
heaven
像个孩子/不亦乐乎。
xiang ge haizi/bu yi le hu.
like   CL kid / extremely
'Nima Ciwan, the-53-year-old Tibetan's eyes are alight with joy / he is extremely happy / as happy as a kid / extremely happy.'

Semantically, SC functions to either describe or evaluate the situation[2] expressed by the verb/verb phrase or adjective before 得 *de* (V/VP/A-得), or to describe the result caused by the event or state denoted by V/VP/A-得. Assuming the verb or

---

1. As the distinction between noun phrases (NP) and determiner phrases (DP) is irrelevant to this study, I'll call both of them NPs.

2. The terms 'situation', 'event' and 'state' are used in the sense of Smith (1991). Events and states are two situation types. Events refer to dynamic situations, whereas states refer to static situations.

adjective immediately preceding 得 is gradable, SC can receive a degree reading, expressing to what extent the state denoted by V/VP/A applies to the subject. In (2), for instance, the AP *feikuai* 'very fast' describes the manner, or more specifically the speed, of the running event by the agent. In (3) with the gradable adjective *gaoxing* 'happy', the CP complement *liangyan faliang* 'eyes are alight' permits two readings: it may express either the resulted situation caused by the experiencer being happy, or the degree of happiness.

It has been noted by 刘 (2001), Lin (2003) and others that some sentences with SC (hereafter SC sentences) permit a generic reading besides an episodic one, even in the absence of contextual information or adverbials marking genericity, e.g., *zongshi* 'always'. At the same time, a good number of SC sentences receive an episodic interpretation with or without context.

(4) 他跑得很快。
    *Ta pao-**de** hen kuai.*
    He run-DE very fast
    a.   'He runs fast (*yesterday/*now).'
    b.   'He ran/is running fast (yesterday/now).'

(4) naturally obtains a generic reading that expresses a generalization over a set of running situations and an episodic reading that depicts a particular non-future instance of running. The sentence is compatible with time phrases like 'yesterday' or 'now' on the episodic reading but not on the generic reading. By contrast, SC sentences in (2) and (3) prefer an episodic interpretation.

The fact that an SC sentence, whether contextualized or isolated, may or may not have a generic reading suggests that its interpretation is under certain constraints besides the linguistic context. However, so far no studies have addressed this issue. To determine what the constraints are and how they come into play in deciding the interpretation of SC sentences is nonetheless essential. Thus this study intends to fill the gap.

Due to the space limit, this paper concentrates on SC sentences with eventive verbs and SC complements fulfilled by APs, as they are the type of SC sentences that favor generic reading the most. Furthermore, according to 吴 (2002), they are the most frequently used type of SC sentences by CFL students as well. Most Chinese examples in the paper are from BCC corpus (北京语言大学语料库), others are from CCL corpus (北京大学中国语言学研究中心语料库) and the internet with or without modification.[3] For the sake of simplicity, I focus on SC sentences with an individual subject in the analysis.

---

3. Chinese examples that do not cite a source are from the internet.

In this paper, I argue that an SC sentence can be interpreted as generic when two conditions are satisfied. First, the verb or verb phrase before 得 *de* (V/VP-得) expresses repeatable or sum events. Second, it is noted, given one's pragmatic knowledge and the semantics of the AP complement, that the property or state expressed by the AP can generally hold true of a relevant event or individual in all events expressed by V/VP-得.

The rest of the paper proceeds as follows. Section 2 discusses the semantics of SC realized by an AP. Section 3 analyzes generic SC sentences within the framework adopted in the study. Section 4 discusses the semantic and pragmatic factors involved in the interpretation of SC sentences with AP complements, and their application to other constructions in Mandarin Chinese. Section 5 briefly concludes the study.

2.   The semantics of AP complements in Chinese SC sentences

2.1   Event-oriented AP complements

Event-oriented SC is realized by an AP and has the semantic function of describing or evaluating an attribute of the event expressed by V/VP-得, such as its manner, time of occurrence, and frequency.

(5)   a.   她跳的虽然不是华尔兹，但是旋转得很优美。
                                                              (BCC-《苦儿流浪记》)
Ta tiao   de suiran   bu shi hua'erzi, danshi xuanzhuan -de hen
She dance DE although not be waltz,   but   turn   -DE very
youmei.
elegant
'Although she was not dancing waltz, she made turns elegantly.'
   b.   春天来得很早，
Chuntian lai   -de hen zao，
Spring   come -DE very early
伴随而来的是几场温暖的春雨。   (BCC- 《飘》)
bansui erlai de shi ji   chang wennuan de chun   yu.
come with DE be several CL   warm   DE spring rain
'Spring came early, and coming with it were a few warm spring showers.'
   c.   平反后他只身一人，虽然有三个儿子，
Pingfan   hou ta   zhishen yiren, suiran   you san   ge erzi,
Rehabilitated after he by himself alone, although have three CL son,
一个义女，但来往得很少。   (BCC-《文汇报》2001-4-2)

yi   ge yinü,   dan laiwang -**de** hen shao.
one CL adopted daughter, but contact -DE very rare
'He lives by himself after being rehabilitated. Although he has three sons
and an adopted daughter, he rarely contacts them.'

The AP complements in sentences in (5), although depicting different aspects of
the events expressed by V-得, are all event-oriented. In (5a), the AP complement
hen youmei 'very elegant' describes the manner of the action expressed by the verb
xuanzhuan 'turn'.[4] In (5b), the AP *hen zao* 'very early' concerns the time when
spring came, while in (5c) the AP *hen shao* 'very few' specifies the frequency of
contacting his children by the father.

Syntactically, the verb in V-得can be intransitive or transitive. While (5a)
above contains a one-place verb *xuanzhuan* 'turn', (6a) below contains a transitive
verb *shuo* 'speak' that takes an object.

(6) a. 他(说)中文说得很流利。
         Ta (shuo) zhongwen shuo -de hen liuli.
         He speak Chinese   speak DE very fluent
         'He speaks Chinese fluently.'
     b. 中文他说得很流利。
         Zhongwen ta shuo   -de hen liuli.
         Chinese   he speak DE very fluent
         'He speaks Chinese fluently.'

When the transitive verb 'speak' takes an object 'Chinese' in (6a), it is repeated
before 得 *de* to form a V-得-XP structure, and becomes optional before the object,
as indicated by the parentheses. The NP 'Chinese' can also precede the subject *ta*
'he', as (6b) shows. In both sentences, the AP *hen liuli* 'very fluent' expresses the
way the agent speaks Chinese.

As 刘 *Liu* (2001) observes, SC sentences with event-oriented AP complements
and repeated verbs like (6a) generally express habitual activities rather than partic-
ular instances. This holds true when the repeated verb is omitted. Yet, generic SC
sentences are not limited to those with (repeated) transitive verbs. An SC sentence
can be generic with an intransitive verb that takes an event-oriented AP comple-
ment too, as illustrated by *pao* 'run' in (4) and *laiwang* 'contact' in (5c).

---

4. The AP *hen youmei* 'very elegant' in (5a) is predicated of an event not a theme even though
the SC sentence can be interpreted as 'Her waltz is elegant'. This is because the NP 'her waltz' in
effect denotes a dancing event, i.e., the event of her dancing waltz.

2.2 Participant-oriented AP complements

SC realized by an AP is not necessarily event-oriented. It can predicate a participant of the event expressed by V/VP-得, including the agent, theme, or experiencer. AP complements of this type describe a state that is true of the event participant, co-existent with or resulted from the said event.

(7) 那人的字写得很漂亮，
Nei ren de zi xie -de hen piaoliang,
That person POSS character write -DE very beautiful,
作文写得出奇好，而且演讲能力也特别强。
(CCL-《从普通女孩到银行家》)
zuowen xie -de chuqi hao, erqie yanjiang nengli ye tebie
essay write -DE surprisingly good, and speech ability too particularly
qiang.
strong
'Characters written by that person are beautiful, and his essays are surprisingly well- written. His public speaking skills are also particularly strong.'

(7) contains two SC sentences. The APs *hen piaoliang* 'very beautiful' and *chuqi hao* 'surprisingly good' each modifies the respective subject NP that takes a theme role. (7) says, "characters written by the person are beautiful and essays written by him are surprisingly good". Both APs in the sentence are theme-oriented, describing a state brought about by the activity denoted by the verb *xie* 'write'.

AP complements in SC sentences may be linked with sentence subjects that are an agent or experiencer, and thus be agent-oriented or experiencer-oriented. Such APs tend to describe a property or mental/psychological state of the agent/ experiencer that either accompanies with or results from the event expressed by the predicate verb, as illustrated below:

(8) a. 上海的工作压力很大，
Shanghai de gongzuo yali hen da,
Shanghai POSS work stress very big,
雪儿工作得很不开心，很累。 (BCC-《都市快讯》 2003-1-8)
Xue'er gongzuo -de hen bu kaixin, hen lei.
Xue'er work -DE very not happy very tired
'It is very stressful to work in Shanghai. Xue'er works unhappily and is tired from work.'

b. [谭茵的话]林永寒听得很不舒服。 (BCC-《心心相拥》)
[Tan Yin de hua] Lin Yonghan ting -de hen bu shufu.
[Tan Yin POSS words] Lin Yonghan listen -DE very not comfortable
'Lin Yonghan became uncomfortable by/when listening to Tan Yin.'

In (8a), the AP *hen bu kaixin* 'very unhappy' talks about how the agent Xue'er feels at work, and *hen lei* 'very tired' describes the state she is in, as a result of working. In (8b), the AP *hen bu shufu* 'very uncomfortable' is predicated of the experiencer subject Lin Yonghan, describing his psychological state accompanying with or resulting from the activity of 'listening'.

刘 *Liu* (2001) points out that one difference between SC sentences with agent/theme-oriented complements and those with event-oriented ones is that the former expresses particular instances, regardless of whether or not they contain a repeated verb taking an object. However, it is not rare for SC sentences with a participant-oriented AP complement to be generic. (7) and (8a) are two examples.

In sum, event-oriented AP complements describe an attribute of the event E expressed by V/VP-得 in SC sentences. An agent/experiencer-oriented AP complement describes a state S_a true of the agent/experiencer in the sentence, which is co-existent with or resulted from E. A theme-oriented AP describes a resultative state S_t that holds true of the theme (i.e., the participant of E playing a theme role) and is subsequent to E. An SC sentence may receive a generic interpretation with all three types of AP complements.

3.   The generic domain and interpretation of Chinese SC sentences

3.1   Generic quantification over situations

The standard formal analysis of a generic sentence is to employ as its logic form (LF) the tripartite quantificational structure in (9), which is proposed by Heim (1982) to analyze adverbially quantified sentences, such as those marked by the adverb *often* or *always*.

(9)   Operator (Restrictor, Nuclear Scope)

The operator, which is an adverbial quantifier (AdvQ), takes a wide scope over a sentence, and the rest of the sentence is partitioned into a restrictor and a nuclear scope. The restrictor specifies a set of situations that restrict the quantification domain of the operator, and can be explicitly expressed by a restrictive clause, such as an *if-clause* or *when-clause*. The nuclear scope contains the main clause without the operator. For example,

(10)   a.   John often/always smokes a cigar after dinner.
       b.   OFTEN/ALWAYS *s* (*s* is a situation after John's dinner) (John smokes a cigar in *s*)

In (10a), the overt AdvQ *often/always* serves as an operator, and the adverbial clause "after dinner" serves as the restrictor, providing a set of after-dinner situations s for the operator to quantify over. The matrix clause "John smokes a cigar" is the nucleus of the sentence. (10b) is the LF representation of the sentence, where *often/always* bounds the situation variable s in the restrictor and the nuclear scope. Under this analysis, (10) means for many/all situations s such that s is after John's dinner, John smokes a cigar in s.

Generic sentences without an AdvQ are generally assumed to contain a null GEN operator with similar function as an AdvQ, like *generally* or *normally*. The GEN operator quantifies over a group of situations in the restrictor, giving rise to the generic interpretation.

The restrictor of a generic sentence is not always specified overtly. When this happens, the restrictor or the quantifying domain of the AdvQ or GEN operator can be determined pragmatically (Schubert and Pelletier 1989; Krifaka et al. 1995; a.o.). von Fintel (2004) argues that a contextually determined discourse topic (i.e., what a discourse is about) can provide the quantifying domain of an AdvQ with an appropriate set of situations. And one can rely on the focus structure of the generic sentence to reconstruct the discourse topic, thereby identifying the quantifying domain. For instance,

(11)   a.   Tai eats with [**chopsticks**]$_F$.
        b.   GEN s (*Tai eats in s*) (Tai eats with chopsticks in s)
        c.   Normally, *when Tai eats*, he eats with chopsticks.

The restrictive domain of the GEN operator in (11a) is understood to contain a set of situations where Tai eats, as shown in the semantic representation in (11b). The restrictor can also be spelled out as a restrictive *when-clause*, as in (11c). According to von Fintel (2004), the covert restrictor is identified pragmatically: in normal contexts, the information focus falls most naturally on *chopsticks*, and this focus structure triggers an inexplicit question *how Tai eats*, which is the discourse topic. It is this discourse topic that introduces a set of eating situations by Tai to the restrictor.[5]

It's important to note that in order for a sentence to have a felicitous generic reading, the situation variable s at its LF cannot be tied to a particular event. In other words, the restrictive domain of the GEN operator must contain multiple episodes to license the GEN operator, and hence a generic reading. Otherwise, the sentence only expresses a single event when a generic reading is not mandatory, or is infelicitous on the intended generic interpretation (Krifka et al.1995). For example,

---

5. This is an oversimplified introduction of von Fintel's analysis, omitting the complex formal mechanisms leading to the conclusion. For more about the analysis, please see von Fintel (2004).

(12) a. John smoked a cigar after dinner today.
    b. [#]When Minette dies, Mary is unhappy.[6]

(12a) reads episodic only, in spite that English past tense is compatible with both episodic reading and generic reading. This is because the *after-clause* with the time adverbial "today" is understood to specify a particular "after dinner" situation to which the variable *s* is tied, given the world knowledge that one normally has one dinner a day. A generic reading is thus not available. (12b) is infelicitous on the intended generic reading (indicated by the symbol #), which reads, "in general, Mary is unhappy each time when Minette dies". However, the restrictive domain introduced by the *when-clause* contains a dying event that can only occur once in normal contexts.

In this paper, I hold that the logic form of a generic sentence without an explicit restrictive clause is a tripartite structure with a covert generic operator quantifying over a set of contextually determined situations. In the absence of further contextual information, the focus structure of a sentence can help reconstruct the discourse topic and thereby the restrictor that supplies appropriate situations to the generic domain. The generic domain must contain multiple situations to license a GEN operator and have a felicitous generic reading.

3.2 The generic reading of Chinese SC sentences

The SC sentence in (13) allows a generic reading when no explicit generic marker or restrictor is present, so we expect the LF representation of the sentence to contain a GEN operator and a contextually determined restrictor in the form of (13b).

(13) a. 这确实是一头出色的骆驼，肯定跑得[**很快**]$_F$。
                                            (BCC-《恐怖的大漠》)
     *Zhe queshi shi yi tou chusede luotuo, kending pao-**de** [**hen kuai**]$_F$*
     This indeed be a CL excellent camel, must run-DE very fast
     'This is indeed an excellent camel. It must run very fast.'
   b. GEN (Restrictor) (Nuclear Scope)

The question is how to identify the restrictor in (13b). According to von Fintel (2004), this can be done by reconstructing the relevant discourse topic, often in the form of a question, with the assistance of the focus structure of the sentence. In normal contexts, the information focus of (13a) falls most naturally on the event-oriented AP *hen kuai* 'very fast' and so triggers an inexplicit question *how (fast) does the camel run*. The discourse topic thus identified concerns the speed of the camel running, which introduces a set of running events by the camel to

---

6. The sentence is adapted from de Swart (1991) cited in Krifka et al. (1995).

the restrictor of the GEN operator. So (13a) is interpreted against a set of situations where the relevant running activity by the camel occurs. (14a) gives its LF representation, and (14b) spells out the restrictor of the SC sentence (ignoring the modal *kending* 'must').[7]

(14)   a.   GEN*s* (the camel runs in *s*) (it runs very fast in *s*)
         b.   When the camel runs, it runs fast.

In the same vein, we can identify the restrictive domain of generic SC sentences with a transitive verb. Take (6a) *Ta shuo zhongwen shuo-de hen liuli* 'He speaks Chinese fluently' for instance. The information focus in (6a) normally falls on the sentence-final AP *hen liuli* 'very fluent', which triggers the implicit question *how well does he speak Chinese* that denotes the discourse topic. Therefore, the set of situations introduced by the discourse topic to the domain of the GEN operator are Chinese-speaking events by the agent. See (15a) for its LF representation, and (15b) for the restrictor identified pragmatically.

(15)   a.   GEN*s* (he speaks Chinese in *s*) (he speaks Chinese fluently in *s*)
         b.   When he speaks Chinese, he speaks it fluently.

Similar analysis goes for SC sentences with participant-oriented complements that permit generic reading. For instance, in the SC sentence in (8a) *Xue'er gongzuo-de hen bu kaixin* 'Xue'er works unhappily', the agent-oriented AP *hen bu kaixin* 'very unhappy' bears the information focus just like an event-oriented AP does. The discourse topic reconstructed through the inexplicit question *how does Xue'er work* denotes multiple working events by the agent, with regard to which, the SC sentence is interpreted. (16) is the LF of the sentence.

(16)   GEN*s* (Xue'er works in *s*) (she is unhappy in *s*)

(16) says, "generally, Xue'er is unhappy when she works". Taking the same steps, we can determine the generic domain of SC sentences with theme-oriented AP complements. I'll not repeat the analysis here.

   Note that in each generic SC sentence discussed above, the generic domain of the GEN operator contains a set of events expressed by the materials occurring before 得 *de*. This is not accidental, as what precedes 得 *de* in an SC sentence is considered conveying known information and so not focused, and as Li and Thompson (1981: 625) observe, the non-focused part must express something "that has already come up in the discussion or one that is 'in the air'". In other

---

7. In this particular context, the modal *kending* 'must' contributes to the generic interpretation as well. Yet the modal is not required for the sentence to read generic.

words, the non-focused materials before 得 *de* provide information about the discourse topic used to reconstruct the situations in the restrictor.[8]

4. The generic/episodic interpretation of Chinese SC sentences with AP complements

4.1 Repeatable events and sum events

We see that in order for a Chinese SC sentence to have a felicitous generic reading, the generic domain at its LF must contain a group of situations, with respect to which the sentence is interpreted. I call it "multi-situation condition". To satisfy this condition, V/VP-得 in a given SC sentence must be able to express either repeatable events or sum events, given that the restrictor of the sentence contains situations expressed by the relevant V/VP. Let's look at verbs expressing repeatable events first. Compare (17) and (18).

(17) a. 祖父起得很早，也睡得很早。   (BCC-《冰心文集》)
      *Zufu   qi -**de** hen zao, ye shui -**de** hen zao.*
      Grandpa get up -DE very early, also sleep -DE very early
      'Grandpa gets up early, and sleeps early too.'
   b. GEN*s* (Grandpa gets up/sleeps in *s*) (Grandpa gets up/sleeps early *s*)

(18) a. 她只知道她的祖父(#经常)死得很早。
                                (Adapted from BCC-《现代聊斋》)
      *Ta zhi zhidao ta de   zufu   (#jingchang) si -**de** hen zao.*
      She only know she POSS Grandpa (often)   die -DE very early
      'All she knows is her Grandpa (#often) died at an early age.'
   b. # GEN*s* (Grandpa dies in *s*) (Grandpa dies at an early age in *s*)

(17) reads generic, whereas (18) reads episodic. As the predicates in the two sentences differ only in V-得, the presence or lack of the generic reading should be ascribed to the verbs. In (17a), the verbs *qi* 'get up' and *shui* 'sleep' denote events that can be performed by the agent more than once, whereas the event denoted by *si* 'die' in (18a) is once-only, which cannot happen repeatedly to the same person in normal contexts. Consequently, in contrast with (17b) whose restrictor consists of multiple situations where Grandpa gets up, the restrictor in (18b) contains a single episode, ruling out the generic reading. On the intended generic reading,

---

8. In addition to the information structure of generic SC sentences, the meaning of AP complements and pragmatic knowledge are also essential in the restrictor identification process. As they are not central to the argument in this paper, I'll not go into details here.

(18b) reads "normally, when Grandpa dies, he dies at an early age". This reading is obviously infelicitous. (18b) also shows that an overt adverbial quantifier *jingchang* 'often' cannot coerce an SC sentence generic if V-得 denotes an event that cannot recur.

Consumption and creation verbs do not denote repeatable events either when they take an object with specific reference (Smith 1991; Rimell 2004). Such an object can be marked by determiner *zhe* 'this' or *na* 'that' in Chinese, or is associated with some entities introduced in the preceding context. An SC sentence expresses an instance with this type of VPs.

(19)   a.   这次任务(#总是)完成得不错，上面很满意。

<div align="right">(BCC-《福建日报》1980-12-10)</div>

Zhe ci   renwu (#zongshi) wancheng -de bu cuo, shangmian hen
This time task   (always)   complete -DE not bad, above   very
manyi.
satisfied
'You (#always) completed this task well. The authorities above us were very satisfied.'

b.   他(写)这篇文章，(#总是)写得很好，

Ta (xie)   zhe pian wenzhang, (#zongshi) xie   -de hen hao,
He (write) this CL   article,   (always)   write -DE very good

值得仔细阅读研究。   (Adapted from BCC-微博)

Zhide zixi   yuedu yanjiu.
worth careful read   study
'He (#always) wrote this article very well. It is worth careful reading and study.'

(19a) has a consumption verb *wancheng* 'complete', and (19b) has a creation verb *xie* 'write'. The object NP of both verbs is marked by a determiner, referring to a specific entity. Both sentences read episodic only, because neither event denoted by the VP-得is repeatable. In (19a), the VP 'complete this task' denotes an event that cannot be accomplished multiple times once it is finished. In (19b), the event 'write this article' is once-only in the sense that the same article cannot be created again after it is done. Even though the agent can write about the same topic more than once, the subsequent articles written cannot be the original article completed. Therefore, the respective restrictor of the two sentences is interpreted as containing a sole episode. Like (18) with the verb *si* 'die', adding an overt adverbial quantifier, e.g., *zongshi* 'always', to either sentence cannot make them generic.

However, this does not mean that verbs taking an object with specific reference in SC sentences cannot express repeatable events at all. In (20a), for instance, the object of the verb before 得 *de* is a definite noun, and in (20b), the object is a

proper noun, referring to the name of a song. VPs in both sentences express events that can recur.

(20) a. 他把那个不好笑的故事讲得很有趣。
          Ta ba na ge bu haoxiaode gushi jiang -de hen youqu.
          He BA that CL not funny   story tell   -DE very interesting
          'He tells that dull story in an interesting way.'
     b. 小城故事邓丽君唱得好听极了。
          'Xiaocheng gushi' Deng Lijun chang -de haoting   jile.
          Small town story Deng Lijun sing   -DE beautiful extremely
          'Deng Lijun sang *the story of a small town* extremely beautifully.'

The AP complements in (20) can be predicated of an event or an individual. On the event-oriented reading, (20a) says "normally, when he tells that dull story, he does so in an interesting way". (20b) says "normally, when the deceased singer Deng Lijun sang *the story of a small town*, she did so in an extremely beautiful way". The generic readings are felicitous because the events expressed by VP-得 in both sentences can be reiterated. Just as the same story can be told multiple times by the same agent, the same song can be sung repeatedly by the same singer as well.

It is worth noting that whether the event expressed by V/VP-得 is repeatable is, more often than not, a pragmatic issue. Take *si* 'die' for instance. It is not difficult to imagine a scenario where one can die multiple times. When playing video games, for example, a player can legitimately say "I died again". It is just that when interpreting an SC sentence in isolation, one tends to evaluate the sentence in normal contexts.

Another way to satisfy the multi-situation requirement is to introduce to the restrictor sum situations, a term borrowed from Krifaka et al. (1995), which consist of similar but not identical individual situations. This can be done when the predicate verb in an SC sentence takes a noun phrase with generic reference, such as a bare noun. For instance,

(21) [他]人有点其貌不扬，但(写)文章写得不错。   (BCC-《抽屉深处》)
      [Ta] ren   youdian qi mao bu yang,   dan (xie) wenzhang xie
       He person a bit   his appear not handsome, but (write) article   write
      -de bu cuo.
      -DE not bad
      'He looks a bit homely, but he writes articles well.'

Differing from (19b) that reads episodic with the NP *zhe pian wenzhang* 'this article', (21) with the bare noun *wenzhang* 'article' is generic, meaning "in general, when he writes an article, he does it well". The reason is that a Chinese bare noun can be generic, representing a set of different entities. This is especially true when

the noun occurs as the object of the main verb in SC sentences. The bare noun 'article' in (21) denotes a group of different articles the agent writes on different occasions. As there are multiple article-writing events by the agent in the generic domain of the sentence, the generic reading is licensed.

Other types of NPs may have generic reference too, and so are associated with generic SC sentences as well.

(22) a. 新版《十万个为什么》虽然定价不菲，
          Xin ban   'Shiwan ge weishenme' suiran   dingjia bu fei,
          New edition '100,000 CL why'   although price   not low,
          仍然卖得很快。   (BCC-《人民日报》2000)
          *rengran mai -**de** hen kuai.*
          still   sell -DE very fast
          'Although the price of the new edition of *100,000 Why* is high, the book still sells fast.'
      b. 人这一辈子没法做太多的事情，
          Ren   zhe yibeizi meifa zuo tai duo   de shiqing,
          Person this life   cannot do too many DE thing
          所以(做)每一件都要做得精彩绝伦。   (BCC-微博)
          suoyi (zuo) mei   yi   jian dou yao   zuo-de jingcai juelun.
          so   (do) every one CL all   must do-DE wonderful
          'One cannot do too many things in one's whole life, so one must do every single thing wonderfully.'

'100,000 Why' in (22a) is a book title. The subject NP 'the new edition of *100,000 Why*' is kind-referring, denoting a group of different copies of the same book to be sold. Thus, the relevant generic domain of the SC sentence consists of a set of book-selling events, with each event containing a different copy of the book. In (22b), the verb *zuo* 'do' in the SC sentence takes a quantifier phrase *mei yi jian* 'every matter' that marks genericity, and so the VP 'do everything' expresses a set of events done by the agent in distinct situations. Thus, both sentences in (22) are interpreted as generic.

To summarize, for a given SC sentence, the multiple situations in the restrictor at its LF can be provided in two ways. One is when V/VP-得 denotes repeatable events, and the other is when the verb takes an object with generic reference and so expresses sum events. However, having a multi-situation-denoting V/VP does not guarantee a generic interpretation of SC sentences, because other factors play a part in this regard as well.

4.2 The role of AP complements and pragmatic knowledge

If the presence of generic reading of an SC sentence is solely determined by whether the "multi-situation condition" is met, then we expect (23) to be generic with all the AP complements, given that the main verb *zou* 'leave' can express repeatable events. However, this is not the case.

(23) 她走得很早/#匆匆忙忙/#很不情愿。
   Ta zou -de hen zao   /# congcong mangmang /#hen bu qingyuan.
   She leave -DE very early/ #hasty   /#very not willing
   'She leaves early/#in haste/#unwillingly.'

(23) permits a generic reading with the event-oriented AP 'very early', but not with the manner AP 'hasty' or the agent-oriented AP 'very unwilling'. In fact, on the intended generic reading, the sentence is semantically well-formed with each of the AP complements. There is nothing peculiar, for example, to say a person normally leaves in a hurry or unwillingly. But such an interpretation is unavailable in the absence of contextual information to the contrary.

Generally speaking, an SC sentence reads episodic when it has an agent/experiencer-oriented AP denoting a mental/emotional state of the agent/experiencer, or denoting a state resulting from the event expressed by V/VP-得.

(24) a. 她还笑得很好看/#很开心/#很凄凉。   (BCC-《微神》)
      Ta hai xiao -de hen haokan /#hen kaixin /#hen qiliang.
      She still smile -DE very beautiful /#very happy /#very desolate
      'She still looks beautiful/#happy/#desolate when she smiles.'
   b. 她洗衣服洗得很勤/#很累/#很烦躁。
      Ta xi   yifu   xi -de hen qin   /#hen lei   /#hen fanzao.
      She wash clothes wash -DE very frequent /#very tired /#very impatient
      'She washes clothes frequently.'
      '#She becomes very tired/impatient from washing clothes.'

All AP complements in (24a) are agent-oriented. The AP *hen haokan* 'very beautiful' describes how the agent looks while smiling, and the APs *hen kaixin* 'very happy' and *hen qiliang* 'very desolate' express the emotional and psychological state of the agent at the time of smiling. The sentence permits a generic reading only with the first AP. In (24b), the AP *hen qin* 'very frequent' is predicated of the event of washing clothes, specifying the frequency of the washing activity. The other two APs *hen lei* 'very tired' and *hen fanzao* 'very impatient' are agent-oriented, which concerns the physical and mental state of the agent resulted from the washing event. The sentence reads generic only with the event-oriented AP, although the

VP-得 is capable of expressing sum events. In contrast with agent-oriented APs with a result reading in (24b), theme-oriented APs may be associated with genericity.

(25) 她洗衣服洗得很干净/#很旧了。
      *Ta xi   yifu   xi   -**de** hen ganjing /#hen jiu le.*
      She wash clothes wash -DE very clean   /#very old SFP
      'The clothes are washed clean/#are worn out from washing.'

In (25), both *hen ganjing* 'very clean' and *hen jiu le* 'very worn out' are theme-oriented, expressing the resulted state of the clothes after being washed. While the sentence reads generic with the first AP, it is episodic with the second.

Sentences in (23)–(25) show that having a multi-situation-denoting V/VP preceding 得 is a necessary but not sufficient condition for SC sentences to read generic, and that the generic/episodic interpretation of SC sentences varies with the AP complements in the absence of further contextual information. An immediate question is how AP complements affect the interpretation of SC sentences. I propose that they do so through what I call "possibility condition" stated in (26). An SC sentence does not read generic if this condition is not fulfilled.

(26) It seems possible in normal contexts, given the semantics of an AP complement and one's pragmatic knowledge, that the property or state expressed by the AP in the nuclear scope applies to the relevant event or event participants in *all* events expressed by V/VP-得 in the generic domain.

Take (23) for instance. When reading generically, (23) with the AP *hen zao* 'very early' means "normally she leaves at an early time in all relevant situations where she leaves". That is, all the leaving events in the restrictor possess the property of being early, as the LF in (27) shows.

(27) GENs (she leaves in s) (she leaves early /#in haste in s)

On the other hand, with the manner AP 'hasty', (27) says "generally whenever the agent leaves, she leaves in a hasty manner". While it is plausible, given one's world knowledge, that the property of being early can be true of all the relevant leaving events the agent is engaged in, such as leaving for work or school, it seems less likely that generally, each time the agent leaves somewhere, she does so in a hasty manner in normal contexts. Accordingly, it is decided that the "possibility condition" is satisfied with the AP 'very early' but not with 'hasty' in (23), hence the different interpretations of the sentence with different APs.

The "possibility condition" is pragmatic in nature, having to do with one's judgment about the level of regularity of the property or state expressed by the AP complement, based on one's world knowledge and the semantics of the AP. Therefore, both the world knowledge and the AP in terms of its semantic

orientation and relationship with the event expressed by V/VP-得 play a crucial part in deciding the generic/episodic interpretation of an SC sentence. Take (24a) and (24b) with agent-oriented APs for example. Their LF representations are in (28a) and (28b) respectively.

(28) a. GEN*s* (she smiles in *s*) (she looks pretty/#happy/#desolate in *s*)
      b. #GEN*s* (she washes clothes in *s*) (she is tired/impatient in *s*)

As (28a) shows, the APs meaning 'pretty', 'happy' and 'desolate' each expresses a state of the agent, which is co-existent with the smiling event expressed by V-得 in (24a). The sentence is not considered generic with the APs 'happy' and 'desolate' because the APs describe the agent's emotional and mental states, which are transient and accidental. Although it is not absolutely impossible that each time when one smiles, one looks happy or desolate, this is not the preferred reading in the lack of contextual information or expressions that force a generic reading, such as *zongshi* 'always'. In (28b), the APs 'tired' and 'impatient' describe the agent's physical and mental states resulting from the clothes-washing event expressed by VP-得 in (24b). When out of context, the resultative state of the agent being tired or impatient is viewed as located in the present time *now*, and so there is no way for the state to be true in all the relevant washing activities in the restrictor in (28b) to fulfill the "possibility condition".

On the other hand, SC sentences with a theme-oriented AP complement can read generic, even though the AP also expresses a resultative state, as (25) illustrates. Its LF representation is in (29).

(29) GEN*s* (she washes clothes in *s*) (the clothes washed are clean/#worn out in *s*)

(29) has a felicitous generic reading with the AP *hen ganjing* 'very clean' expressing the final state of the clothes washed. The reason is that when evaluating the truthfulness of the sentence, we consider the cleanness of all the clothes being washed in one or more than one washing episode. The sentence will receive a generic reading if, given the world knowledge, it is plausible for all the clothes washed by the agent in a particular instance to be clean, or if the clothes being washed become clean each time after the agent performs the washing activity. As either case seems plausible, a generic reading is expected. (29) reads episodic with the AP *hen jiu* 'very worn out', as it is unlikely that all the clothes become worn out after a single wash, or they get worn out each time being washed.

It needs to be pointed out that the semantics of AP complements becomes irrelevant with regard to the interpretation of SC sentences when contextual information or some explicit expression makes it clear that the sentences express a generic situation.

(30) a. 王小丽也不和我们说话，…
Wang Xiaoli ye bu he women shuohua,…
Wang Xiaoli also not with we   talk
上课下课仍然走得匆匆忙忙。   (BCC-《寻找春天的百合花》)
shangke xiake *rengran zou* -***de*** *congcong mangmang.*
go to class get off class still   leave -DE hasty
'Wang Xiaoli doesn't talk to us either… she still leaves in haste when class is over.'

    b. 在别人面前常常笑得很开心，
Zai bieren mianqian ***changchang*** *xiao* -***de*** *hen kaixin,*
At other front   often   smile -DE very happy
一个人时却很落寞。   (BCC-微博)
yi ge ren shi que hen luomo.
one CL person when but very lonely
'When in front of other people, he often smiles happily. But when he is alone, he feels empty and lost.'

In (30a), the SC structure with the manner AP 'hasty' is interpreted as generic with the presence of an overt restrictor *shangke xiake* 'come to class and get out of class', which indicates that the action of leaving (the classroom) in a hasty manner is not accidental. In (30b), the adverb *changchang* 'often' overtly marks the V-*de*-AP 'smile-*de* happy' as generic, in spite of the fact that the AP *kaixin* 'happy' normally expresses a transient mental state.[9]

    In addition, when the main verb in an SC sentence denotes an activity that is understood to happen regularly, such as *guo* (*rizi*) 'live', *shenghuo* 'live', *gongzuo* 'work', the sentence favors a generic interpretation regardless of the semantics of the AP complements. Similar verbs include *xue(xi)* 'study', *xiangchu* 'get along', *fendou* 'fight for an aim', and so on.

(31) a. 日子过得简单，却也充实。   (BCC-《面的司机》)
Rizi guo -***de*** *jiandan, que ye chongshi.*
Days live -DE simple, but also fulfilling
'He lives a simple but fulfilling life.'

    b. 你生活、工作得快不快乐、累不累，
Ni ***shenghuo, gongzuo*** -***de*** *kuai bu kuaile, lei bu lei,*
you live   work   -DE happy not happy, tired not tired

---

9. Examples in (30) do not exhaust the types of contextual information contributing to the generic/episodic interpretation of SC sentences. Due to the limited space, I will leave the issue as it is and will elaborate it on another occasion.

只有你自己知道。   (Adapted from BCC-微博)
zhiyou ni   ziji zhidao.
only   you self know
'Only you know whether you live a happy life or are exhausted by life,
and whether you are happy at work or feel exhausted after work.'

Due to the habitual nature encoded in the verbs *guo* 'live', *shenghuo* 'live' and *gong-zuo* 'work', the attribute or state expressed by the AP complements is interpreted against a set of living or working situations separated by an interval of a day. In other words, the restrictors of these SC sentences are seen as having multiple situations expressed by these verbs, no matter what the features of the AP complements are.

Before leaving the discussion, I'd like to point out that the AP complements discussed in the paper do not contain degree adverbs or temporal adverbs with deictic information, such as *zhen* 'really', *zheme* 'like this', and *zheng* 'in the process of'. AP complements with these adverbs, e.g., *zheme kuai* 'this fast; so fast', *zheng gaoxing* 'being happy', describe a property or state that is often tied to an ongoing or completed episode. Accordingly, SC sentences with such an AP complement are interpreted as episodic.

4.3   The generic interpretation of other constructions in Chinese

SC sentences with state complements realized by non-APs may permit a generic reading as well (see 32), so may sentences without SC complements (see 33).

(32)   a.   她笑得像个孩子/#喘不过气来/#肚子疼。
            *Ta xiao -**de** xiang ge haizi /#chuan bu guo qi lai /#duzi   teng.*
            She laugh -DE like   CL kid   /# out   of breath   /#stomach hurt
            'She laughs like a kid/#was breathless with laughter/#laughed so hard
            that her stomach hurt.'

        b.   #这一顿饭吃得宾主尽欢。
            *Zhe yi   dun fan   chi -**de** bin   zhu jin huan.*
            This one CL   meal eat -DE guest host all happy
            #'Both the guests and the hosts thoroughly enjoyed the meal.'

(33)   他走路上班/#送朋友回家。
       Ta zoulu shangban / # song pengyou hui jia.
       He walk go to work/# take friend   go home
       'He walks to work/# walked a friend home.'

The availability of a felicitous generic reading in sentences like (32) and (33) is also subject to the "multi-situation condition" and the "possibility condition" slightly modified as follows:

(34) a. *Multi-situation Condition*: the generic domain at LF of a sentence must contain a group of situations, with respect to which the matrix clause in the nuclear scope is interpreted.

b. *Possibility Condition*: it seems possible in normal contexts, given the semantics of the matrix clause and one's pragmatic knowledge, that the situation expressed by the matrix clause is true in *all* situations in the relevant generic domain.

Recall that in the absence of further contextual information, the non-focused materials of a sentence can help reconstruct the situations in the generic domain, while the elements bearing the information focus constitute a major part of the matrix clause in the nuclear scope. In the case of SC sentences where the information focus invariably falls on the XP complements by default Xu (2004), the XP complements appear in the nuclear scope regardless of their semantic orientation, whereas the non-focused V/VP-得 introduces situations to the generic domain. As for non-SC sentences like (33), their generic domain and nuclear scope at LF are established in a similar fashion.

Take (32) for instance. The generic domain of sentences in (32) contains events expressed by V/VP-得, and the nuclear scope has the VP/CP complements in the matrix clause. In (32a), while the "multi-situation condition" is satisfied by the repeatable event expressed by the verb *xiao* 'laugh' before 得, the generic interpretation is ruled out by the "possibility condition" when the complement is the VP 'be out of breath' and the CP 'stomach hurt', as it seems unlikely, given our pragmatic knowledge, that generally one is breathless or has a stomach ache whenever one laughs. (32b) doesn't permit a generic reading because the VP-得 'eat the meal' denotes a particular eating event, violating the "multi-situation condition".

In contrast with (32), the focus in (33) naturally falls on the first VP *zoulu* 'walk', which describes the means of transportation and enters the nuclear scope. The sentence reads generic with the non-focused VP 'go to work' denoting repeatable events, but reads episodic with the VP 'take a friend home' capable of denoting sum events and fulfilling the "multi-situation condition". The generic/episodic interpretation of (33) with the two VPs is ascribed to our judgment of the likelihood of the intended habitual situations: while it is not rare for one to commute to work by walking regularly, it is a lot less likely that in all situations where one needs to take a friend home, one does so by walking.

5.   Conclusion

This paper examines the generic/episodic interpretation of Chinese sentences with a state complement that is realized by an adjectival phrase. It shows that the presence of a generic reading in such sentences is subject to two conditions: the "multi-situation condition" and the "possibility condition", which involve both the semantic properties of relevant expressions in the sentences and pragmatic knowledge. The two conditions are applicable to the generic/episodic interpretation of many other constructions in Chinese as well. The study suggests that in Mandarin Chinese, in which genericity is not marked by any overt generic markers or tense-aspect markers, both semantic and pragmatic factors contribute to the generic interpretation of sentences.

References

Carlson, Greg. 1989. "On the Semantic Composition of English Generic Sentences". In *Properties, Types and Meaning. Volume II: Semantic Issues*, ed. by Chierchia, G., B. H. Partee, and R. Turner, 167–192. Dordrecht, Boston, London: Kluwer Academic Publishers.

Heim, Irene. 1982. The semantics of definite and indefinite noun phrases. PhD dissertation. University of Massachusetts, Amherst.

Krifka, Manfred, et al. 1995. "Genericity: An Introduction". In *The Generic Book*, ed. by G. N. Carlson and F. J. Pelletier, 1–124. Chicago and London: University of Chicago Press.

Li, Charles. N., and Sandra. A. Thompson. 1981. *Mandarin Chinese: A Functional Grammar*. Berkeley: University of California Press.

Lin, Jo-Wang. 2003. "Aspectual selection and negation in Mandarin Chinese". *Linguistics* 41 (3): 425–459.

Rimell, Laura. 2004. "Habitual Sentences and Generic Quantification". In *WCCFL23 Proceedings*, ed. by G. Garding and M. Tsujimura. Somerville, 663–676. MA: Cascadilla Press.

Schubert, L. and F. Pelletier. 1989. "Generically Speaking". In, *Properties, Types and meaning*, vol.2, ed. by G. Chierchia, B. Partee and R. Turner, 193–268. Dordrecht: Kluwer.

Smith, Carlota. 1991. *The Parameter of Aspect*. Dordrecht, Boston, London: Kluwer Academic Publisher.

von Fintel, Kai. 2004. "A Minimal Theory of Adverbial Quantification". *Context Dependence in the Analysis of Linguistic Meaning: Current Research in the Semantics/Pragmatics Interface*. Vol 11, ed. by Barbra Patee, Hans Kamps, 137–175. Elsevier.

Xu, Liejiong. 2004. "Manifestation of informational focus". *Lingua* 114: 277–299.

胡裕树. 2011.《现代汉语》. 上海: 上海教育出版社.

刘月华等. 2001.《实用现代汉语语法》. 北京: 商务印书馆.

吕叔湘. 2009.《现代汉语八百词》. 北京: 商务印书馆.

陆丙甫等. 2015. "状态补语是汉语的显赫语法成分".《中国语文》第三期.

吴颖. 2002. " '动词+得+补语'的分类和语义特征分析".《苏州大学学报》第二期.

朱德熙. 1982.《语法讲义》. 北京: 商务印书馆.

CHAPTER 10

Kinship metaphors in the Chinese construction A *shi* B *zhi fu/mu*

Biology and culture as conceptual basis

Lin Zhu

University of Oregon

This corpus-driven study focuses on two metaphorically used kinship terms in Modern Chinese, 父 *fu* 'father' and 母 *mu* 'mother'. Under investigation are two constructions [A *shi* B *zhi fu*] 'A is the father of B' and [A *shi* B *zhi mu*] 'A is the mother of B'. It is found that the figurative meanings expressed by *mu* (mother) are more conventionalized than those expressed by *fu*. The study shows that *mu* has higher metaphoricity, and I argue that the degree of metaphoricity of the two kinship terms in Chinese is a function both of the experiential basis of cognition in terms of universal biological phenomenon and of cultural constraints, especially Confucian thoughts, on conceptualization.

Keywords: kinship metaphor, the Chinese language, corpus linguistics, culture, semantics, cognitive linguistics, experiential basis of cognition

1. Introduction

The study of semantics has demonstrated that a word can have more than one meaning (Cowie 2012: 25; Gries 2015) and probably in most cases, as Lakoff and Johnson (1999: 499) point out, "polysemy is the norm". Commonly quoted examples of multiple meanings of word include "game" (Wittgenstein 2001: 27) and "mother" (Lakoff 1987: 74). Metaphor is an important way of meaning creation because new meanings of a word can be derived via metaphorical extension (Johnson 1987: 192; Lyons 1995: 59). Cognitive linguists believe that metaphor is a way we think which is prevalent in our life (Lakoff and Johnson 2003). This belief goes beyond the traditional view that metaphor is no more than a rhetorical device. Many researchers have proposed explanations for origins of metaphor's powerful capability of meaning creation. Experiential basis and cultural constraints

are arguably two major factors that provide a solid foundation for metaphor as figurative language use.

Experiential basis refers to the belief that metaphor is deeply rooted in human experience (Jing-Schmidt 2008; Johnson 1987: 15); the use of metaphor is not arbitrary but motivated by "physical, social and cultural experience" of the language community (Chandler 2002: 129). Cognitive semantics demonstrates that "the metaphor system is not arbitrary, but is also grounded in experience" (Lakoff and Johnson 1999: 497). This is consistent with the viewpoint of a larger framework that Lakoff and Johnson formulate in their study of the "embodied mind": "Abstract concepts are largely metaphorical. This means that most of our non-physical (social, psychological, etc.) reality is conceptualized via physical reality, that is, in terms of physical domains of experience" (Kövecses 2005: 10). In other words, metaphor is the "indirect method" we use to understand abstract concepts to which we do not have direct access (Yule 2014: 157); metaphor is not directly based on the reality, the "really real", but on the reality which is experienced by humans (Alverson 1991: 98). A study from the evolutionary perspective has also shown that "(to) integrate sensory and affective experiences" is an important ability that underlies metaphor (Seitz 2005: 82).

Another claim within the cognitive linguistics paradigm, the encyclopedic view of lexicon, is also consonant with metaphor's experiential basis. As Evans and Green (2006: 221) maintain, "the encyclopedic model views lexical items as points of access to encyclopedic knowledge". This view reflects in the other way around that human experience is behind the metaphor. For example, the acceptability of the +LOVE IS A JOURNEY+ metaphor arises from the solid foundation of human experience of journey and the signification of this metaphor is supported by human "motor-sensory image base" which in this case is journey-related events (Hiraga 2005: 6).

Regarding cultural constraints, it is linguists who first realize that culture can sanction word meaning (Quinn and Holland 1987: 23). For metaphors, which have figurative meanings, culture also plays a pivotal role in metaphorical usage and especially in metaphor selection (Wee 2006). Kövecses' study lends strong support to the cultural constraints on metaphorical usage within a certain language (Kövecses 2005). He points out although universality of metaphor can be detected, "cross-cultural variation" and "within-culture variation" do exist in metaphorical usage. In other words, types of metaphor can vary according to different cultures. Ning Yu (2007; 2011) study of metaphors related to the concept of heart and Chinese Beijing opera respectively further further substantiate this view. Also, consonant with the idea of cultural constraints on metaphor, Steen (2011) touches upon the concept of conventionalization of metaphor and adopts the view that metaphor can be regarded as a cultural phenomenon rather than an individual one. Metaphor as a complete novel coinage cannot be understood by

other interlocutors unless in line with a certain cultural convention. Furthermore, according to semiotics, connotation is "the socio-cultural and 'personal' associations (ideological, emotional, etc.) of the sign" (Chandler 2002: 140) and thus a metaphor as a sign is constrained by culture and a metaphor needs to conform to connotations associated with culture so that it is possible to be accepted by the language community (Tarasti 2016).

Finally, kinship metaphor is an ideal test field for us to view experiential basis of and cultural constraints on metaphor because of both the human beings' common, concrete, and physical experience of kinship upon which metaphors can be grounded (Turner 1987) and markedly different viewpoints on kinship by members of different cultures. Matisoff (1991: 299) conducted a comprehensive study on kinship metaphors, especially those expressed by "mother", in East and Southeast Asian languages. He found that expressions for "child" and "father" as well as "mother" in these languages have been grammaticalized and "are often metaphorically opposed as diminutive vs. augmentative, or masculine vs. feminine". Furthermore, he maintains that the metaphorical extensions and grammaticalization of "mother", instead of "father", seem to be a general pattern and thus he coins the term "mothermorphs".

This study attempts to overcome the shortcomings of Matisoff's introspection data by drawing on a linguistic corpus. The focus is on two metaphorically used kinship terms in Modern Chinese, *fu* (father) and *mu* (mother), which are two major kinship terms for Chinese people. It is argued that such metaphors are motivated by experiential basis and sanctioned by cultural constraints.

2.   Data for analysis

The data are obtained from the Modern Chinese section of CCL Corpus.[1] Under investigation are two constructions of two Chinese kinship terms, 父 *fu* 'father' and 母 *mu* 'mother': [A是B之父] *[A shi B zhi fu]* 'A is the father of B' and [A是B之母] *[A shi B zhi mu]* 'A is the mother of B'. Turner's (1987) study on kinship metaphors in English, which also focused on constructions of "A is the father of B" and "A is the mother of B", served as important guidance for the choice of these two constructions in Chinese. In the current study, each construction is summarized from the meaning of each concordant sentence. For some sentences, the meaning is rather straightforward in that they state "A is the father/mother" in its literal form. For others, semantic abstraction and simplification are needed. For example, one sentence from the data pool is "'杂交水稻之父'、全国政协委员

---

1. http://ccl.pku.edu.cn:8080/ccl_corpus/

袁隆平7日在全国政协十届二次会议第二次全体会议上作大会发言时提出".
Although this is a long sentence, the focus of this study is only on the apposition
part, that is

(1) '杂交水稻 [之父]'、全国政协委员袁隆平…
    zajiaoshuidao zhi fu   quanguozhengxieweiyuan   yuan
    hybrid rice   of father CPPCC National Committee member Yuan
    longping
    Longping
    'Yuan Longping, a CPPCC National Committee member, is the father of
    hybrid rice'

In terms of the construction, we could simplify it as 袁隆平是杂交水稻之父
yuan longping shi zajiaoshuidao zhi fu 'Yuan Longping is the father of hybrid rice'
in which A = 袁隆平 (Yuan Longping) and B = 杂交水稻 (hybrid rice).[2] Because
the relationship between "Yuan Longping" and "father of hybrid rice" are apposi-
tional, we can make that simplification. Therefore, in this study, sentences are first
simplified into such construction before they are coded for statistical analysis.

    In the two aforementioned constructions, A refers to the expression of "par-
ents" (*fu* or *mu*) while B refers to "offspring". This can be explained by the reciproc-
ity of the kinship term in Chinese (Blum 1997: 361). When father or mother is
mentioned, the child or the offspring is implied. For example:

(2) 田径是运动[之母]
    tianjing   shi yundong zhi mu
    track and field be sports   of mother
    'Track and field is the mother of sports'

Although the construction only literally says "track and field" is *mu* 'mother', we
could infer that "sports" is offspring. So in this example, A = 田径 (track and
field) = mu (mother) and B = 运动 (sports) = offspring.

    For the data from the CCL corpus, the keywords are 之父 *zhifu* 'father of' and
之母 *zhimu* 'mother of'. The keywords are marked between brackets "[ ]" in each
case when demonstrated in this paper. The span of each concordant sentence is
100 Chinese characters on both left and right sides of the key words. In the initial
stage, 1426 concordant sentences related to *fu* and 630 related to *mu* were found.
Sentences concerning religious figures and beliefs were then excluded because of
the extreme difficulty in identifying their referents. Also, repeated and "wrong"
data were excluded.

---

2. This and all the following translations from Chinese to English are made by the author of
this paper.

"Repeated" means that some sentences, despite that they are different super-ficially, express the same meaning in terms of the schematic construction [A *shi* B *zhi fu/mu*].[3] One example of this has been mentioned above, that is, 袁隆平是 杂交水稻[之父] 'Yuan Longping is the father of hybrid rice'. More than 10 tokens from the data are related to the information about Yuan in terms of his innova-tion of the hybrid rice, but they as a whole are only counted as one type since they can be simplified and abstracted into the same construction: A = 袁隆平 (Yuan Longping) and B = 杂交水稻 (hybrid rice) and A is the father of B. For this rea-son, the somewhat 10 sentence tokens about Yuan are counted as one sentence type (compared with the variable "Metaphor Type" mentioned in Methodology section) and it is based on the sentence types that metaphor types are identified.

"Wrong" means that some sentences have "之父" or "之母" in them, but they do not convey the meaning of "father of" or "mother of", as is shown in …王羲[之父]子…, which contains the keyword but it is actually the juxtaposed two char-acters belonging to two different consecutive phrases.

To sum up, with the repeated and "wrong" data deleted, the final data pool consists of 922 sentence types (including both metaphorical and non-metaphori-cal) with 666 related to *fu* and 256 related to *mu*.

3.   Methodology

3.1   Metaphor identification

It is indeed necessary to clearly state the criteria for metaphor identification for this study. According to 现代汉语词典 *xiandai hanyu cidian* 'The Dictionary of Modern Chinese' (the 6th edition) (Chinese Academy of Social Sciences Institute of Languages Dictionary Editorial Room 2013), the basic meaning of *fu* and *mu* is "father" and "mother" respectively. Therefore, in this study the criterion "A and B have the relationship of procreator and procreatee" is used to judge whether a construction is metaphorical or not. If a construction satisfies this criterion, it will be counted as a non-metaphorical or literal token; otherwise, it will be considered metaphorical.

To make a detailed distinction between *fu* and *mu*, this study benefits from the componential analysis (Yue 2012; Yule 2014: 111) in semantics by using two semantic components [+human] and [+specific] to further demonstrate that the metaphoricity of *mu* is different from that of *fu*. By this token, 牛顿 *niudun*

---

3. By "schematic", I mean this schematic construction can have two variations, [A *shi* B *zhi fu*] and [A *shi* B *zhi mu*].

'Newton' will be coded as having both semantic components; 导弹 *daodan* 'missile' having [+specific] but not [+human]; 几何 *jihe* 'geometry' having neither component. The foremost uses of "mother" and "father" refer to human relationships so that their implied offspring should also be human in its basic sense. Accordingly, if A or B in the construction does not have the semantic component [+human], it means the sentence with the construction is metaphorically used in that humans are drawn on to denote nonhuman entities. For another component [+specific], it can be understood that the basic sense of *fu* or *mu* denotes specific entities because human father and mother are specific entities in their basic biological sense and so are the implied offspring. Therefore, if A or B in the construction does not have the semantic component [+specific], we can believe sentences with this construction are metaphorical in nature. This judgment is also in line with the commonly accepted belief about metaphor: via metaphor, abstract things are understood by specific things (Lakoff and Johnson 2003). Consequently, it is believed that the less A or B in [A *shi* B *zhi fu/mu*] contains these two components, the more metaphorical the construction will be.

Therefore, the criterion for metaphor identification in this study can be stated as whether A and B have the relationship of procreator and procreatee. Two semantic components [+human] and [+specific] will be adopted to further demonstrate the distinction between the metaphoricity of *fu* and *mu*.

3.2   Statistical procedure

For statistical analysis, seven variables are established in the SPSS software: (1) Research Object; (2) A has [+human] or not; (3) B has [+human] or not; (4) A has [+specific] or not; (5) B has [+specific] or not; (6) A and B have the relationship of procreator and procreatee or not; (7) Metaphor Type (i.e. conceptual metaphors summarized from sentence types).

For variable (1) the possible values are 1 = "fu (父)" and 2 = "mu (母)" and for variables (2) to (6) are 1 = yes and 2 = no. Variable (7) have 11 possible values because 11 metaphor types are identified and the specific values will be presented in the Results section. As mentioned earlier, since for metaphorical sentence types, A and B in [A *shi* B *zhi fu/mu*] does not have the relationship of procreator and procreatee, the different metaphor types are summarized based on (1) nuanced different relationships between A and B (for example, "A creates B" and "A is the most distinguished of all B" are regarded as different relationships and thus different metaphor types); (2) different kinships, father or mother (for example, the relationship "A (as father) creates B" is different from "A (as mother) creates B"); and (3) different semantic referents of A and B (for example, "A (A has [+human]) makes a primary contribution to an undertaking" is different from "A (A

does not have [+human]) makes a primary contribution to an undertaking"). For reasons of space, one example is provided here. 张丽珠教授是中国试管婴[之母] *Zhanglizhu jiaoshou shi zhongguo shiguanying zhi mu* 'Professor Zhang Lizhu is the mother of China's test tube baby' is coded as the metaphor type "A PERSON WHO MAKES PRIMARY CONTRIBUTION TO AN UNDERTAKING IS MOTHER" because the sentence type is about *mu* 'mother' and the meaning of this sentence can be summarized as "Professor Zhang makes a primary contribution to the undertaking of research in China's test tube baby".

4. Results

The results show that the majority of corpus sentences are metaphorical (75.9%, 700 in total for metaphorical sentence types) in that A and B do not have the relationship of procreator and procreatee. This indicates that sentences with the schematic construction [A *shi* B *zhi fu/mu*] are often metaphorical. Table 1 and Table 2 show the conceptual Metaphor Types for *fu* and *mu* identified from the data.

We can see that for *fu* the predominant metaphor type is + A PERSON WHO MAKES PRIMARY CONTRIBUTION TO AN UNDERTAKING IS FATHER+ and it takes up 94.9% of all metaphor types for *fu*. By contrast, the data for *mu* show a much more various distribution of metaphor types. The most frequent metaphor type for *mu* is +AN ENTITY THAT CREATES ANOTHER ENTITY IS MOTHER+, but it only takes up 46.8%. Obviously, in this sense, in

Table 1. Metaphor types for *fu*

Metaphor type for *fu*		Frequency	Valid percent
Valid	A PERSON WHO MAKES PRIMARY CONTRIBUTION TO AN UNDERTAKING IS FATHER	502	94.9
	AN ENTITY THAT MAKES PRIMARY CONTRIBUTION TO AN UNDERTAKING IS FATHER	11	2.1
	AN ENTITY THAT IS THE MOST DISTINGUISHED OF ALL IS FATHER	8	1.5
	AN ENTITY THAT CREATES ANOTHER ENTITY IS FATHER	6	1.1
	A PERSON WHO IS THE MOST DISTINGUISHED OF ALL IS FATHER	2	.4
	Total	**529**	**100.0**

Table 2. Metaphor types for *mu*

Metaphor type for *mu*		Frequency	Valid percent
Valid	AN ENTITY THAT CREATES ANOTHER ENTITY IS MOTHER	80	46.8
	AN ENTITY THAT MAKES PRIMARY CONTRIBUTION TO AN UNDERTAKING IS MOTHER	32	18.7
	A PERSON WHO MAKES PRIMARY CONTRIBUTION TO AN UNDERTAKING IS MOTHER	28	16.4
	AN ENTITY THAT IS THE MOST DISTINGUISHED OF ALL IS MOTHER	18	10.5
	AN ENTITY THAT IS CRUCIAL TO ANOTHER ENTITY IS MOTHER	11	6.4
	A PERSON WHO IS THE MOST DISTINGUISHED OF ALL IS MOTHER	2	1.2
	Total	**171**	**100.0**

the schematic construction [A *shi* B *zhi fu/mu*], *mu* has more evenly distributed multiple referents.

Below are examples of those Metaphor types, *fu* and *mu* respectively with one example for each type:

Metaphor for *fu*
+A PERSON WHO MAKES PRIMARY CONTRIBUTION TO AN UNDERTAKING IS FATHER+

(3) 欧几里得是几何[之父]
   Oujilide shi jihe   zhi fu
   Euclid   be geometry of father
   'Euclid is the father of geometry'

+ A PERSON WHO IS THE MOST DISTINGUISHED OF ALL IS FATHER +

(4) 卡鲁索是意大利男高音之父
   Kalusuo shi Yidali nangaoyin   zhi fu
   Caruso be Italian tenor singers of father
   'Caruso is the father of Italian tenor singers'

+ AN ENTITY THAT MAKES PRIMARY CONTRIBUTION TO AN UNDERTAKING IS FATHER +

(5) 休斯公司是当代卫星[之父]
　　 xiusi　　gongsi　　shi dangdai　　　　weixing zhi fu
　　 Hughes company be contemporary satellite of father
　　 'Hughes Company is the father of contemporary satellite'

+AN ENTITY THAT CREATES ANOTHER ENTITY IS FATHER+

(6) 金沙泉水是茶[之父]
　　 jinshaquan　　shui　shi cha zhi fu
　　 Jinsha Springs water be tea of father
　　 'The water of Jinsha Springs is the father of tea'

+AN ENTITY THAT IS THE MOST DISTINGUISHED OF ALL IS FATHER+

(7) 慕士塔格峰是冰山[之父]
　　 mushitage　　feng　shi bingshan zhi fu
　　 Muztagh Ata peak be iceberg of father
　　 'Muztagh Ata is the father of iceberg'
　　 Metaphor for *mu*

+A PERSON WHO MAKES PRIMARY CONTRIBUTION TO AN UNDERTAKING IS MOTHER+

(8) 张丽珠教授是中国试管婴[之母]
　　 zhanglizhu　　jiaoshou　shi zhongguo shiguanying　　zhi mu
　　 Zhang Lizhu Professor be China　　test tube baby of mother
　　 'Professor Zhang Lizhu is the mother of China's test tube baby'

+ AN ENTITY THAT MAKES PRIMARY CONTRIBUTION TO AN UNDERTAKING IS MOTHER +

(9) 碱是工业[之母]
　　 jian　shi gongye　zhi mu
　　 alkali be industry of mother
　　 'Alkali is the mother of industry'

+ AN ENTITY THAT CREATES ANOTHER ENTITY IS MOTHER +

(10) 山是水[之母]
　　 shan　　shi shui　zhi mu
　　 mountain be water of mother
　　 'Mountain is the mother of water'

+ AN ENTITY THAT IS CRUCIAL TO ANOTHER ENTITY IS MOTHER+

(11) 爱是教育[之母]。
      Ai shi jiaoyu   zhi mu
      love be education of mother
      'Love is the mother of education'

+ AN ENTITY THAT IS THE MOST DISTINGUISHED OF ALL IS MOTHER +

(12) '十二木卡姆'是维吾尔族"音乐[之母]"
      shiermukamu shi weiwuerzu yinyue zhi mu
      Mukam   be Uighur   music of mother
      'Mukam is the mother of Uighur music'

+ A PERSON WHO IS THE MOST DISTINGUISHED OF ALL IS MOTHER +

(13) 南丁格尔是护理人员之母
      nandinggeer shi hulirenyuan zhi mu
      Nightingale be nursing staff of mother
      'Nightingale is the mother of nursing staff'

To sum up, as Finding 1, sentences with the schematic construction [A *shi* B *zhi fu/mu*] are often metaphorical in that A and B in this kind of construction generally are not in the relationship of procreator and procreatee. For sentences with *fu*, it is normative that *fu* refers to "some person makes a primary contribution to an undertaking" while *mu* can have relatively evenly distributed multiple referents.

Now we can take a close look at the distinction of metaphorical usage by the two components [+human] and [+specific]. We will first look at the distribution of each semantic component in the two constructions.

For the semantic component [+human], A in [A *shi* B *zhi fu/mu*] (77.1%) Often has this semantic component. As Table 3 shows:

Table 3. The frequency of A having the semantic component [+human]

		Frequency	Valid percent
Valid	+human	540	77.1
	−human	160	22.9
	Total	**700**	**100.0**

This is equal to say that A in the construction often denotes human referents. For example:

(14) 莱特兄弟是飞机[之父]
    laite   xiongdi shi feiji   zhi fu
    Wright brothers be airplane of father
    'Wright brothers are the father of airplane'

In this case, A = 莱特兄弟 (Wright brothers) and B = 飞机 (airplane).

The situation is different for B in [A *shi* B *zhi fu/mu*] because in most cases B does not have the semantic component [+human]. Table 4 tells us this information:

Table 4. The frequency of B having the semantic component [+human]

		Frequency	Percent	Valid percent	Cumulative percent
Valid	+human	18	2.6	2.6	2.6
	−human	682	97.4	97.4	100.0
	Total	**700**	**100.0**	**100.0**	

It is shown that in 97.4% cases B does not denote human referents, but nonhuman entities. For example,

(15) 南丁格尔是护理学[之母]
    nandinggeer shi hulixue zhi mu
    Nightingale be Nursing of mother
    'Nightingale is the mother of Nursing'

In this case, B = Nursing and B does not denote a human but a discipline.

The situation is similar for the semantic component [+specific]. Table 5 shows that in 86.3% of all cases, A denotes specific entities including human being(s) or other specific entities such as 牛顿 *Niudun* 'Newton', 碱 *jian* 'alkali', etc.:

Table 5. The frequency of A having the semantic component [+specific]

		Frequency	Percent	Valid percent	Cumulative percent
Valid	+specific	604	86.3	86.3	86.3
	−specific	96	13.7	13.7	100.0
	Total	**700**	**100.0**	**100.0**	

On the contrary, Table 6 shows that in 74.9% of all cases, B in [A *shi* B *zhi fu/mu*] does not denote specific entities:

Table 6. The frequency of B having the semantic component [+specific]

		Frequency	Percent	Valid percent	Cumulative percent
Valid	+specific	176	25.1	25.1	25.1
	−specific	524	74.9	74.9	100.0
	Total	**700**	**100.0**	**100.0**	

B frequently denotes abstract entities such as 力学 *lixue* 'mechanics', 欧洲音乐 *ouzhou yinyue* 'European music', etc. but not specific entities – this is the Finding 2: For metaphorical sentences with schematic construction [A *shi* B *zhi fu/mu*], A in the construction tends to denote human beings or specific entities while B tends to denote nonhuman and nonspecific entities, i.e. the "offspring" tends to be nonhuman and/or abstract, such as a discipline, a concept, etc.

Next, for the difference between [A *shi* B *zhi fu*] and [A *shi* B *zhi mu*] in the metaphorical cases (700 in total), the two semantic components [+human] and [+specific] are used to explain. Chi-square analysis shows three significant associations between research objects and the variables whether the two research objects have the two semantic components respectively, which will be discussed in the following: (1) research object and A has [+human] (X^2 (1) = 446.924, $p < .01$. Cramer's $V = .799$); (2) research object and B has [+human] (X^2 (1) = 13.466, $p < .01$. Cramer's $V = .139$); (3) research object and A has [+specific] (X^2 (1) = 247.725, $p < .01$. Cramer's $V = .595$); (4) research object and B has [+specific] (X^2 (1) = 1.148, n.s.).

4.1   Research object and A has [+human]

Table 7 shows that construction with *fu* and that with *mu* are different in whether A in each construction has the semantic component [+human] or not:

Table 7. Research Object * A has [+human] crosstabulation

			+human	−human	
Research Object	fu (父)	Count	509	20	529
		Expected Count	408.1	120.9	529.0
		Standardized Residual	5.0	−9.2	
	mu (母)	Count	31	140	171
		Expected Count	131.9	39.1	171.0
		Standardized Residual	−8.8	16.1	
Total		**Count**	**540**	**160**	**700**
		Expected Count	540.0	160.0	700.0

The majority of A in [A *shi* B *zhi fu*] denote "human" (509 out of 529), which means *fu* (father) in these cases are generally human. On the contrary, A in [A *shi* B *zhi mu*] frequently does not denote "human" (140 out of 171); in other words, *mu* (mother) is not human in most cases. Further examination of the data provides example sentences such as,

(16) 土地是财富[之母]
      tudi shi caifu   zhi mu
      earth be wealth of mother
      'The earth is the mother of wealth'

(17) 市场是企业[之母]
      shichang shi qiye   zhi mu
      market   be enterprise of mother
      'The market is the mother of enterprises'

The fact that *mu* (mother) in metaphorical cases denote nonhuman entities indicates that *mu* is more metaphorical than *fu*. Explanations in Chinese dictionaries can also lend support to this judgment, as is shown in Table 8:

Table 8. Dictionary entries for 父 and 母

	Explanation in 《辞源》 (The Commercial Press 2007)	Explanation in 《现代汉语词典》 (Chinese Academy of Social Sciences Institute of Languages Dictionary Editorial Room 2013)
fu 父	（1） 父亲 'father'; （2） 男性长辈的通称 'the general term for seniors in one's family'.	（1） 父亲 'father'; （2） 家族或亲戚中的长辈男子 'a term for male seniors in a family or among relatives'.
mu 母	（1） 母亲 'mother'; （2） 称女性尊长。如叔母、外祖母等 "a term for female seniors, e.g. 'the wife of one's uncle', 'mother's mother'"; （3） 老妇的通称 'the general term for female who is in her old years'; （4） 雌性 'the term for female creature'; （5） 本源 *'origin or source'*; （6） 币有大小轻重,其大者重者称"母",小者轻者称"子" 'the term for the bigger the currency dyad in ancient China'; （7） 能使他物滋生者 *'a term for something that can create other things'*.	（1） 母亲 'mother'; （2） 家族或亲戚中的长辈女子 'a term for female seniors in a family or among relatives'; （3） 属性词 'a term for quality'. （禽兽）雌性的（跟"公"相对） "a term meaning opposite to '公' (usually used to describe beasts)"; （4） 指一凸一凹配套的两件东西里的凹的一件 'a term for the concave part of a concave-convex dyad'; （5） 有产生出其他事物的能力或作用的 *'of something that can produce something else'*; （6） 姓 'a surname'.

Explanation for *fu* is invariably "father", and this should be its basic sense. At the same time, entry 5 and 7 from 辞源 *ciyuan* and entry 5 from 现代汉语词典

xiandai hanyu cidian demonstrate that in the sense of "production, procreation" (not confined to humans' procreation) *mu*'s metaphorical usage has been highly conventionalized and thus already complied into dictionaries. This finding that *mu* enjoys higher metaphoricity is also in line with cross-linguistic data provided by Matisoff (1991). He found that in Thai, mɛ̂ɛ (mother) can have the metaphorical sense of 'chief' and, in this sense, it can be extended to refer to "people of unspecified or even usually male gender", as in "mɛ̂ɛ-tháp" which means 'general, commander in chief'. Also, nā 'mother' in Green Hmong dialect can metaphorically mean 'big, large, great' which can be extended to be used in gender-neutral expressions, for example, "nā-klě" which means 'big dog (either male or female)'. Finally, since the data in this study suggest that a predominant metaphorical usage for *fu* is + A PERSON WHO MAKES PRIMARY CONTRIBUTION TO AN UNDERTAKING IS FATHER+, this meaning, i.e. "a person who makes a primary contribution to an undertaking" might be added as one new dictionary entry to the explanations of *fu* because of its high frequency shown in the data (See Table 1).

4.2   Research object and B has [+human]

A weak association exists between research objects and the semantic component [+human] (Cramer's V = .139).

Table 9. Research object * B has [+human] crosstabulation

			+human	−human	
Research Object	fu (父)	Count	7	522	529
		Expected Count	13.6	515.4	529.0
		Standardized Residual	−1.8	.3	
	mu (母)	Count	11	160	171
		Expected Count	4.4	166.6	171.0
		Standardized Residual	3.1	−.5	
Total		**Count**	**18**	**682**	**700**
		Expected Count	18.0	682.0	700.0

As is shown in Table 9, *fu* is slightly likely to have a nonhuman offspring (i.e. B in [A *shi* B *zhi fu*]) while *mu* is slightly likely to have a human offspring. However, since this study focuses on metaphors related to father and mother, not the offspring, and the association is weak, arguably this association will not be included as a criterion for metaphoricity judgment of the two kinship terms and thus I will not press this point further.

4.3 Research object and A has [+specific]

Table 10 shows the third significant association between research object and A in [A *shi* B *zhi fu/mu*] has [+specific].

Table 10. Research object * A has [+specific] crosstabulation

			+specific	-specific	
Research Object	fu (父)	Count	518	11	529
		Expected Count	456.5	72.5	529.0
		Standardized Residual	2.9	−7.2	
	mu (母)	Count	86	85	171
		Expected Count	147.5	23.5	171.0
		Standardized Residual	−5.1	12.7	
Total		**Count**	**604**	**96**	**700**
		Expected Count	604.0	96.0	700.0

This means in most cases (518 out of 529), A in the construction [A *shi* B *zhi fu*] has the semantic component [+specific] and therefore *fu* frequently denotes specific entity or a (specific) human being. As the previous Example (7) and this Example (18) show,

(18) 牛顿是力学[之父]
niudun shi lixue   zhi fu
Newton be mechanics of father
'Newton is the father of mechanics'
Only a few cases involving *fu* denote nonspecific entities such as,

(19) 美丽是圣洁[之父]
meili   shi shengjie zhi fu
beauty be purity   of father
'Beauty is the father of purity'

(20) 劳动是财富[之父]
laodong shi caifu   zhi fu
labor   be wealth of father
'Labor is the father of wealth'
Things are different for *mu*. For 85 out of 171 cases, A does not have the semantic component of [+specific] which means *mu* in those sentences denotes nonspecific, abstract entities. Examples include,

(21)   教育是科技[之母]
jiaoyu   shi keji   zhi mu
education be science and technology of   mother
'Education is the mother of science and technology'

(22)   需要是发明之母
xuyao   shi faming   zhi mu
necessity be invention of mother
'Necessity is the mother of invention'

In summary, this is Finding 3: sentences with *mu* are more metaphorical than those with *fu* in that, regardless of the status of offspring, *mu* normally denotes a nonhuman entity and that entity is much more likely than *fu* to be abstract, non-specific. Therefore, metaphorical sentences with *mu* can mean "a nonspecific thing is the mother of something else." whereas *fu* predominantly denotes specific entities.

5.   Discussion

Findings 1 and 2 suggest that metaphorical sentences with the [A *shi* B *zhi fu/mu*] construction generally mean that a human being or a specific entity is the father/mother of a nonhuman and/or nonspecific entity. This is the prototypical form of this metaphor in such sentences. Also, in the typical usage, the choice of *fu* or *mu* depends on the biological classification between the human referent in terms of gender (Yule 2014: 274). In other words, if the human being as A in [A *shi* B *zhi fu/mu*] is male or female, *fu* or *mu* is chosen respectively.

However, three things are worth mentioning here. First, exceptions for the gender-based selection are found from the data, as is shown below,

(23)   鲁迅是现代中国的国民文化[之母]
luxun   shi xiandai   zhongguo de   guomin wenhua zhi
Lu Xun be contemporary China   GEN(genitive) people culture of
mu
mother
'*Lu Xun* is the mother of people's culture in contemporary China'

(24)   亨德尔是音乐[之母]
hengde'er shi yinyue zhi mu
Handel   be music of *mother*
'Handel is the mother of music'

Lu Xun and Handel were males, but they are still referred to as "the mother of…" in metaphorical sentences. Although such cases are rare, it seems that *mu* enjoys higher metaphoricity because it can be used to denote such "male mother" while none "*female father" is found.

Second, as is mentioned in the Results section, *fu* frequently refers to "some person makes a primary contribution to an undertaking" because in 230 out of 244 cases a person making primary contribution to an undertaking is called "father" for the sake of establishing a discipline or theory, inventing something, founding a nation or organization, creating a fictional figure or playing a major role in a spirit or belief. At the same time, *mu* metaphors have a rather evenly distribution which means no one kind of *mu* metaphor has predominantly more instantiations than others. This might remind us of the dominance-approach in sociolinguistic studies on gender (Eckert and McConnell-Ginet 2003) which believes that male plays a dominant role in society and that causes the linguistic gender difference. Since in most undertakings, be it establishing a discipline or founding a nation, we always have "founding fathers" instead of "founding mothers", data in this study show that the *fu* metaphor invariably concerns a male who makes the primary contribution to an undertaking. Then we can say this is an example that social motivation underlies metaphorical usage. In addition, this distinction fits well into the Chinese cultural context. In the traditional patriarchal discourse, the division of labor is that Chinese fathers as breadwinner focus on making a living and therefore Chinese mothers are mainly responsible for child nurturing; it is believed that Chinese mothers have a higher level of maternal involvement in child-rearing and child development than European-American mothers (To 2015). At the same time, although in the Chinese context male plays a dominant role in society, it is actually female who "has a commanding force" at home (Gardner 2014: 110). Among others, a mother is the primary birth-giver, caretaker, child's first teacher, and household manager (ibid.: 111). Despite Chinese "father" makes a primary contribution in the social realm, "mother" has more multiple and important parts to play in the domestic realm. Chinese mother is omnipresent in child's life. As a result, for "mother" (*mu*) metaphor, it is likely to have been conceptualized to extend to more dynamic target domains than in the case of "father" (*fu*) metaphor. We have seen in the Results section that several evenly distributed metaphor types can be found for "mother".

Third, careful scrutiny of the carrier sentences reveals that the *fu* and *mu* metaphors often have positive connotations. As is shown in the following example: 陈霜莹…被誉为"中国的富士[之母]" *Chen Shuangying…beiyuwei zhongguo de fushi zhi mu* 'Chen Shuangying…is known as the mother of Fuji apple in China'. The expression "被誉为…" *beiyuwei* 'be known as' is used which means 'somebody or something has a positive reputation for…'. This common usage suggests

that people associate *fu* and *mu* metaphors with positive things or at least non-negative ones. Other similar expressions that have been found in the data pool include 雅称 *yacheng*, 号称 *haocheng*, 荣称 *rongcheng*, etc. One possible explanation is the influence of filial piety which is "a centerpiece of the teachings of Confucius" (Gardner 2014: 97). Under the influence of Confucianism, great significance is attached to showing respect for one's parents in the cultural realm and this attitude is exemplified by the metaphor usage under investigation. The meanings conveyed by the metaphors are also positive overall such as "making primary contribution" and "most distinguished".

Finding 3 suggests that *mu* is more metaphorical than *fu* in that *mu* usually denotes nonhuman, nonspecific entities. This can be explained by experiential basis of and also a cultural constraint on metaphor.

For experiential basis, antenatal development for female parent is universal for all human beings so that it is reasonable to believe that Chinese people tend to rely more on *mu* instead of *fu* to conceptualize abstract phenomena via metaphor. In this sense, metaphor related to *mu* can powerfully serve as mediator between physical experience and abstract understanding (Johnson 1987: XV).

In terms of the cultural constraint, characteristics of Chinese culture also lends support to the claim that *mu* is more metaphorical than *fu*. Lin (2000: 150) argues that the expectation of a Chinese woman is to be a mother and "propagate the race". In the Confucian tradition, both human and animal mothers were assumed to have natural feelings of compassion and concern for their babies (Birdwhistell 2007: 112). This cultural norm can be supported by the affectionate name entitled to Yellow River, a river believed to be the origin of Chinese civilization, as 母亲河 *muqin he* "(literally) 'mother river'" rather than as *父亲河 *fuqin he* "(literally) 'father river'", which is in contrast with German: Rhein is literally called "Father Rhein" (Mercurio 2016, 'Tales of Ritona'). The examples of metaphors related to *mu* and the affectionate name given to the Yellow River reflect the influence of cultural knowledge (Quinn 1991) on metaphor selection. When the two terms *fu* and *mu* have equal status in a pure linguistic sense, the one that "fit(s) a preexisting and culturally shared model" (ibid.: 60), that is *mu*, is preferred and selected in metaphorical usage (Mischler III 2013: 12). From the perspective of semiotics, Tarasti (2016: 121) claims, metaphors are not only signs, but also "signs plus something". Thus, the selection of father or mother in the source domain of metaphors in this Chinese case buttresses the existence of that "something", which is the set of cultural constraints. To press this point further, apart from the foundational role of culture relative to metaphor, as is claimed by Wee (2006) that linguistic expressions can be manifestations of an underlying cultural model, this study indicates that culture can also play a constraining role in the genesis of metaphor when alternatives occur, father or mother in the kinship case.

To sum up at this point, in the Chinese language, it is easier for Chinese people to accept a procreator, whether a physical one or a metaphorical one, as *mu* than *fu*, probably because of the common sense of mother giving birth and the image of a mother shaped by Chinese culture. *Mu* metaphor seems to be what Steen (2011: 72–73) called "metaphor in grammar" since the metaphorical meaning has been in dictionaries while *fu* metaphor seems to be "metaphor in usage" with no dictionary entry about the metaphorical meaning found yet. Therefore, echoing the concepts of "cross-cultural variation" and "within-culture variation" proposed by Kövecses (2005), this study again shows that metaphor usage is characterized by within-cultural variation, the domain of kinship in particular and if future studies can systematically compare the kinship metaphor usage in Chinese with those in other languages, it is envisaged that cross-cultural variation in this regard can also be found.

6.   Conclusion

This study focuses on the two kinship terms in Chinese 父 *fu* 'father' and 母 *mu* 'mother' in the schematic construction [A *shi* B *zhi fu/mu*]. The two terms are frequently metaphorically used. A closer look at the data demonstrates that the metaphoricity of *mu* is higher than that of *fu* in that *mu* in the data generally denotes something that is not only nonhuman but also nonspecific while *fu* metaphor is always "founding-father-like" because *fu* denotes predominantly someone making a primary contribution to an undertaking. Two possible reasons can account for this distinction. One is the universal reality of female's antenatal development which is the prototypical scenario of giving birth and this serves as the primary experiential basis for such kinship metaphor usage. The other is the Chinese cultural constraint. A typical Chinese mother image is the parent who is supposed to play multiple roles in family life while a father as male is in a dominant position in matters regarding social realm. These two factors are thought to "give birth to" the high metaphoricity of *mu* and this high metaphoricity is also identified in other East and Southeast Asian languages (Matisoff 1991).

Limitations exist in this study. The metaphor types are all summarized based on my own understanding of sentences from the corpus without reference to other native speakers. At the same time, this has been ameliorated by referring to information of contexts for difficult sentences, which is empowered by the corpus. Also, the coding process is all done by myself. Therefore, a checking process was conducted to ensure high-quality coding.

Overall, this study again shows the soundness of a non-Objective theory of meaning (Johnson 1987: 175) which believes meaning is not a product of checklist

but experientially and culturally motivated (Yu 2009). Also importantly, this study shows that metaphorical usages of constructions "A is the father/mother of B" can shed light upon the nexus of language, culture, and society in a specific language, which is again a possible research topic for cross-linguistic comparison.

References

Alverson, Hoyt. 1991. "Metaphor and Experience: Looking over the Notion of Image Schema." In *Beyond Metaphor: The Theory of Tropes in Anthropology*, 94–117. Stanford University Press.

Birdwhistell, Joanne. 2007. *Mencius and Masculinities: Dynamics of Power, Morality, and Maternal Thinking*. Albany: State University of New York Press.

Blum, Susan D. 1997. "Naming Practices and the Power of Words in China." *Language in Society* 26 (3): 357–79.

Chandler, Daniel. 2002. *Semiotics: The Basics*. London ; New York: Routledge.

Chinese Academy of Social Sciences Institute of Languages Dictionary Editorial Room. 2013. 现代汉语词典 [The Contemporary Chinese Dictionary]. Beijing: The Commercial Press.

Commercial Press, The. 辞源 [The Etymology]. 2007. Beijing: The Commercial Press.

Cowie, A. 2012. *Semantics*. Shanghai Foreign Language Education Press.

Eckert, Penelope, and Sally McConnell-Ginet. 2003. *Language and Gender*. Cambridge University Press.

Evans, Vyvyan, and Melanie Green. 2006. *Cognitive Linguistics: An Introduction*. Lawrence Erlbaum Associates Publishers.

Gardner, Daniel. 2014. *Confucianism: A Very Short Introduction*. Oxford University Press.

Gries, Stefan. 2015. "Polysemy." In *Handbook of Cognitive Linguistics. Vol. 39*, ed. by Dąbrowska, Ewa, and Dagmar Divjak. Walter de Gruyter GmbH & Co KG.

Hiraga, Masako K. 2005. *Metaphor and Iconicity: A Cognitive Approach to Analysing Texts*. Palgrave Macmillan.

Jing-Schmidt, Zhuo. 2008. "Much Mouth Much Tongue: Chinese Metonymies and Metaphors of Verbal Behaviour." *Cognitive Linguistics* 19 (2): 241–82.

Johnson, Mark. 1987. *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. University of Chicago Press.

Kövecses, Zoltán. 2005. *Metaphor in Culture: Universality and Variation*. Cambridge University Press.

Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. University of Chicago Press.

Lakoff, George, and Mark Johnson. 1999. *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. New York: Basic Books.

Lakoff, George, and Mark Johnson. 2003. *Metaphors We Live by*. Chicago: University of Chicago Press.

Lin, Yutang. 2000. 吾国与吾民 [Wǔguó Yǔ Wǔmín]. Beijing: Foreign Language Teaching and Research Press.

Lyons, John. 1995. *Linguistic Semantics: An Introduction*. Cambridge University Press.

Matisoff, James. 1991. "The Mother of All Morphemes: Augmentatives and Diminutives in Areal and Universal Perspective." In *The First Annual Meeting of the Southeast Asian Linguistics Society*.

Mercurio, Deo. 'Tales of Ritona', 2016. http://www.citethisforme.com/guides/chicago-note-bibliography/how-to-cite-a-website.

Mischler III, James. 2013. *Metaphor across Time and Conceptual Space: The Interplay of Embodiment and Cultural Models*. John Benjamins Publishing.

Quinn, Naomi. 1991. "The Cultural Basis of Metaphor." In *Beyond Metaphor: The Theory of Tropes in Anthropology*, edited by James W. Fernandez, 56–93. Stanford University Press.

Quinn, Naomi, and Dorothy Holland. 1987. "Culture and Cognition." In *Cultural Models in Language and Thought*. Cambridge University Press.

Seitz, Jay A. 2005. "The Neural, Evolutionary, Developmental, and Bodily Basis of Metaphor." *New Ideas in Psychology* 23 (2): 74–95.

Steen, Gerard B. 2011. Metaphor in Language and Thought: How Do We Map the Field? In *Cognitive Linguistics: Convergence and Expansion* edited by Mario Brdar, Stefan Th. Gries, and Milena Žic Fuchs, 67–86. Amsterdam, Netherlands: Benjamins.

Tarasti, Eero. 2016. "Metaphors, semiotics and futures studies". *Futures* 84: 120–123.

To, Siu-Ming. 2015. "Development and Validation of a Quantitative Measure for Chinese Sources of Parental Meaning". *Journal of Child and Families Studies* 24: 3580–3594.

Turner, Mark. 1987. *Death Is the Mother of Beauty: Mind, Metaphor, Criticism*. University of Chicago Press.

Wee, L. 2006. "The Cultural Basis of Metaphor Revisited." *Pragmatics & Cognition* 14 (1): 111–28.

Wittgenstein, Ludwig. *Philosophical Investigations*. 3rd ed. Translated by G. E. M. (Gertrude Elizabeth Margaret) Anscombe. Oxford UK; Malden, Mass.: Blackwell, 2001.

Yu, Ning. 2007. "The Chinese Conceptualization of the Heart and Its Cultural Context: Implications for Second Language Learning." In *Applied Cultural Linguistics: Implications for Second Language Learning and Intercultural Communication*, ed. by Farzad Sharifian and Gary Palmer, 65–85. Amsterdam, Netherlands: Benjamins.

Yu, Ning. 2009. "Metaphor, Body, and Culture: The Chinese Understanding of Gallbladder and Courage." *Metaphor and Symbol* 18 (1). Lawrence Erlbaum Associates, Inc.: 13–31.

Yu, Ning. 2011. "Beijing Olympics and Beijing Opera: A Multimodal Metaphor in a CCTV Olympics Commercial." *Cognitive Linguistics* 22 (3): 595–628. https://doi.org/10.1515/cogl.2011.023.

Yue, Wenhui. 2012. "义素及义素分析方法浅谈 [Semantic Component and Semantic Component Analysis: A Brief Introduction]." *Journal of Shanxi Normal University(Social Science Edition), Special Issue For Graduate Students* 37: 84–85.

Yule, George. 2014. *The Study of Language*. Cambridge University Press.

The classification of Chinese time expressions from Systemic Functional Linguistics Perspectives

Linda Tsung and Lubei Zhang

The University of Sydney, Australia / Southwest Jiaotong University, China

Time expressions are one of the fundamental concepts of human cognition and communication and thus have been the major concern in many linguistics, applied and developmental psycholinguistic studies (e.g. Klein & Li 2009; Li & Bowerman 1998; Shirai, Slobin, & Weist 1998). A study of time expressions in Modern Chinese (Mandarin) was conducted to explore the common patterns, system networks and realizations from a Systemic Functional Linguistics (SFL) perspective. The results indicated that (1) time expressions can be classified as extent or location; definite or indefinite, (2) within extent a distinction can be made between duration and frequency; within location there are subcategories of Absolute and Relative; of Rest and Motion (3) the forms used to realize these time elements of time are nominal groups, adverbs and pre-verbal phrases, though not all subcategories of temporal elements are realized by all these forms.

Keywords: time expressions, Chinese circumstantial elements, classifications

1. Introduction

As one of the fundamental concepts of human cognition and communication, time expressions have been a major concern in many linguistics, applied and developmental psycholinguistic studies (Klein & Li 2009; Li & Bowerman 1998; Shirai, Slobin, & Weist 1998). It is generally believed that there are six types of devices used to express the circumstances of time in a language, namely, tense, aspect, aktionsart, temporal adverbials, temporal particles, and discourse principles (Klein 2009). Impressive findings have been made in the studies of time expressions during the past decades, however, strong biases have been also found. Of the 90% of the world's languages, the understanding of their time expressions was mostly

based on descriptions, which just tried to find analogues to tense and aspect in these languages (Klein 2009). Most works were done on tense and aktionsart of a few Indoeuropean languages (Comrie 1985; Dahl 2000; Ebert and Zuniga 2001), such as Greek, Latin, English, etc. The existing researches yield a very unbalanced and incomplete picture.

However, in languages as Chinese, time is not marked as tense or aspect. Particular elements are used to locate a situation in time. Many Chinese researchers maintain that in Chinese, time as a circumstantial element may be classified as expressing extent (referring to some period or length of time, analogous to distance in space) or location (referring to a point or points of time) (Yu 2006; Liu 2005; Chen 1982). Chen (1982) once pointed out that Chinese time expressions can be divided into three groups: one expresses extent, like nian 年 (year)/yue 月 (month)/ri 日 (day); one expresses specific point of time, like zhou 昼 (daytime)/ye 夜 (nighttime)/chun 春 (spring)/xia 夏 (summer)/shuo 朔 (the first day of the lunar month)/zuo 昨 (yesterday)；and others express non-specific time, like chu 初 (early)/gu 古 (ancient)/jin 今 (present)/xi 昔 (past). Later, Yu (2006) made a relatively more comprehensive classification, and categorized the time expressions into 7 groups, namely, (1) indefinite time or time extent; (2) year; (3) Season; (4) month; (5) day; (6) period within a day; and (7) hour.

More recently, classifications of Chinese time expressions were made from the perspective of discourse and information processing. Long (2009) classified time expressions into temporal nominal, temporal adverbs and temporal preposition. He maintained that temporal nominal can convey the most abundant message, while temporal adverb and temporal preposition are also important for understanding the temporal information in the sentence. He further classified the temporal nominal into accurate, relative and event-related, temporal adverb into past, current, future, and end, temporal preposition into preposition introducing temporal nominal and preposition convey time sequence by itself. However, with the focus on the time sequence of discourse, its classification method has greater limitations as it didn't consider the quantitative structure of time expressions. Shi's (2010) study on the automatic identification of time expressions categorized the common time expressions into 9 groups. Although it has considered formation rules of common time words, it neglects the combination of time words with other grammatical structures, for example, the combination of time with locations and other quantitative structures.

Considering the merits and demerits of various classification methods and the effectiveness for information processing, the present paper tries to analyze the sub-branches of common time expressions in Chinese based on a structure put forward by Hallliday (1985) from the perspective of Systemic Functional Linguistics (SFL). Within the framework of SFL, time together with other elements such

as place, cause and manner has been classified as circumstantials derived from a relational process. Halliday has identified 9 types of circumstantial elements (Extent, Location, Manner, Cause, Contingency, Accompaniment, Role, Matter, and Angle), among which Extent and Location have formed a four-term set for spatial and temporal expressions. As Halliday (1987) has noted, "there are close parallels between temporal and spatial expressions …", in particular, both may be classified as expressing extent or location; and as being definite or indefinite. Based on this structure, further distinctions of Chinese time expressions are made. Within extent a distinction is made between duration and frequency, while within location subcategories of Absolute and Relative, of Rest and Motion are identified (see Figure 1).
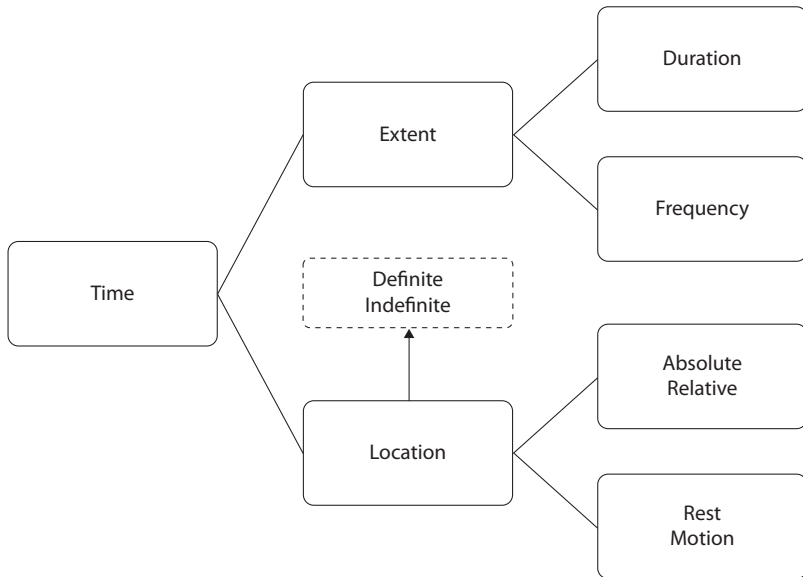

Figure 1. Network: Time expressions

The data for the present study is selected from the following 2 corpora: BJKY (Beijing Kouyu Yuliaoku, corpus of Beijing spoken language, developed by Beijing Language and Culture University) and CCL (modern Chinese corpus developed by Center for Chinese Linguistics PKU). The data for description and interpretation is mainly selected from the standard, colloquial speech of modern Chinese rather than from written sources, since evidence shows that the potential of the language system is more richly developed, and more fully revealed in speech (Halliday 1987). The unconscious nature of spoken language renders it respond continually to the small but subtle changes in its environment, and thus exhibits a rich pattern of semantic and grammatical variation that does not explored in

writing. Hence, it is believed that spoken language is more natural and representative for discussing the circumstances.

It is expected that SFL would provide us with a new perspective to approach time expressions in Chinese.

2. Time expressions: Extent

According to Halliday (1994), time expressions of extent, involving the use of a unit of measurement of time (for example, nian 年 "year"; yue 月 "month"; xiaoshi 小时 "hour"), may be either definite or indefinite. A distinction can also be drawn between duration and frequency.

In the following part, Chinese time expressions are analyzed from two perspectives: definite & indefinite time, duration & frequency time. We are using the term pre-verb in the same way Halliday uses the term prepositive verb[1] "A prepositive verb of this type combines with following N. into a complex group which normally has direct pre-verbal position, occasionally initial position, in the clause".

2.1 Definite time

In Chinese, definite time is usually expressed by nominal groups, adverbs, and pre-verbal phrases,[2] as discussed below.

Nominal groups expressing definite temporal extent often consist of a numerative, a measure word (together with the Head word) as in (1) and (2); or use a modifier followed by a time word such as shihou 时候 "time", nianjian 年间 "during the years …", as in (3).

---

1. According to Halliday (2014), "verbal words in Chinese can be classified into 4 categories: (i) lexical verbs (including adjectives, which in Chinese are a kind of verb, not a kind of noun as in English); (ii) modal auxiliary verbs, which construe the likelihood and the desirability of the process; (iii) postpositive verbs, which show phases of the process, and (iv) prepositive verbs, which construe the relation between the process and some particular entity. (For comparison with English, prepositive verbs are similar to prepositions and postpositive verbs are similar to post-verbal adverbs.)"

2. As a pre-verb cannot realize a circumstantial element by itself, only when it precedes a nominal group or nominal groups as a pre-verb + nominal group(s) phrase can it express a circumstantial element. For such a phrase we will use the term pre-verbal phrase.

(1) San ge yue mei xia yu le.
三个月没下雨了。
three M. month not drop rain P.
It has not rained for three months.

(2) Wo zai Shanghai zhu le liangnian.
我在上海住了两年。
I at Shanghai live P. two years
I lived in Shanghai for two years.

(3) Zhe ge huaping shi qianlong nianjian zuo de.
这个花瓶是乾隆年间做的。
this M. flower-vase is Qianlong years make P.
The vase was made during the years of Qianlong.

In (1) <u>san ge yue</u> 三个月 "three months" is a participant as a circumstantial element. In (2) <u>liangnian</u> 两年 "two years", the Head directly follows a numerative without a measure word in between, because the Head <u>nian</u> 年 itself denotes measurement, in the sense that some words like <u>nian</u> 年 "year" and <u>tian</u> 天 "day" can function as both nouns and measures words.In (3) <u>nianjian</u> 年间 "the years" denotes the period of time as a circumstantial element.

The interrogative forms for temporal extent are <u>ji</u> 几, in <u>ji nian</u> 几年 "how many years", <u>ji tian</u> 几天 "how many days", <u>ji ge yue</u> 几个月 "how many months", <u>ji ge xingqi</u> 几个星期 "how many weeks", <u>ji ge xiaoshi / zhongtou</u> 几个小时／钟头 "how many hours"; and <u>duo</u> 多 in <u>duojiu</u> 多久 "how long", <u>duochang shijian</u> 多长时间 "how long a time", <u>duoshao ci</u> 多少次 "how many times", etc.

Adverbs as realizations of definite temporal extent are <u>yizhi</u> 一直 "all along", <u>yongyuan</u> 永远 "for ever", <u>shizhong</u> 始终 "always", <u>zhanshi</u> 暂时 "for the time being", <u>conglai</u> 从来 "at all times", etc., as in (4) and (5).

(4) Ta yizhi zai bangzhu wo.
她一直在帮助我。
she all long in help me
She has been helping me all along.

(5) Wo conglai mei tingshuo guo zhe shi.
我从来没听说过这事。
I never not hear P. this thing
I never heard about that.

In (5) <u>conglai</u> 从来 "at all times" is the same as <u>lilai</u> 历来 and <u>xianglai</u> 向来 in meaning, but <u>conglai</u> 从来 is often used in negative sentences and <u>xianglai</u> 向来 is

mainly used in affirmative sentences as in (6), <u>lilai</u> 历来 never appears in negative sentences as in (7).

(6) Xiaohua xianglai hen laoshi.
小华向来很老实。
Xiaohua at all times very honest.
Xiaohua is very honest at all times.

(7) Women lilai tichang ziligengsheng.
我们历来提倡自力更生。
we at all time promote self-reliance
We promote self-reliance all the time.

Pre-verbal phrases can be used as circumstantial elements of definite temporal extent as well. As Chao (1968) commented that a time word can fill the following positions:

zai 在…	"to be at …"
dao 到 …	"reach (the time of)"
dengdao 等到 …	"by …" or "by the time when …"
cong … -qi 从 起	"start from …"

Besides , the frames <u>dao … weizhi</u> 到…为止 "reach … as stopping place" /"as far as" and <u>cong … qi</u> 从…起 "start from …" can be filled by both place and time words, as well as words for quantity.

To express circumstances of temporal extent, pre-verbal phrases <u>zai … zhong</u> 在…中 "during … (the time of)" and <u>dao le</u> …到了" reach … (the time of)" are the most commonly used ones. For example:

(8) Zai liang ge xiaoshi zhong, ta wen le ba ge wenti.
在两个小时中，他问了八个问题。
during two M. hour among, he asked P. eight M. questions.
He asked eight questions during the two hours.

(9) Dao le yi kezhong, xiyiji hui zidong ting xialai.
到了一刻钟，洗衣机会自动停下来。
reach P. one quarter washing machine can automatically stop down
The washing machine will automatically stop after running for a quarter of an hour.

In Chinese, the time: <u>xiaoshi</u> 小时 "hour", <u>kezhong</u> 刻钟 "quarter", <u>fenzhong</u> 分钟 "minute" are usually employed to express circumstances of temporal extent.

In colloquial Chinese, we also use <u>zhongtou/zhongdian</u> 钟头／钟点 to substitute for <u>xiaoshi</u> 小时, but they must be used together with the measure word <u>ge</u>个, while <u>ge</u> 个 is optional in the nominal group when <u>xiaoshi</u> 小时 is used, for example:

> Liang (ge) xiaoshi 两个小时
> "two hours"
> Liang ge zhongtou/zhongdian 两个钟头／钟点
> "two hours"

From the examples above, we can find that in Chinese circumstantial elements of definite time extent can be realized by nominal groups, adverbs, and pre-verbal phrases.

2.2 Indefinite time

Indefinite temporal extent are also usually realised, in Chinese, by nominal groups, adverbs and pre-verbal phrases.

Indefinite temporal extent realized with a nominal group contains <u>ji</u> 几 "a few or a little" plus the Head: <u>jitian</u> 几天 "a few days", <u>jinian</u> 几年 "a few years", as in (10). <u>Ji</u> 几 with <u>yue</u> 月"month", <u>xingqi</u> 星期 "week", and <u>xiaoshi</u> 小时 "hour" or <u>zhongtou</u> 钟头 "hour" require measure words in between, for example, <u>ji ge yue 几个月</u> "a few months", <u>ji ge xiaoshi / zhongtou</u> 几个小时／钟头 "a few hours", as in (11).

(10) Wo ji nian qian jian guo ta.
      我几年前见过他。
      I few year ago see P. him
      I saw him a few years ago.

(11) Wo deng le ta ji ge xiaoshi.
      我等了他几个小时。
      I wait P. him few M. hour
      I had waited for him a few hours

As mentioned in Section 2.1 above, ji 几 can be used in a question, as an interrogative, meaning "how many". In Examples (10) and (11), however, it is used as a declarative form to express indefinite temporal extent. There is another word <u>duo</u> 多 "many" that is used with <u>nian</u>年 "year", but <u>duotian</u> 多天 "many days" is not acceptable in Chinese. <u>Xuduo</u> 许多 "many" can play this part, <u>xuduo tian 许多天</u> "many days".

(12)   Duonian lai ta yizhi yanjiu peitai yizhi de jishu.
多年来他一直研究胚胎移植的技术。
many years P. he all long research embryo transfer P. technology
He has been researching the embryo transfer technology for many years.

Lai来, a particle in (12), is often used with time words mentioned above to express indefinite temporal extent, such as ji ge yue lai 几个月来 "about a few months" and xuduo tian la i许多天来 "about many days", etc.

Adverbs such as henjiu 很久 or haojiu 好久 "long before" as in (13) , bujiu 不久 "before long", changchang 常常 "often", wangwang 往往 "usually", jinlai 近来 "recently", youshi 有时 "sometimes", ou'er 偶尔 "occasionally" as in (14) , are also used as circumstantial elements to express indefinite temporal extent. For example:

(13)   Ta diaozou henjiu le.
他调走很久了。
he transfer long before P.
He was transferred long before.

(14)   Ta ou'er qu kai yi ci hui.
他偶尔去开一次会。
he occasionally go attend one P. meeting
He attended meetings only occasionally.

From the analysis above, we can find that circumstantial elements of indefinite time extent are usually realized by nominal groups, adverbs and pre-verbal phrases as well.

2.3   Duration time

From another perspective, time expressions of extent can also be divided into duration time and frequence time. The temporal duration expresses a period of time. They may be either definite or indefinite. Three forms – nominal groups, adverbs and pre-verbal phrases – can be used as circumstantial elements of temporal duration, as in Examples (15), (16), and (17).

(15)   Ta zou le san tian le.
他走了三天了。
he go P. three day P.
He has been away for three days.

(16) Lao Wang jie yan duonian le.
老王戒烟多年了。
Old Wang stop smoking many years P.
Old Wang gave up smoking many years ago.

(17) Cong xinzhongguo chengli yilai renmin de shenghuo shuiping zai tigao.
从新中国成立以来人民的生活生活水平在提高。
since new China found P. people's living level at increase
Since the founding of new China the people's standard of living has been going up.

In (15), a nominal group is used to express definite duration; in (16) indefinite duration is realised by the nominal group <u>duonian</u> 多年 "many years"; and in (17) <u>cong xinzhongguo chengli iy lai</u> 从新中国成立以来 "since the founding of new China", a pre-verbal phrase is employed to show a long, definite period of time.

Duration can also be realised by adverbs such as <u>henjiu</u> 很久 "long before" and <u>conglai</u> 从来 "at all times, since". The examples in 2.1 and 2.2 illustrate adverbs expressing definite and indefinite duration.

2.4 Frequency time

The temporal frequency expresses how many times or how often the process takes place. They can be realized by nominal groups which express the multiplicity, i.e., the number of times, this involves the numerative plus a verbal measure word , such as <u>liang ci</u> 两次 "twice", <u>si hui</u> 四回 "four times", <u>ji ci</u> 几次 "a few times" and <u>duo ci</u> 多次 "many times".

For example:

(18) Wo quguo Xianggang liang ci.
我去过香港两次。
I go P. Hong Kong twice
I have been to Hong Kong twice.

(19) Wo duo ci qingjiao ta zhe ge wenti.
我多次请教他这个问题。
I many times consult he this M. problem
I have consulted him on this problem many times.

In (18) <u>liang ci</u> 两次 "twice" expresses definite multiplicity; while in (19), <u>duo ci</u> 多次 "many times" indicates indefinite frequency (multiplicity) as circumstances. <u>Ci</u> 次 is a verbal measure word, which follows the numerative to express the circumstance of temporal frequency. <u>Hui</u> 回 is also used in this way. These two verbal measure words are mostly interchangeable.

Circumstances of temporal frequency can also be realized by adverbs to express how often the process takes place. They have three degrees of value: high, medium and low. High frequency can be expressed by <u>yongyuan</u> 永远 "forever", <u>yizhi / shizhong</u> 一直／始终 "all along", <u>yixiang</u> 一向 "consistently", <u>zongshi</u> 总是"always", etc., as in (20) . Medium frequency is typically expressed by <u>jingchang</u> 经常 "often", <u>changchang</u> 常常 "often", <u>shichang</u> 时常 "often", <u>tongchang</u> 通常 "usually". <u>Wangwang</u> 往往 "frequently", <u>wangchang</u> 往常 "habitually in the past / used to", as in (21). Low frequency can be realized by <u>youshi</u> 有时 "sometimes", <u>ou'er</u> 偶尔 "occasionally", <u>shi'er</u> 时而 "sometimes", as in (22) below.

(20)   Dayu yizhi xia le liangtian.
　　　 大雨一直下了两天。
　　　 big rain all long drop P. two day
　　　 It has been raining heavily for two days without let up.

(21)   Wo tongchang liu dian qichuang.
　　　 我通常六点起床。
　　　 I usually six o'clock get up
　　　 I usually get up at six o'clock.

(22)   Ta youshi ye xie ji ju shi.
　　　 他有时也写几首诗。
　　　 he sometimes also write a few P. poem
　　　 Sometimes he also writes a few lines of poetry.

The three examples above show the three different degrees of frequency as circumstances: high in (20), medium in (21), and low in (22). Table 1 summarises the adverbs used as temporal frequency in Chinese.

Table 1. Degrees of frequency as time elements

Frequency	High:	Yongyuan 永远 "forever", shizong 始终 "all along", zongshi 总是 "always", yizhi 一直 "all along", yixiang 一向 "consistently", conglai 从来 "never"
	Medium:	jingchang 经常 "often", wangwang 往往 "frequently", tongchang 通常 "usually", wangchang 往常 "habitually"
	Low:	Youshi 有时 "sometimes", shier 时而 "from time to time", ou er 偶尔 "occasionally", cengjing 曾经 "ever"

As we discussed above, time expressions of extent can be either definite or indefinite, expressing a time duration or frequency. They are realized mostly by nominal groups or adverbs.

3. Time expression: Location

Temporal location, some point in time, is similar to temporal extent in that it may be definite or indefinite. However, several other dichotomies, not relevant to temporal extent, may also be elucidated, namely, absolute and relative – and within relative, near and remote; and rest and motion. Within motion there is a further dichotomy – towards and away from (see Figure 2).
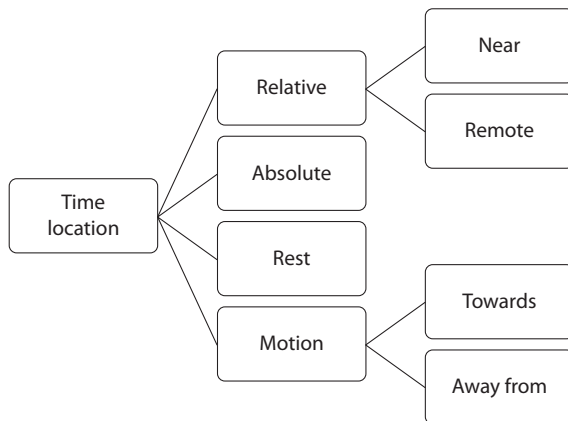

Figure 2. Time expressions: Location

The forms in Chinese used to realize temporal location are the same as those used for temporal extent, nominal groups, adverbs, and pre-verbal phrases. In the following part, we are going to analyze the temporal location from three perspectives, namely, definite and indefinite temporal location, absolute and relative temporal location, rest and and motion temporal location.

3.1 Definite temporal location

Definite temporal location, such as the times of years, months, days, weeks, seasons, hours, festivals, and so on, is realized by nominal groups, adverbs, and pre-verbal phrases. For example, <u>1985 nian 1985 年</u> "1985", <u>wuyue 五月</u> "May", <u>shi'er hao 12号</u> "the 12th", <u>xingqisi 星期四</u> "Thursday", <u>chuntian 春天</u> "Spring", <u>zai sandian 在三点</u> "at three o'clock", <u>zhongqiujie 中秋节</u> "the mid-Autumn Festival". Apart from these, which are common to many languages, there are some special time words probably unique to Chinese: <u>tian gan 天干</u> "heaven's stems" (ten), <u>dizhi 地支</u> "earth's branches" (twelve), and <u>ershisi jieqi 24 节气</u> "the beginning dates of the twenty-four subdivisions of the solar year". All of these are commonly used as circumstantial elements of definite temporal location, as illustrated in the following examples:

(23)   Jintian (shi) xingqi liu.
今天是星期六。
today (is) Saturday
Today is Saturday.

(24)   Beijing ren zai Dongzhi zhe tian chi jiaozi.
北京人在冬至这天吃饺子。
Beijing people at Winter Solstice this day eat dumpling
Beijingese eat dumplings on the day of the Winter Solstice.

(25)   Lu Xun sheng yu 1881 nian.
鲁迅生于 1881 年。
Lu Xun born 1881 year
Lu Xun was born in 1881.

It can be seen that in (23) both <u>jintian</u> 今天 "today" and <u>xingqiliu</u> 星期六 "Saturday" are temporal circumstances. <u>jintian</u> 今天 "today" is a participant. As a circumstantial element: <u>jintian</u> 今天 and <u>xingqiliu</u> 星期六 can be reversed without changing the meaning: <u>Xingqiliu shi jintian</u> 星期六是今天 "Saturday is today".

The pre-verbal phrase <u>zai dongzhi zhe tian</u> 在冬至这天 "on the day of the Winter Solstice" is used in (24). <u>Dongzhi</u> 冬至 is from the twenty-four subdivisions of the solar year, and is frequently used in everyday speech in China to refer to time. In (25), <u>yu</u>于 "in" can be replaced by the pre-verb <u>zai</u>在 "in or at", but <u>yu</u> 于is from classical Chinese, it seems more literary than <u>zai</u>在.

As Chinese does not have a tense system, the use of time words as circumstantial elements is more important than it is in English. In Chinese, there are quite a lot of words used to express the circumstances of definite temporal location . The following is a list of those commonly used.

1.   Past time
qiannian 前年 "the year before last"
qiantian 前天 "the day before yesterday"
qunian 去年 "last year"
zuotian 昨天 "yesterday"
shangge yue 上个月 "last month"
shang ge xingqi 上个星期 "last week"
. …tian/nian/dian yiqian 天／年／点 以前 "day/year/o'clock ago"
…ge yue/xinggi yiqian 个月／星期 以前 "…month/ week ago"
2.   Present time
Jintian 今天 "today"
Jinnian 今年 "this year"
xianzai 现在 "now"

zhe ge yue 这个月 "this month"
zhe ge xingqi 这个星期 "this week"
muqian 目前 "at present"
yanqian 眼前 "at present"

3. Future time
mingtian 明天 "tomorrow"
mingnian 明年 "next year"
xia ge yue 下个月 "next month"
xia ge xingqi 下个星期 "next week"

The interrogative forms for temporal location are to answer the question "What time?" "When?" In Chinese <u>shenme shihou</u> 什么时候 "what time or when" is the most common form of this kind. Also used are <u>jishi</u> 几时and <u>duohuir</u> 多会儿 "when or what time", for example:

(26)   Ni duohuir lai Xini de?
你多会儿来悉尼的？
you when come Sydney P.?
When did you come to Sydney?

In this example we can also use <u>jishi</u> 几时 to replace <u>duohuir</u> 多会儿. <u>Duohuir</u> 多会儿 is commonly used in speech, but <u>jishi</u> 几时 is mostly used in writing, such as in poems.

However, in some cases, <u>Duohuir</u> 多会儿 and <u>jishi</u> 几时 can also be used to express indefinite temporal location. For example:

(27)   Duohuir ni you kong jiu lai ba.
多会儿你有空就来吧。
whenever you have free time then come P.
Please drop in whenever you are free.

As has been discussed above, a lot of expressions in Chinese can be found to express the circumstances of definite temporal location: nominal groups, adverbs, and pre-verbal phrases.

3.2   Indefinite temporal location

Indefinite temporal location is typically realized by adverbs, such as <u>gang cai</u> 刚才 "just now", <u>yi jing</u> 已经 "already", <u>zui jin</u> 最近 "recently", <u>cengjing</u> 曾经 "ever" etc., for example:

(28)   Ta gangcai zai tushuguan kanshu.
他刚才在图书馆看书。
he just now at Library read book
He read a book in the Library just now.

(29)   Wo zuijin jian guo ta.
我最近见过他。
I recently meet P. him
I saw him recently.

Nominal groups expressing indefinite temporal location, often use time words plus <u>qianhou 前后</u> "before or after／around", <u>zuoyou 左右</u> "left or right /about" as in <u>liang dianzhong zuoyou</u> 两点钟左右 "about two o'clock", <u>guoqingjie qianhou 国庆节前后</u> "before or after National Day". For example:

(30)   Tamen Guoqing jie qianhou jiehun.
他们国庆节前后结婚。
they National Day before after marry
They will marry sometime around the National Day.

Pre-verbal phrases can use <u>zai 在</u> plus the time words and <u>qianhou 前后</u> or <u>zuoyou 左右</u> "at about that time", <u>yue zai 约在</u> plus time words, or <u>yue yu 约于</u> plus time words "at about that time".

(31)   Qimo kaoshi yue yu xiazhou jinxing.
期末考试约于下周进行。
end term exam about at next week hold
The end of term exam will be held about next week.

Yue yu 约于 "at about …" can be replaced by yue zai 约在 "at about …" in this example, but it is more formal.

As is seen above, besides using adverbs as a typical type of indefinite temporal location, nominal groups and pre-verbal phrases plus <u>qianhou 前后</u>, <u>zuoyou 左右</u>, or yue 约 can also be used to express indefinite temporal location.

3.3   Absolute temporal location

Absolute temporal location as a circumstance expresses a unique time reference, not dependent on the moment of speaking. In a broad sense, absolute temporal location is quite similar to definite temporal location because a circumstantial element of absolute temporal location must also express definite temporal location, for example <u>1986 nian 1986年</u> "1986". However, a circumstantial element of definite temporal location does not necessarily express absolute temporal location. For

example, xingqiliu 星期六 "Saturday" can be any Saturday in a month, or in a year. Xingqiliu 星期六 can express absolute temporal location only when it is used in a nominal group which indicates the particular xingqiliu 星期六 "Saturday", such as 1952 nian de di er ge xingqiliu 1952 年的第二个星期六 "the second Saturday in 1952", or any other unique Saturday.

To distinguish between absolute and definite temporal location, note that the former is a unique time, which may include time of day, day of week and day of month, as well as year; while the latter is not. For example, the nominal group 1985 nian wuyue si hao xingqiliu si dian 1985 年 5 月 4 号星期六 4 点 "4 o'clock on Saturday, May 4, 1985", each of the constituent nominal groups expresses absolute temporal location; but without 1985 nian "in 1985" (which establishes a unique time, in years), the other constituent nominal groups express only definite temporal location (as they do also when used alone). The following four examples illustrate this:

(32) Ta shi 1958 nian 3 yue 8 hao sheng de.
      她是 1958 年 3 月 8 号生的。
      she is 1958 year March 8 date born P.
      She was born on 8 March 1958.

(33) Ta shi 1958 nian sheng de.
      她是 1958 年生的。
      she is 1958 year born P.
      She was born in 1958.

(34) Ta shi san yue ba hao sheng de.
      她是 3 月 8 号生的。
      she is March 8 date born P.
      She was born on the 8th of March.

(35) Ta shi ba hao sheng de.
      她是 8 号生的。
      she is 8 date born P.
      She was born on the 8th.

It is clear that in (32) and (33) both 1958 nian san yue ba hao 1958 年 3 月 8 号 "8th March, 1958" and 1958 nian 1958年 "1958" are circumstantial elements of absolute temporal location (and they obviously express definite temporal location as well); while san yue ba hao 3 月 8 号 "8th March" in (34) and ba hao 8 号 "the 8th" in (35) are circumstantial elements of definite temporal location, but they do not convey absolute temporal location.

In the examples above, absolute temporal location is realized by nominal groups. However, it can also be expressed by pre-verbal phrases. These pre-verbal

phrases consist of nominal groups, which, by themselves would express absolute temporal location, preceded by a pre-verbs such as <u>zai</u> 在 "on／at/in" , or optionally, succeeded by <u>de shihou</u> 的时候 "of time". For example:

(36) Zai 1960 nian de shihou, women dou chi bu bao.
      在 1960 年的时候，我们都吃不饱。
      In 1960 year P. time we all eat not enough
      None of us had enough to eat in 1960.

In (36), <u>zai 1960 nian de shihou</u> 在 1960 年的时候 can also be <u>zai 1960 nian</u> 在 1960年 "in 1960". <u>de shihou</u> 的时候 is often used in the spoken form.

3.4 Relative temporal location

Relative temporal location as circumstances expresses a deictic time reference, relative to time "now". It may be either near or remote. Relative temporal location is also typically expressed by adverbs, as in the following examples:

(37) Ta zuijin zai Deguo.
      他最近在德国。
      he recently at Germany
      He was in Germany recently.

(38) Ta congqian shi chushi.
      他从前是厨师。
      he before is cook.
      He was a cook before.

In English, time can be described by the verb tense. However, there is no tense system in the Chinese language, it uses the circumstantial elements of relative time to show the time either near as in (37), or remote as in (38). Other adverbs which express relative temporal location are <u>gang cai</u> 刚才 "just now", <u>muqian</u> 目前 "at present", etc., as near relative location; <u>ranhou</u> 然后 "then", <u>jinhou</u> 今后 "from now on", <u>yiqian</u> 以前 "before", <u>yihou</u> 以后 "later", etc., as remote relative location. Relative time is also indefinite time in temporal location.

3.5 Rest temporal location

Temporal elements of rest express a static time when the process takes place. Pre-verbal phrases with the pre-verb <u>zai</u> 在 "at, in or on" and <u>dang</u> 当 "when" express temporal rest, such as in <u>zai xingqiliu</u> 在星期六 "on Saturday", <u>zai qi dian</u> 在 7 点 "at seven o'clock", <u>zai … de shiqi</u> 在…的时期 "during … period", <u>zai / dang … de shihou</u> 在／当…的时候 "when …". For examples:

(39)   Tamen zai xingqiliu juxing hunli.
他们在星期六举行婚礼。
they at Saturday hold wedding.
They will hold their wedding on Saturday.

(40)   Zai Wenhua Da Geming de shihou wo hai shi ge xuesheng.
在文化大革命的时候我还是个学生。
at culture big revolution P. time I still is M. student
During the Cultural Revolution I was still a student.

In the examples above the pre-verbs can be omitted without changing the meaning, though a circumstantial element of rest temporal location is more emphatic when the pre-verb is used. Both <u>dang</u> 当 and <u>zai</u> 在 can be used to express temporal location of Rest, however, <u>dang</u>当 can only be used in a clause, while <u>zai</u> 在 can be used in both clauses and pre-verbal phrases, for example:

(41)   Dang wo huilai de shihou, ta yijing zou le.
当我回来的时候，他已经走了。
when I come back P. time he already leave P.
When I came back, he had already left.

(42)   Zai 1978 nian, wo shang le da xue.
在 1978 年我上了大学。
At 1978 year I enter P. university
In 1978 I entered university.

(*42)   Dang 1978 nian wo shang le daxue.
当 1978 年我上了大学。
When 1978 I entered the university.
I entered in the University in 1978.

In (41) either <u>dang</u> 当 or <u>zai</u> 在 can be used in the clause of temporal location, but in (42) <u>dang</u> 当 cannot replace <u>zai</u> 在, because <u>1978 nian 1978年</u> is a nominal group, therefore (*42) is an ungrammatical sentence in Chinese.

3.6   Motion temporal location

Temporal motion indicates the kinetic time when the process takes place. They may be either motion towards or motion away, and is typically realized by preverbal phrases. These pre-verbal phrases, unlike those which realize temporal rest, have pre-verbs <u>dao</u>到 … "to", <u>zhidao</u> 直到 "till", <u>zai / dang yiqian</u> 在／当…以前 "before…", etc., which express motion towards and <u>cong</u> 从 "from", <u>zicong</u> 自从 "since", <u>da … shihou qi</u> 打…时候起 "from that time …" , <u>zai / dang … yihou</u> 在／

当…以后 "after …" and <u>zi自</u> "from", etc., which express motion away from. It is illustrated in the examples given below.

(43) Zhidao qunian ta cai zhaodao ta de nüer.
      直到去年他才找到他的女儿。
      until last year he yet find he P. daughter
      He didn't find his daughter until last year.

(44) Cong qunian ta jiu shi laoshi le.
      从去年他就是老师了。
      since last year she yet is teacher P.
      She has been a teacher since last year.

In (43) <u>zhidao 直到</u> "until" expresses motion towards <u>qunian 去年</u> "last year", i.e. <u>qunian 去年</u> "last year" is the end point of a period of time which began before then; in (44) <u>cong从</u> "since" indicates motion away from <u>qunian 去年</u> "last year",i.e., <u>qunian 去年</u> "last year" is the starting point of a period of time that stretches forwards away from then.

4.   Time expressions and classfications

In this paper we examined the circumstantial elements of time in Chinese from different perspectives. Figure 3 below summarises the relationship between the (parent) category, Time, and its "child" subcategories, as well as between the (parent) subcategories Extent and Location and their "child" subcategories

It can be seen that this categorization scheme has four levels. The categories on each level (except the highest) form pairs, as shown in Figure 4.

From the figure above, we can find that each of the two pairs in level 4 is only related to one element of a set in level 3, i.e. Near & Remote is only related to Relative and Towards & Away from is only related to Motion. This makes it possible that the level 4 categories can be absorbed in the level 3, forming a doublet (two member set) and two triplets (three member set), as follows:

Level 3. (Duration, Frequency)
        (Absolute, near, remote)
        (Rest, towards, away from).

Since the distinctions of subcategories at the same level are made from different perspectives, we can find there are crossings between them. To clarify the relationships between different subcategories of intra- and inter-level, it is helpful to consider them as matrices, the dimensionality of the matrix being the number of sets
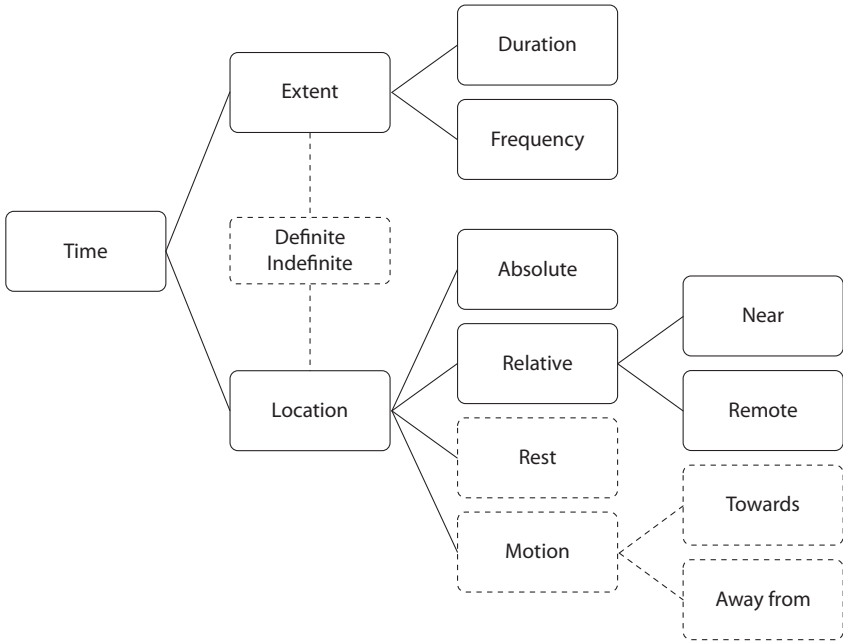
Figure 3. Classification time scheme
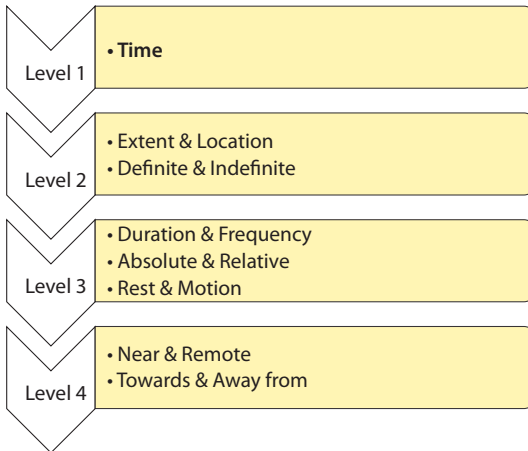
Figure 4. Levels of classification

in a level, the number of rows (columns) being the number of elements in each set. For example, the level 2 matrix is a 2 × 2 matrix. Realizations of the four possible classes are given below (see Table 2).

Table 2. Subcategories in level 2

	Extent of time	Location of time
Definite	san tian 三天 "three days"	san hao 三号 "3rd (date)"
Indefinite	ji tian 几天 "a few days"	yi hou 以后 "later"

The level 3 matrix would appear to have 18 distinct elements ($2 \times 3 \times 3$), but the set (Duration, Frequency) is unrelated to the sets (Absolute, near, remote) and (Rest, towards, away from) as the former is a subcategory of Extent of time only, and the latter of Location of time only. Thus there are two distinct level 3 matrices, one is 2×1 (see Table 3) and one is 3×3 (see Table 4).

Table 3. Subcategories of time extent

Duration	Frequency
si tian 四天 "four days"	si ci 四次 "four times"

Table 4. Subcategories of time of location

	Absolute	Near	Remote
Rest	zai 1985 nian 在 1985 年 "in 1985"	zai zhe shihou 在这时候 "at this time"	zai na shihou 在那时候 "at that time"
Toward	dao 1985 nian 到 1985 年 "till 1985"	dao zhe shihou 到这时候 "till this time"	dao na shihou 到那时候 "till that time"
Away from	cong 1985 nian 从 1985 年 "from 1985"	cong zhe shihou 从这时候 "from this time"	cong na shihou 从那时候 "from that time"

From the tables above, we can see in level 2 there are 4 subcategories of time expresions, while in level 3 there are 11 subcategories, 2 of time extent and 9 of time location. However, to complete the discussion of the relationships between subcategories, it is necessary to examine the relationships between sets at different levels. From the discussion above, we can see that the set (Extent, Location) in level 2

Table 5. Relationships between level 2 and level 3

	Definite	Indefinite
Absolute	1985 nian 1985 年 "in 1985"	
Near	xianzai 现在 "now"	zuijin 最近 "recently"
Remote	na shihou 那时候 "that time"	yihou 以后 "later"
Rest	zai xingqi si 在星期四 "on Thursday"	zai yihou 在以后 "(at) later"
Towards	dao xingqi si 到星期四 "till Thursday"	dao yihou 到以后 "(till) later"
Away from	cong xingqi si 从星期四 "from Thursday"	cong na yihou 从那以后 "since then"

is related to the sets in level 3 in a simple way, i.e. Extent related to Duration & Frequency and Location related to Absolute & near & remote and Rest & towards & away from. This leaves (Definite, Indefinite) to form a two 2 × 3 matrix, as presented in Table 5.

Using the matrix above, we can identify 11 subcategories of temporal expressions between level 2 and level 3, since there are no temporal elements expressing indefinite absolute time in Chinese as in any other languages.

As we discussed above, circumstantials of time in Chinese are usually expressed by adverbs, nominal groups, or pre-verbal phrases, Table 6 summarized the typical examples of each kind.

Table 6. Realizations of temporal elements in Chinese

			Adverbs	Nominal groups	Pre-verbal phrases
Time	Extent	Definite	Yizhi 一直 "all long"	san tian 三天 "three days"	zai san tian 在三天 "in three days"
		Indefinite	henjiu 很久 "long before"	ji tian 几天 "a few days"	zai ji than 在几天 "in a few days"
		Duration	Henjiu 很久 "long before"	san ge yue 三个月 "three months"	zai san ge yuezhong 在三个月中 "in three months"
		Frequency	Jingchang 经常 "often"	san ci 三次 "three times"	
	Location	Definite	Muqian 目前 "now"	san dian 三点 "three o'clock"	zai san dian 在三点 "at three o'clock"
		Indefinite	Gangcai 刚才 "just now"	Yihou 以后 "later"	<u>zai…qianhou</u> 在…前后 "at about…"
		Absolute		1985 nian 1985 年 "1985"	zai 1985 nian 在 1985 年 "in 1985"
		Relative	Zuijin 最近 "recently"	zhe shihou 这时候 "at this time"	zai the shihou 在这时候 "at this time"
		Rest			zai xingqi yi 在下星期一 "on Monday"
		Motion		Yihou 以后 "later"	dao xingqi yi 到星期一 "till Monday"

5.   Conclusion

As a fundamental concept of human cognition and communication, time expressions have long become the major concern of many studies in different languages. Although previous studies on the classification of Chinese temporal expressions have yielded some important results and have provided insights into the construction of Chinese temporal expressions, the limitations are also obvious. None of them can provide a comprehensive categorization. Adopting a sturucture proposed by Halliday, this paper have analyzed the temporal expressions in Chinese from the perspective of SFL. Considering the special characteristics of Chinese time expressions, further distinctions have been made under the structure of time extent and time location put forward by previous researchers, which have made the categories clearer and more organized.The results indicated that (1) Chinese time expressions can be classified as extent or location; definite or indefinite, (2) within extent a distinction can be further made between duration and frequency; within location there are subcategories of Absolute and Relative; of Rest and Motion (3) the forms used to realize these time elements are mainly nominal groups, adverbs and pre-verbal phrases, though not all subcategories of temporal elements are realized by all these forms. A network of Chinese temporal expressions from the perspective of Systemic Functional Linguistics (SFL) has been established. By identifying the subcategories of Chinese temporal expressions and their relationships, the present study gives us a more comprehensive classification of Chinese temporal expressions. It cannot only help us clarify the relationships among the complex Chinese time expressions, but also fill in the research gap by offering a new perspective to approach time expressions in Chinese. It is hoped that this study could provide us an enlightenment on studies of other circumstancial elements in Chinese.

References

Chao, Y. R. 1968. *A Grammar of Spoken Chinese*. LA: University of California Press.

Chen, C. Z. 1982. *Guowenfa Caochuang* [Pioneering Study of Chinese Grammar]. Beijing : Commercial Press.

Comrie, B. 1985. *Tense*. Cambridge: Cambridge University Press.

Dahl, Ö. (ed.) 2000. *Tense and aspect in the languages of Europe*. Berlin/New York: Mouton de Gruyter.

Ebert, K. and Zuniga, F. (eds.) 2001. *Aktionsart and Aspectotemporality in Non-european languages*. Zürich: ASAS.

Halliday, M. A. K. 1985. *Language, context, and text: Aspects of language in a social-semiotic perspective*. Geelong: Deakin University.

Halliday, M. A. K. (1987) Spoken and written modes of meaning. In Rosalind Horowitz & S. Jay Samuels (eds.) *Comprehending Oral and Written Language*, pp. 55–87. San Diego, CA: Academic Press.

Halliday, M. A. K. 1994. *An introduction to functional gramma (2nd ed.)*. London: Edward Arnold.

Halliday, M. A. K. (2014) *Halliday's Introduction to Functional Grammar (4th Edition)*. London: Routledge.

Klein, W. 2009. How time is encoded. In W. Klein & P. Li (Eds.), *The expression of time* (pp. 39–82). Berlin/New York: Mouton de Gruyter.

Klein, W., & Li, P. 2009. *The Expression of Time*. Berlin: Mouton de Gruyter.

Li, P., & Bowerman, M. 1998. The acquisition of lexical and grammatical aspect in Chinese. *First Language*, 18(54), 311–350.

Liu, Y. R. 2005. Xiandai Hanyu Shijianci de Yuyi Fenxi [Analysis of Semantics of Temporal Expressions in Modern Chinese]. *Jiamusi University Socail Science Journal*, 4: 66–68.

Long, C. Y. 2009. Minxiang pianzhang de shijian xulie he jianmo jiqi zai chanpin sheji zhong de yingyong. [Application of time series and modeling of time expressions in the text]. Thesis published by Xidian University.

Shirai, Y., Slobin, D. I., & Weist, R. E. 1998. Introduction: the acquisition of tense-aspect morphology.

Shi, H. F. 2010. Jiyu CRF de Zhongwen Mingming Shiti Shibie Yanjiu [Entity Identification Study of Chinese Names Based on CRF]. Thesis published by Suzhou University.

Yu, D. T. 2006. *Xiandai Hanyu Shijianci Yanjiu*. [Study of Temporal Expressions in Modern Chinese]. Wuhan: Huazhong Normal University Press.

Being a Kam in China

Ethnic identity in narratives

Wei Wang,[1] Lisong Jiang[2] and Meishu He[2]
[1]The University of Sydney, Australia / [2]Southwest University, China

In light of the growing interest in investigating the ethnic minority Kam people in China, this paper offers a sociolinguistic analysis to explore how Kam people's identity is represented and negotiated in spoken narratives with outside researchers. Drawing on sociolinguistic approaches to identity analysis (Bucholtz and Hall 2005; De Fina et al. 2006; Blommaert 2005) and membership categorisation analysis (Sacks 1972a & b, 1992; Baker 2004; Fitzgerald and Housley 2015), this paper explores the relationships between Kam people's sense of membership in their ethnic community and social practices that define this sense of membership. It focuses on the self-representation of a former Kam village head in a remote village in Southern China, Guizhou Province, and explores his way of conceptualizing being a Kam with a view to examining the relationship between his representation of the ethnic identity and the sociocultural impacts on this identity construction process.

Keywords: the Kam people, ethnic identity, narrative, sociolinguistics, membership categorisation

1. Introduction

Against a backdrop of increased interest in investigating the ethnic minority people in China, the Kam people, famous for the Kam Grand Choirs ("侗族大歌" in Chinese), have drawn the attention of scholars from a range of fields including sociology, anthropology, culture and music studies (e.g. Geary et al. 2003; Ingram 2012). The Kam (or 侗 "Dong" in modern Chinese) are a minority ethnic group in China with the main branches residing in the border areas of Guizhou, Hunan and Guangxi in the subtropical hilly areas of southern China. According to the 2010 China national census, the Kam population was 2.87 million with the majority living in rural areas practising traditional intensive farming. From a geographical

point of view, the Kam villages interlace with other villages of the Han, Miao and other ethnic groups. The geographical features of this ethnic relationship are an important basis for the study of the Kam people's identity.

At present, most Kam people are bilingual in at least the Kam language and standard Mandarin Chinese (or some local Chinese dialect). The social structure of the Kam people is mainly patrilineal in that the male dominates the community. Historically, various Kam villages (and sometimes other ethnic groups) formed a customary law-based village alliance – "Dong Kuan" ("侗款" in Chinese). This system still plays an important role in the current rural governance of villagers' autonomy in this region of China. Traditionally, there is no Kam orthography, and the Kam people keep historical records only through oral accounts, especially through the Kam Grand Choirs which is their main form of cultural heritage.

Zhanli Village, where the current project is physically based, consists of 168 households of 791 inhabitants. It is located in Congjiang County, Qiandongnan Miao and Kam Autonomous Prefecture in Guizhou province. According to local oral histories, the Kam's ancestors migrated from Guangxi and other southern places about 700 years ago.

In recent years, the socio-economic environment in the Kam areas has transformed rapidly and profoundly. On the one hand, people mobility in its modern sense still exists with a large number of the rural population moving to the city for a better life. On the other hand, due to the rapid development of infrastructure, especially in roads and transport, the Kam areas have become important destinations for "rural tourism" in Southern China. Along with this development, population mobility across these areas has accelerated in its scope and frequency in comparison with the past. People in this changing world are seeking and/or renegotiating their cultural and ethnic identities. In addition, with the universalization of education, institutionalized education has been changing the basic knowledge structure of the Kam people. In particular, an increasing number of the younger generation have received higher education and have extensive exposure to the outside world. All these inevitably bring about profound impact on the development of their cultural and ethnic identity.

Against such a backdrop, this paper sets out to examine the relationship between Kam people's sense of membership in their changing community, the beliefs and social practices that define this sense of membership, and its expression and manifestation in social behaviour. Thus, the research question of this paper is formed as "How Kam people's identity is represented and negotiated in narratives, crucially, what being Kam people means to them?" This question leads to more specific questions: how do the Kam people conceptualize being Kam? What kinds of self-representations are associated with this category in discourse? As argued by Silverman (1993: 108), "by analysing how people talk to one another,

one is directly gaining access to a cultural universe and its content of moral as-sumptions". This paper will, then, offer a small window through which to view the contemporary status of the Kam people and its relationship with other ethnic groups, especially the Han.

2. Identities in narrative

In sociolinguistic approaches to discourse and identity, identity is considered to be a fluid, multidimensional, social and cultural phenomenon that emerges from linguistic and other semiotic practices. It is often a two way construction that in-volves not only how one views oneself, but also how others view one (Paltridge 2012). According to Blommaert (2005), the context, "space" and "place", and the purpose of communication also shape identity construction. Blommaert (2005) also argues that people don't have identity, but produce, enact, or perform identi-ties. This means that identity is not an inherent trait of human beings, but a so-ciocultural construct that involves constant self-representation and manipulation. Following this anti-essentialist view of identity construction, this paper accepts the impossibility of representing a fully constituted, separate and distinct identity. It maintains that categories of identity are culturally constructed and are always in a constant state of change.

For the analysis of identity using linguistic interaction as data, Bucholtz and Hall (2005) proposed a framework with a key principle, that is, identity is the product rather than the source of linguistic and other semiotic practices. The so-ciocultural linguistic perspective on identity holds that identity is fundamentally a social and cultural phenomenon that emerges from linguistic and other semi-otic practices. Identity in this framework is not simply viewed as a psychologi-cal mechanism of self-classification that is reflected in people's social behaviour but rather as something that is constituted through social action, and especially through language (Bucholtz and Hall 2005).

Ethnic identity is a term used to describe people belonging to a certain ethnic or cultural group. It is the cognitive construct viewed by themselves and/or others as being associated with an ethnic group through feelings, beliefs, sociocultural customs and behaviours. Ethnic identity can be mutually constructed, negotiated and represented by the people themselves and by others. This paper examines a Kam village head's ethnic identity construction through investigating the narra-tives constructed by the interviewee in the interviews and daily interactions with the interviewers (i.e. the researchers). The foci will be his self-representation of the interior and exterior world, the mechanism of his membership classification and the performance devices that he draws on in the interactions.

In discourse studies, sociolinguistics, and anthropology, narrative is considered as the "basic", most "essential", mode of human communication (Riessman 1993, 2008). Narratives are everywhere. Narrative is ubiquitous in myths, fables, legends, epics, history, movies, novels, poems, dramas, conversation, interviews, news reports, meetings …it exists in nearly every aspect of human communication. As argued by Barthes (1977: 79), "Narrative … is able to be carried by articulated language, written or spoken, fixed or moving images, gestures, and the ordered mixing of all these substances". As a basic human way of making sense of the world, narrative is experience-as-told and as made social (Blommaert 2005: 84). As Riessman (1993) argues, people are leading "storied lives". In other words, narrative is constitutive of reality as well as of identity/subjectivity.

While narrative is regarded as a privileged window on human experience, Hymes (1996) insists that narrative is a universal function of language. But because of its deep context-embeddedness, its often "irrational" or emotive stance, and its connection to non-generalizable individual experience, it has been overlooked as a format of knowledge production and reproduction (Blommaert 2005: 84).

In Fludernik's (2009) definition,

> A narrative … is a representation of a possible world in a linguistic and/or visual medium, at the centre of which there are one or more protagonists of an anthropomorphic nature who are existentially anchored in a temporal or spatial sense and who (mostly) perform goal directed actions (action and plot structure). It is the experience of these protagonists that narratives focus on, allowing readers to immerse themselves in a different world and in the life of the protagonists
> (Fludernik 2009: 6)

From such a perspective, conversational and written (non-literary/literary) narratives are not as distant as they may seem, since the focus is not on events and actions per se but on the way humans experience and react to them.

3.   Interview participants

The interviewee is a robust male Kam in his late forties, who has lived in Zhanli village for his whole life with some occasional visits to Beijing, Shanghai, Hangzhou, Guiyang, and other big cities in China. He used to be the village Party secretary (i.e. the real village head), and a key spokesman of Zhanli when hosting outside visitors and investigators. In addition, he is often the first contact for visitors to the village as he usually provides accommodation for them. With only primary school education, he has been highly praised and respected for his adept skills in the Kam Grand Choirs at his home community. His father was one of the first

graduates of Guizhou *Minzu* University and served in various departments of local governments in Congjiang County. His sister and mother are the village's traditional medicine masters who hold the mysterious ability of "determining the sex of infants" of Kam people. This mystery is widely known and becomes an attraction for outside visitors, but the actual results are unclear. His son is a college graduate majoring in English with a one-year internship experience in Singapore; his daughter has completed undergraduate studies and is now preparing for the postgraduate admission test. The interviewee's background suggests that he holds a high position at his home community and is among the first of Kam people who have had rich experience in interacting with outsiders.

The primary data of this paper were drawn from over 10 hours of tape-recordings of the day-to-day interactions and interviews with him during the interviewers' two visits to the village during which the interviewers lived in his house. The first visit lasted for four days in May 2016 and the second lasted for seven days across August and September in 2016. During the two visits, the interviewers conducted four interviews with him. With varying attitudes toward our interviews, the interviewee responded quite passively to the first interview in May, but in the three interviews in August/September he was quite active and open to communication with the interviewers, in spite of intermittent expressions of indignation and helplessness. All the interviews with him on these occasions were permeated with his ambivalence, hesitation, and struggles in performing his ethnic identity. We will explore the reasons behind this in the following sections along with the detailed analysis of the interactions with him. Following De Fina's (2006) analytical model for identity construction, this paper investigated four dimensions of the sociolinguistic construction of the interviewee's ethnic identity, namely, the power relations in the interaction, the self-representation of the interior and exterior world, the membership categorization analysis, and the performance devices.

Prior to the formal interviews, the interviewers and the interviewee have exchanged information about their basic backgrounds, such as age, ethnicity, education, social and professional status, and purposes of the interviews. Concerning the interactions, both the interviewers (i.e. the Han) and the interviewee are considered to be competent members of their own communities (the Han versus the Kam) and to interact with each other to generate cultural knowledge of the interviewee's community (the Kam). It is assumed that the interviewee's ethnic identity is represented and negotiated in these interactions. The analyst plays the role of post hoc ethnomethodologist or discourse analyst in analysing the recorded data.

4. Power relations in interaction

In May 2016, two young postgraduate students including one junior chief investigator in this project visited the interviewee in Zhanli village. Both of them were junior researchers with limited fieldwork experience in ethnic minority areas in China, which provided the interviewee with opportunities to hold a dominant position during the interviews. As the former village Party secretary with frequent communication with outside visitors, the interviewee was very knowledgeable about Kam culture and history and excellent in singing the Kam Grand Choirs. These backgrounds ensured his relatively high social status in his ethnic community. As a resident and practitioner of Kam culture, he had the strong advantage of experiencing the culture. His personal experience and the cultural resources he has possessed were the main interests of the interviews, which to some extent also contributed to his relatively high status in the interview conversations.

However, if we place this project in a larger social context, this interviewee can be positioned as holding a weaker status. First, although the start and end of the interview as well as some topics of the interview were governed by him, the research interviews were formally designed by the interviewers. In one way or another, he was forced to respond to their questions. Second, he was under the pressure of having to speak in *Putonghua* (i.e. the standard spoken form of Modern Chinese) which was not his native language for the interviews. He was unable to express himself fully in that language especially when talking about the Kam culture. Moreover, with just primary school education, he mentioned more than once in the interviews that his Chinese language was very limited. The communication with the interviewers broke down several times due to the interviewee's inaccurate pronunciation, wrong lexicon used and bad sentence structures. These kinds of constraints and pressures as well as his role as a member of an ethnic minority pushed the interviewee to a weaker position in the interaction with the interviewers.

In the subsequent three meetings and interviews in August/September 2016, there was a major change in the power relation between the participants. These two rounds of interviews form a good comparison for understanding the identity building of the interviewee. Even though the interviewers of the first round participated in the second and third interviews, the main interviewer in the August/September interviews changed to being another chief investigator, a fifty-year-old male associate professor in the field of ethnic studies. This professor had rich experience in ethnographic fieldwork research and had been to Zhanli several times before. With deep knowledge of the Kam people in Zhanli and its culture, he had a clear advantage over the previous interviewers and he also possessed a higher social position in his own community. Consequently, the power relations between

the interviewers and the interviewee in these interactions changed drastically with the interviewer holding absolute power.

It is worth noting that even in these circumstances the interviewee still had the incontrovertible advantage of being the expert in Kam culture. During the interaction process, the interviewee appeared cooperative and active in response to the interviewers' questions. In addition, he provided more details by elaborating topics with additional information and he even initiated some new topics to share with the interviewers.

The comparison of changes in power relations across these two rounds of interviews enhances our understanding of the ethnic identity of the Kam people in this changing world. On the one hand, the power that the interviewers and the interviewees held was endowed by their statuses and the social communities to which they belong. On the other hand, along with the change of interview participants and the settings of interviews, the Kam man was skilfully negotiating his ethnic identity as well as manifesting his power in the interviews. In this case, the interviewing activities as social interactions between the interviewers and interviewees, to some extent, have also become a showground of interactions between a predominant ethnic group and a peripheral group in the society. The representation and negotiation of power relations across different "times" and "spaces" in the interviews have epitomized the ways of how the Kam people construct their ethnic identity in the society.

5.   Representations of the exterior and interior worlds

This section considers how the interviewee represents his interior (states of mind) and exterior (descriptions of social settings) worlds in the interviews with the interviewers. In this dimension, he reveals his group culture, classification mechanism, and evaluation criteria when describing cultural events, especially when mentioning the "others" (i.e. people outside his home community). The interviews themselves are social interactions in which both parties represent their own social backgrounds and identities. Consequently, the interviews collectively embody the group classification, attitudes and behaviours of both parties. However, the interviewers are more likely to play the moderator's role in guiding the communication, while both work together to generate cultural knowledge of the Kam community.

In his narratives with the interviewers, the interviewee's interior world generally appears passive, ambivalent and full of hidden conflicts. He even expressed a kind of anger, resistance and helplessness in the interviews in August/September (see Extract (2) for details). His hidden conflicts were revealed in two areas. First concerns his attitudes towards the Kam Grand Choirs. On the one hand, he is

pleased to see that the Kam Grand Choirs and its underlying culture are valued by the outside world and he is very proud of it. And he recognises this as an important resource to enhance the ethnic identity of Kam people. He also focuses on the Grand Choirs when introducing his people to outside visitors. But, on the other hand, he believes that the real Kam traditional culture or many of its original features are disappearing and what the outside world has been presented with is a "fake" phenomenon. Second, he overtly describes a peaceful, friendly and self-perceived equal relationship with the Han people. However, at the same time, a hidden perception of much alienation, inequality and tension was unconsciously revealed in his narratives.

The following extract is from the interview with him in May 2016, talking about the Kam ethnic culture and the Grand Choirs with the two junior interviewers. (Please be noted that for the verbatim extracts of the interviews in this paper, pseudonyms are used. W stands for the interviewee; H and Z, the junior interviewers; and J, the senior interviewer.)

> Extract (1)
> 1. **H:** 但现在像一些会说，现在反而国家重视保护这个民族文化、倡导这个民族文化 – [but now some will say, on the contrary the State attaches great importance to protect this ethnic culture, to advocate this ethnic culture …]
> 2. **W:** 虽然这么说，毕竟也是很多东西都要失传！[In spite of this, after all, lots of things are to be extinct.]
> 3. **Z:** 就是说他只是有这个口号并没有提出实际的行动来做这些事情，是吗？[That is to say that the State has only formulated some slogans but not put into practice, isn't it?]
> 4. **W:** 那只是一个掩饰而已！[It's just putting on a good face!]
> 5. **Z:** 掩饰？[A good face?]
> 6. **W:** 掩饰而已！就是像唱歌嘛！假如是现在的《多彩的贵州》，它还有民族一些文化在里面，但是变了，不是真实的，都是一些商业。[Yes, it's just a pretence. Just like songs. For instance, the Gala Show "Colourful Guizhou", there are some ethnic cultures in it. However, it has been changed. It's not real and very commercial.]
> 7. **Z:** 商业化？[Commercialization?]
> 8. **W:** 对！[Yes!]
> 9. **Z:** 它是把有利于商业的那部分发展了，就是不能发展商业的那部分它就不管？[Does that mean some part that is useful for business has developed, but others are to be neglected?]
> 10. **W:** 嗯，对对对！[um, right, right, right!]
> 11. ……

12. **W:** 那，那种传播只能让外面对我们中国有这个民族而已，这个东西保持不下来！有这么回事而已！[So, this kind of dissemination can only let the outside world know that China had this ethnic minority group and this thing cannot be kept up. It's just let people know there is such a thing!]

This extract shows that the interviewee tends to disagree with the social actions imposed on the Kam people, saying the current Kam Grand Choirs presented in the gala show is somewhat fake, only for commercial purposes with little essence being retained. Also in the interview, he claimed that the young people including his son don't really understand the lyrics of the Kam Grand Choirs, not to mention the deep cultural implications in the Grand Choirs.

In illustrating his exterior (descriptions of social settings) world, he described that Kam people are recognized and have become well-known mainly through the publicity of their performance of Kam Grand Choirs and dances in recent years. The inferior status that Kam people and other ethnic minorities experienced in relation to the Han people is now changing as a result of the increasing influence of the Kam culture and the implementation of national policies on ethnic minority protection.

He agrees that the identity of Kam is defined through their clothing, language, culture (e.g. the Kam Grand Choirs), and architecture (e.g. the drum tower). This definition concurs with the accepted standards of both the academics and the general public. Also, he emphasizes that the Kam Grand Choirs is the essence of Kam culture and is an important symbol that makes the Kam stand out from other ethnic groups. In practice, he is also actively involved in, and contributes to, the teaching and dissemination of the magnificent Kam Grand Choirs. However, the current forms and contents of Kam culture that are disseminated to the outside world are what he disagrees with, as in his view those are not the genuine Kam culture. He believes that the real Kam culture is vanishing and is being disrupted by lots of "outsiders" and "others" in the name of protecting the traditional culture. In his mind, that is actually an act of sabotage, which is most obvious in the destruction of the Kam Grand Choirs. In his narratives, his original world (both physical and spiritual) is disrupted by outside forces in the face of the economic boom in his village.

In addressing people outside his home village, he prefers to use geographical means rather than the official classification of ethnicity to differentiate ethnic groups in his daily life. That is he prefers to name other communities and people by linking them with their locations (i.e. towns, villages) rather than their ethnic groups. Only when associating with Han people and highlighting the renowned Kam cultural symbols will he emphasizes his identity of Kam. Otherwise, he tends to call himself a member of the "minorities" ("少数民族" in Chinese) rather than

a "Kam". It appears that he chooses to downplay his ethnic identity in interaction with outsiders. This issue will be further elaborated in the membership categorisation analysis below.

Extract (2) is from the second round of interviews in August/September 2016 mainly with one of the senior interviewers, talking about the State advocated "Traditional Village Construction" project that has impacted on the interviewee's home village.

Extract (2)

1. **H:** 当时工程队进村的时候你们有没有开群众会告诉你们这个事儿啊？ [When the construction team entered the village, did you have a public meeting to notify you what will be happening ah?]

2. **W:** 开群众会也开过！但是很多方面很多事情没有在大会上提到！就是现在做的事情都没有在大会上提。[Yes, we had public meeting! But many aspects, many things have not been mentioned in the meeting! What is happening now has not been mentioned in the meeting.]

3. **J:** 哦…… [ah…]

4. **W:** 是这样的！ [Yes, it is so!]

5. **J:** 嗯！[hm!]

6. **W:** 开呢肯定要说一些……把一些家啊、房子啊什么的收整！房前屋后这些东西都……都搞好！搞好哪个不愿意！你为我们好谁都愿意！谁都同意！"你们同不同意？""嗯！我们同意！"然后我们同意过后，还有很多事情没有在会上听说的都做起来了！所以群众、大部分的群众都有些不理解！ [At the meeting it should say something … to renovate some homes ah, the houses ah! To refine around the house! … who will say no to renovations! Everyone will say yes if you do good for us! Everyone agree! "Do you agree?" "Well, we agree!" But after we agreed, many things that were not mentioned at the meeting have been done! So the people, most of the people did not understand!]

7. **J:** 哦……那这种事情，这个寨老有没有发言权呢，现在？ [ah… as for those thing, whether the village head can say no, now?]

8. **W:** （pause 4 seconds）发言根本不起作用！ [It's no use to say no!]

9. **J:** 哦。[ah]

10. **W:** 没有，没有作用！ [No, no use!]

11. **J:** 嗯，那现在作用最大的是谁啊？是村两委还是县里面？ [mh, who has the right to say? Is it the village or the county?]

12. **W:** 现在来做的就是由县、乡来安排，什么事情都他们来做的。就是好像村里面都是随从！[Now everything is arranged by the county or the township, everything's done by them. It seems that the village is just following!]

13. **J:** 哦…… [ah…]

14. **W:** 我们本村好像没有自己的一些建议！[it seems that our village doesn't have our say!]
15. **J:** 哦，呵呵呵……[ah, hehe…]
16. **W:** 不存在！[no existence!]
17. ……
18. **W:** 其他就是做我们的耕田耕地，都是拿挖机来挖了我们都不知道、不知情！[Then, they took our farming land, to dig the land with the digging machine, we didn't know!]
19. **J:** 哦！[ah!]
20. **H:** 耕田耕地挖了都 – [Digging the farming land…]
21. **J:** 就上面那个停车场就是 – [Is that the parking lot up there…]
22. **W:** 对！就停车场！来当停车场用了！[Yes! It is the parking lot! Makes it the parking lot!]
23. **J:** 你们不知道啊？[You didn't know that?]
24. **W:**（pause）村里面的人都不知道啊！[No one in the village knew that before!]
25. (pause 2 seconds）
26. **H:** 村委知道吗？[how about the village committee?]
27. **W:**（略顿）应该是不知道！或者他们知道我们也不知道！不清楚！(pause) [Should not know! Or they knew but we didn't know! Not clear!]
28. **H:** 那没人找……闹事儿什么的？[Then nobody made some trouble (protest)?]
29. **W:** 闹不起嘛！你去哪儿闹？[You cannot make it! Where can you make it?]
30. **H:** 或者说就 – [or then…]
31. **W:** 你到乡里面……我们只有到乡里面去闹！县里面……[You can make it in the town, or in the county…]
32. **H:** 嗯，就是上访嘛！[mh, that is to make a petition!]
33. **W:** 县里面又……我们还到哪里去闹？[The county has… (pause) where can we make it?]
34. **H:** 就赔偿款什么都没跟你们说？[Did they mention compensation to you?]
35. **W:** 有说！他赔得太少了吧？一亩才多少？我听说一亩才两万！还是多少？[Yes, but compensated too little? One mu for about twenty thousand or what?](Note: 1 mu =666.7 m^2)
36. …
37. **W:** 其实征用这个地不是纯属是……国家来占地啊！不是啊！要是国家一定要……或者是开公路或者什么事情，要过去，那我们没话说！那是他一个高级宾馆用来修个停车场，那算是国家吗？[In fact, this land requisition is not purely … … by the State ah! no! If the State must … … or

build the road or something, in the past, then we have nothing to say! But this time that is a luxury hotel used it to build a parking lot, does that count as the State?]

In this interview, a story of land requisition without the concurrence of the Kam people unfolded during the progress of the conversation. At the beginning the interviewee just complained about the actions done by outsiders without the agreement or full understanding of the locals. He mentioned that something unstated in the public meeting had been done without the agreement of the villagers. Later on, he hesitantly expressed his indignation with many pauses and struggled to find a proper way to give vent to his annoyance (e.g. L.8, L.22, L.25). But in his view there would be no use to do anything about it and nobody will prepare to make petitions in the public (e.g. L.8, L.10). In addition, the real status of the institutional power behind the construction of the parking lot still confused him. On the one hand, it appears that he is not averse to State power and expressed gratitude to everyone who "did good for them"; on the other hand, he was infuriated by the social actions imposed on them without appropriate notification and consensus (e.g. L.12, L.34).

In his narratives, the world in which his people have been living for generations has been invaded and disturbed by the State power as well as other outside powers whose legitimate statuses are vague. With ostensibly good intentions, these foreign powers bring about changes and developments in the village. However, it seems the outsiders' real concern is not whether these changes and developments to the village are really welcomed and accepted by the local villagers. In his eyes, his people are scared of making public complaints about these changes. However, his narrative on this shows that, after hesitation and struggling in mind, he still chose to express his annoyances to the interviewers as outsiders.

While discussing his attitude towards State power, his attitude toward commercialization should not be neglected. This is because the interviewee explicitly expresses his negative attitude towards the commercialization of their culture as well as the transformation of local landscape due to commercialization. As evident in his words "那是他一个高级宾馆用来修个停车场，那算是国家吗?", the concern over commercialization is even greater than the concern over underconsulted social actions brought forth by the State power.

In addition, it is noticed that the interviewee was more reserved in Extracts (1) and (3) when talking with the junior investigators, but more open in Extract (2) with the presence of the senior interviewer. It appears that the people tend to be more willing to speak "the truth" when talking to someone who they believe is "an outsider or, especially, a senior", not involved or not part of their omnipresent power struggle, maybe with the hope that their bad situation or misfortune can be reversed by "the senior outsider".

6. Membership categorisation

Following Sacks' (1972a & b, 1992) original concepts or tools for membership categorisation analysis which include "membership categorisation devices", "membership categories", "category bound activities", and "the rule of economy and consistency", the following section teases out an analysis of how the interviewee conceptualised his membership as a Kam in China. Membership categories are references to persons such as "father", "mother", "teacher", and "Kam" and "Miao" in this paper. Membership categorisation "devices" refer to collections of these categories with organisational or social relevance. For instance, the categories "father" and "mother" may be used to form co-membership with other categories in an organisationally and situationally relevant "device", e.g. a family. Similarly, references to categories such as "Kam and Miao" can connect with the device "ethnic groups", which has social relevance here. That means, particular categories like "Kam" and "Miao" can go together within the social device "ethnic groups" as they did in this instance. Further, these categories involve working knowledge of background expectations such as their anticipated activities (i.e. category bound activities) in relation to each other. In this way, categories are selected, used and configured by members with an orientation to the topic at hand, so their use in any particular situation is purposeful or practical for that topic, rather than being simply abstract references following the rule of economy and consistency.

In examination of the membership categories that the interviewee and the interviewers draw on in the interactions, it is found that there are at least three membership categorisation devices in use in these situations. The first can be called the "state vs. people" device that both the interviewee and the interviewers employed when referring to "our country" ("我们国家" in Chinese) and "our China" ("我们中国" in Chinese) versus "us". Here "we/us" refers to all people in China including all ethnic and geographical groups. It is closely related to the speakers' recognition of the nationality identity and the political awareness of solidarity under the current government of China. The second is the "ethnic group" device such as "the Han", "the Kam and the Miao" that both the interviewers and interviewee draw on in referring to ethnic identity. The third is the "geographical group" device such as "Zhanli ren" ("占里人" in Chinese), or "Fuzhong ren" ("付中人" in Chinese) that refers to village peoples being identified by their geographic locations rather than their ethnic origins. These three membership categorisation devices intersected in the narratives in which any one device might predominate in a particular situation for a specific purpose.

Due to the multiplicity of personal identity, the membership category assigned to a person might vary in different contexts and narratives Baker 2004, Fitzgerald and Housley 2015. With regards to relations with outside people, the

interviewee conventionally chose the concepts of "we" vs. "they" which can refer to different people in different situations. For instance, the interviewee sometimes directly equates "they" to "the Han people", and sometimes expresses this reference implicitly, or even unconsciously, but the connotation of "we" covers the Han people merely when he intends to highlight the nationality identity in "our country" or "our China". Sometimes, he used "they" to refer to anyone outside his village regardless of ethnic or geographical groups. In general, the concept of the Han people is normally expressed as the "other" or "they" by the interviewee in the narratives.

Extract (3)

1. **H:** 就以您的经历来说，给我们讲一下您的故事。[Can you tell us your own story?

2. **W:** 凡是我小的时候，还没有接触外面的时候太多的时候，<u>我们</u>见到汉族<u>我们</u>觉得<u>我们</u>少数民族很渺小，很没有自信。就是……不敢那个……凡是见<u>他们</u>，我们觉得<u>我们</u>民族非常非常的低级这种意思啊。[When I was little I had not been exposed much to the outside world. We felt we were very trivial when seeing Han Chinese. I was not confident at all, which means … I was not able to … anytime when I saw them, which means we felt our ethnic group is very very inferior.]

3. **H:** 什么时候会有这种感觉？[When did you feel that?]

4. **W:** 就是那时候我们那个汉话都还不会说，人家说话的话都觉得<u>我们</u>这个、那个啊……[It was the time when we were not able to speak Han Chinese. What they said could mean we were somewhat …]
……

5. **W:** 就是通过我接触外面的人多了，也是在那个人家……现在<u>我们国家</u>也是逐渐是，就是保护一些民族的文化啊。然后我们就觉得<u>我们的民族</u>在<u>我们的中国</u>也有地位！[That is through more contacts with outside people, that people… now (we know) our country is gradually working on protection of our ethnic cultures. Then we feel that our ethnic group has a position in our China!]

6. **H:** 很独特哈？[That is unique?]

7. **W:** 嗯，是！[eh, yeah!]

8. **H:** 您有没有印象比较深刻的关于这方面的一些事儿？[Did you remember something very impressive in this regard?]

9. **W:** 最深刻的就是……还是少数民族的歌、舞这些东西，可以就是说我们……<u>我们</u>占里的话还没有出名嘛！那都是小黄啊、还有黎平那些，还有岩洞啊那<u>些</u>他们都出国啊！这些东西都是我们侗族的文化，可以包含！也算是<u>我们中国</u>的一些民族的东西，也是在<u>我们中国</u>也有影响。[The most impressive is … the songs and dances of the minority people. That said that we, "zhanli" was not famous! But people

from "Xiaohuang", "Liping" and "Yandong" (NB: all villages and counties of the Kam people) have been abroad. Those are all our Kam cultural things, may include! That should be our China's ethnic things. Those are influential in our China.]

As argued by Bilik (2010), ethnic groups are inherently unequal whenever the predominant ethnic group leads the national ideology. In the extract above from the interview in May 2016, at least two membership categorisation devices are in use in different places. Initially, the interviewee identified himself as "we ethnic minority" versus "the Han" who appear superior and more developed. In this context, the ethnic device is in use. But towards the end, he tends to acknowledge State power in promoting and protecting ethnic culture. He tends to use "our country" and "our China" instead. This is a clear example of the interviewee's deft shift between his national identity and his ethnic identity. He knows perfectly well when he should highlight which identity.

In talking about the relationship with the Han people, the interviewee indicates a clear transformation of the traditional "powerful" vs. "less powerful", or the "dominating" vs. "dominated" relation. In his narrative, he tends to pay more attention to "our" (the Kam people's) feelings and behaviours with relatively little attention to the other participating interviewers' (the Han people) feelings and attitudes either consciously or unconsciously. In this scenario, it appears that "the Han people" as "outside people" have not done anything particularly good or bad, but just provide a reference for him to compare with his or his group's behaviours. It is also evident that the contact with the outside world has helped build the self-confidence of the Kam people as an ethnic minority. However, sometimes the interviewee purposely frames this relationship as conflictual, especially when the clash is typically associated with circumstances imposed by the Han people in which case they are referred to as "they". These occasions covertly demonstrate that the Han people have played a dominating and oppressing role by imposing their needs and ideology on "us" and reshaping "our" culture. "We", in this relationship, play the role as victims. For instance, "we" had to compromise as we are "scared" even though we feel disaffected and want to resist. However, in other situations, the interviewee accepts what has been imposed and takes them as being a way of "protection" or of "making them well-known" by the outside world. Meanwhile, he abhors the reality that everything is changing beyond their control and the traditional Kam culture is disappearing. This kind of intricacy is well demonstrated by his use of the membership categories of "we" and "they" in different situations.

In the interviews concerning another ethnic minority group, the "Miao", as participants, the interviewee labelled them as "brothers" or "big brothers" and attested to this specific relation by associated social behaviours, such as "we stood

aside to let them pass" or "we will have their approval first", otherwise "It's bad luck". It is to be noted that the close relations between Kam and Miao in the interviewee's narrative seem to be restricted to the area geographically adjacent to each other such as the villages of Zhanli and Fuzhong. Further proof is required to make this relation in his narratives generalizable to the overall Kam and Miao relation.

It is evident also that the interviewee takes it for granted that he gets used to representing Zhanli village group as he equals "we" to "Zhanli ren" in answering questions about the Kam identity when interviewers use "you" (plural) to refer to "you Kam People" in their questions. In this regard, the interviewers might not have realised the interviewee would misunderstand their intention in treating "you" as "the Kam people". When this occurs, there is a mismatch between the membership categorisation devices used by the interviewers and those used by the interviewee as the interviewee inclines to flout the rule of consistency. Indeed the interviewers tend to draw on the ethnic device by using "you" to refer to "the Kam people", while the interviewee tends to draw on the geographic device by understanding that "you" only refers to the "Zhanli ren". For instance, in a conversation acknowledging the "Fuzhong ren" (who are the Miao) as their "brothers", the interviewee indicates that the "Meide" and "Xiaohuang" (i.e. two Kam villages) groups are only their "friends", even though these villages are located nearby (but not as close as "Fuzhong") and contain people who are also officially classified as Kam. His statement astonished the interviewers and challenged credibility. This can be clearly detectable from the tone and contents of the interviewer's questions.

This exchange stimulated us to think further about questions such as when the identification of the minority groups commenced in China, under what conditions it becomes significant in social life, and in what circumstances the interviewee would consciously acknowledge his Kam identity. The categorisation of ethnic groups such as the Kam, the Miao and the Han was initiated and enacted by the State system. However, the interviewee's narratives indicate that these ethnic categories such as the Kam are more likely an identity label enforced on the group that he belongs to, while, in fact, the geographic membership categorisation device is still predominant in their social interactions and communication with people outside.

His narratives also indicate that there is no doubt that "we", as ethnic minorities, are closer to each other than to the Han people. As opposed to the concept of "they" (often referring to the Han), "we" and "our ethnic minorities" are found to be interchangeable in many of his narratives. The concept of "we" does not simply include the Kam people but also the Miao (and might also include other minorities in his mind), even though his narratives showed that he had no experience in dealing with people other than the Kam, the Miao and the Han.

7.   Performance devices

This section considers the performance devices that both the interviewers and interviewees draw on during these interactions and interviews. In the first interview in May, it was always the interviewers who initiated the conversation with a lot of fillings, reminders and questions. The interviewee seemed to be just a passive interlocutor responding to the questions, but his responses were always positive, readily providing indiscriminate affirmative answers such as "eh, yeah, yes, etc." (See details in Lines 8 and 10 of Extract (1) and Line 7 of Extract (3)). But later in the conversations, he tends to reshape his views by further elaborating the details. In most cases, he used a relatively calm and neutral tone as if he was describing a factual situation in an objective way. However, he also used authoritative voices that non-Kam people find hard to refute, for example, "we think", "we are like this".

In the second round of interviews, due to the change in the main participants, the performance of the interviewee altered; he became more proactive and cooperative. He took initiatives to elaborate the details instead of passively responding to the interviewer's questions (see details in Lines 6, 33, 34 of Extract (2)). In addition, compared to the objective expression style in the first interview, he tended to be more emotional and subjective with occasional rhetorical and disjunctive questions in response to the interviewers' questions. He chose a more formal and serious tone to enlist the interviewer's recognition and affirmation by asking rhetorical questions. From the interviewers' perspective, the senior interviewer put up questions in a rational way and responded to the interviewee's questions in a timely and affirmative manner, indicating that he was engaged with the interviewee's narrative world with the purpose of encouraging him to continue with more specific details.

8.   Conclusion

This project investigated the ethnic identity of the former Kam village head through four dimensions of sociolinguistic analysis. It is found that the power of the interview participants is much endowed by their social status and the community to which they belong. These power relations can also change along with many personal and social factors, such as personalities and social status of participants, settings of the interviews, and even the topics under discussion. This paper also finds that the interviewee's self-representation of identity is filled with ambivalence and hidden conflicts. In the membership categorisation analysis, this paper reveals that there are at least three membership categorisation devices in use in the interactions. They are, the "state vs. people" device, the "ethnic groups" device,

and the "geographical groups" device. The interviewee's deft use of these devices is analysed with examples. In addition, the performance devices employed by both the interviewee and the interviewers have been analysed in light of the differences of "time" and "participants" in the interviews.

Overall, it is evident that the interviewee accepted being labelled a Kam and this acceptance was intensified through the interactions and interviews. However, the Zhanli village where the interviewee and his people have lived for generations is more like a small self-contained community with a closed operating system historically. First, this sociolinguistic analysis of their ethnic identity reveals that traditionally they tend to draw on a geographical categorisation device rather than the ethnic device in identifying other groups and they form relations with other people on the basis of the geographical device rather than the ethnic one, which, to some extent, is imposed by outside forces. This might have further confirmed the claims by Gidden (1990) and Appadurai (1996) that in premodern times, the sense of belonging to a community was tied to a specific place; while in contemporary society, the spatial relation between people and their communities is radically transformed. Second, as a result of advances in transportation, tourism and other socio-economic factors in recent years, more "others" make an impact on the Kam "minority" community. In addition, the establishment of a unified multiethnic state requires them to participate in the greater national society as citizens with blurred geographic categorisations.

Driven by these two main social forces, the Kam people are obliged to establish a new categorisation device (i.e. the ethnic one) to differentiate themselves from "others" with an aim of providing a unique identity in this wider society. Consequently, their Kam ethnic identity has been foregrounded in order to differentiate themselves from other ethnic groups, drawing on public representations of their cultures and customs, such as the Kam Grand Choirs.

Similarly, the interviewee's alliance with other ethnic minorities rather than the Han might be a consequence of the following social factors. First, the Han-dominated State discourse has classified all non-Han ethnic groups as ethnic minorities as opposed to the majority Han. This social practice has been enforced by State minority protection policies, including regional designated autonomy and the administration right of self-governance in the ethnic minority regions. Despite the public consensus of identifying Kam people by their exterior characteristics and culture, this paper argues that the social factors discussed above are the main driving forces in constructing the ethnic identity of the Kam people in contemporary China, and this process is under constant negotiation and manipulation.

References

Appadurai, Arjun. 1996. *Modernity at Large: Cultural Dimensions of Globalization*. Minneapolis: University of Minnesota Press.

Baker, Carolyn. 2004. Membership categorization and interview accounts. In *Qualitative research: Theory, method and practice (2nd edition)*, ed. by David Silverman, 162–176. London: Sage.

Barthes, Roland. 1977. Introduction to the structural analysis of narratives, in *Image-Music-Text*, selected and translated by Stephen Heath. London: Fontana.

Bilik, Naran. 2010. Yuyan yishixingtai: yuyanrenleixue xinpian ("Language ideology: A new approach in linguistic anthropology"). *Shijie Minzu* (Ethnicity in the World) 4: 38–41.

Blommaert, Jan. 2005. *Discourse*. Cambridge: Cambridge University Press.

Bucholtz, Mary and Kira Hall. 2005. "Identity and interaction: A sociocultural linguistic approach." *Discourse Studies* 7:4–5.

De Fina, Anna. 2006. Group identity, narrative and self-representations. In *Discourse and Identity*, ed. by A. De Fina, D. Schiffrin and M. Bamberg. 351–375. Cambridge: Cambridge University Press.

De Fina, Anna, Deborah Schiffrin, and Michael G. W. Bamberg. 2006. *Discourse and identity*. Cambridge: Cambridge University Press.

Fitzgerald, Richard and Housley, William (ed.). 2015. *Advances in membership categorisation analysis*. London: Sage.

Fludernik, Monika. 2009. *Introduction to narratology*, London/New York: Routledge.

Geary, Norman, Ruth B. Geary, Ou Chaoquan, Long Yaohong, Jiang Daren, Wang Jiying. 2003. *The Kam people of China: Turning nineteen*. London/New York: RoutledgeCurzon.

Giddens, Anthony. 1990. *The Consequences of Modernity*. Stanford, CA: Stanford University Press.

Hymes, Dell. 1996. *Ethnography, linguistics, narrative inequality: Toward an understanding of voice*. London: Taylor & Francis.

Ingram, Catherine. 2012. Tradition and Divergence in Southwestern China: Kam Big Song singing in the village and on stage, *The Asia Pacific Journal of Anthropology*, Vol. 13:5, 434–453.

Paltridge, Brian. 2012. *Discourse analysis: an introduction* (2nd edition). London: Bloomsbury.

Riessman, Catherine Kohler. 1993. *Narrative analysis*. Newbury Park, CA: Sage Publications.

Riessman, Catherine Kohler. 2008. *Narrative methods for the human sciences*. Thousand Oaks, CA: Sage Publications.

Sacks, Harvey. 1972a. An initial investigation of the usability of conversational data for doing sociology. In *Studies in social interaction*, ed. by D. Sudnow, 31–74. New York: Free Press

Sack, Harvey. 1972b. On the analyzability of stories by children. In *Directions in sociolinguistics: The ethnography of communication*, ed. by J. J. Gumperz & D. Hymes, 325–345. New York: Holt, Rinehart, and Winston. (Reprinted in R. Turner (Ed.), 1974, Ethnomethodology. Penguin)

Sacks, Harvey. 1992. *Lectures on conversation* (2 vol. ed.), ed. by G. Jefferson, introduction by E. Schegloff. Oxford: Blackwell.

Silverman, David. 1993. "Beginning Research". *Interpreting qualitative data. Methods for analysing talk, text and interaction*. London: Sage Publications.

Specialised corpora for Chinese language education in Singapore

Hock Huan Goh, Jinzhan Lin and Chunsheng Zhao
SCCL, Nanyang Technological University, Singapore

Corpus linguistics is crucial to language education, but many corpora do not pay enough attention to curriculum and pedagogical needs. To address this issue and in view of Singapore's unique language environment, the Singapore Centre for Chinese Language built two specialised corpora for Chinese language education in Singapore, which comprise a Written Corpus and a Spoken Corpus. The Written Corpus provides information on Chinese characters, vocabulary words and sentence structures used in written materials daily; while the Spoken Corpus provides guidelines for attainable spoken proficiency of primary school students at different academic levels. With these corpora, curriculum developers can design syllabi with greater precision on the language content and address the learning gap for Chinese language proficiency. As for teachers, an online resource platform developed based on the Written Corpus provides them with authentic materials and practical applications (such as the text grading module) as reliable tools and resources for lesson preparation and learning assessment.

Keywords: Specialised Corpora for Language Education, Singapore Daily Written Chinese Corpus, Singapore Primary School Children Spoken Chinese Corpus, Singapore Chinese Language Teaching Resource Platform, corpus application

1. Introduction

Corpus linguistics and language teaching have long been considered complementing disciplines. Since the 1990s, efforts have been made to advocate the construction and application of corpora for language education (Leach 1997; Granger, Hung and Petch-Tyson 2002; Sinclair 2004; McEnery and Xiao 2010; Zhang and Tao 2018). However, many corpora, especially Mandarin Chinese corpora, are found to be less concerned about the teaching of the language (Lin et al. 2015; Yu

2009), because these corpora are built mainly for the analysis of linguistic phenomenon and have overlooked the need for curriculum development and language teaching (Braun 2007; Cook 1998; Gavioli and Aston 2001). To tap on the affordance of a corpus for mother tongue language teaching in Singapore, the Mother Tongue Languages Review Committee (MTLRC 2010) advocated the development of a corpus of daily vocabulary and sentence structures for all mother tongues, including Mandarin Chinese (henceforth "Chinese"). Echoing this initiative, the Singapore Centre for Chinese Language (SCCL) has, since 2011, committed to developing specialised corpora for Chinese language (CL) education.

Specialised corpora for language education, in short, refer to corpora designed for the purpose of language teaching and curriculum development. Leech (1997) and McErny and Xiao (2010) maintained that such corpora should serve three broad functions: (1) indirect use for teaching (e.g. reference/dictionary publishing, syllabus design and development, and language testing); (2) direct use in teaching (e.g. teaching linguistics or foreign language in universities); and (3) teaching-oriented support (e.g. learner corpora and interlanguage analysis). With reference to these three functions, the SCCL corpus development team coined the concept of "Specialised" which consists of two aspects: Firstly, the building and construction of such corpora, data sampling and processing should, to its best possibility, take into consideration the needs of local syllabus design and teaching. Secondly, upon the completion of the corpora, practical features should be further developed so that educators can access information in the corpora for reference and professional development. Useful applications should also be developed to support educators' daily work, such as the selection of supplementary teaching materials and assessment instruments.

With such curricular initiatives taken by the MTLRC (2010) two specialised corpora for CL education, namely Singapore Daily Written Chinese Corpus (henceforth "Written Corpus") and the Singapore Primary School Children Spoken Chinese Corpus (henceforth "Spoken Corpus") were launched. The Written Corpus was constructed to identify the level of CL Singapore Chinese students of primary and secondary schools should achieve, in terms of vocabulary and sentence patterns. Whereas the Spoken Corpus outlined the basic CL vocabulary and sentence patterns that primary school Chinese students can produce for their topic of interest, at different ages. There are three stages to the development of both corpora: (1) data collection; (2) data processing; and (3) data analysis. The first part of this paper describes the construction process of these two corpora (see Section 3 and 4). While the second part analyses the corpora and elaborates on the key findings relevant to CL education in Singapore (see Section 5 and 6). The language information of these two corpora will provide curriculum developers with the range of Chinese characters, vocabulary and sentence patterns to be covered

in their curricula. Educators will also have a reference for the basic language proficiency primary school students of different ages and home language backgrounds can attain.

In addition, based on the Written Corpus, the research team developed the Singapore Chinese Language Teaching Resources Platform for CL teachers in 2013 (see Section 7). This platform provides teachers with a function to search for Chinese characters, words and sentence patterns in the corpus; key linguistic information such as character structures, word collocation etc. are also provided together with authentic examples. Furthermore, this platform provides teachers with a text grading parser to electronically input any Chinese text and assess if the text is suitable as supplementary reading material or as test items for students.

2.   Specialised corpora for language education

Ever since the first computer corpus, the Brown Corpus, was created in the early 1960s, considerable development has been made in corpora construction and application. In many disciplines of linguistics, corpora have proven to be valuable resources: they are used for creating dictionaries, studying language changes and variations, facilitating contrastive analyses of languages, advancing developments in automated translation and translation memory, spearheading studies in natural language processing (NLP) and artificial intelligence, and uncovering the process of language acquisition etc. (McEnery and Xiao 2010; Meyer 2002).

Generally, corpora are developed for different purposes. For instance, the Brown Corpus is a balanced corpus with the purpose of permitting both systematic studies of individual genres and comparison between genres of written English. The Penn Treebank corpus was created for linguists with computation interests so that they could conduct research in NLP, an area of study that involves the computational analysis of human language behaviour and cognition. Corpora are also widely applied in language education and one field of application is lexicography, in which a number of dictionaries, such as the Collins COBUILD English Dictionary, have been produced based on corpora such as the Birmingham Corpus and the Bank of English Corpus. Furthermore, some corpora were developed to facilitate studies on language acquisition, such as the Child Language Data Exchange System (CHILDES) and the International Corpus of Learner English (ICLE).

Apart from the aforementioned applications of corpora, the development of learners' corpora has also benefited language education. Such corpora that comprise learners' output could be used to address key learning issues and to develop teaching strategies which improve learning outcomes. For example, Altenberg and Tapper (1998) discovered via corpus analysis that Swedish English learners

overused informal connectives in their writing and they proposed that the best teaching strategy for Swedish learners is to expose them "to a greater range of registers and to a more extensive training in expository writing." In short, by using learners' corpora, teachers have a more accurate assessment of how their students use the language. Such information can be used to develop specific teaching strategies which can be incorporated into textbooks or lessons to aid language learners.

Besides developing teaching strategies using learners' corpora, studies have advocated the learning of languages through corpora by the students themselves. This method is known as "data-driven learning (DDL)", and it is asserted that DDL helps learners observe language patterns in a second language (Johns 1991; Hadley 2002). This method attempts to implicitly inculcate linguistic knowledge by providing ample samples of relevant authentic language materials for language learners to tease out the grammatical patterns or word usages. It stands in contrast to the traditional approaches of teaching grammatical patterns or word usage explicitly through textbooks that make use of non-authentic examples and materials. Johns (1991) likens the language learner (on the DDL model) as a researcher, analysing target language data and becoming familiar with the language through the regularities and consistencies encountered. Kettemann (1995) further demonstrated how students can make use of a concordance programme to identify various kinds of grammatical distinctions in English, for instance, the difference between the past perfect tense and present perfect tense.

Despite the aforementioned efforts, corpora have not been fully applied in language education. Römer (2006: 121) pointed out that "despite the progress that has been made in the field of corpus linguistics and language teaching, the practice of ELT has so far been largely unaffected by the advances of corpus research". There seems to be a mismatch between the growing number of applied native corpus research and corpus-based teaching materials and everyday teaching practices (Meunier 2011). One reason for this mismatch could be that these corpora are usually constructed by university research teams and most applications in these corpora are meant for academics. The focus of these corpora and their applications require specialised knowledge in linguistics, and hence do not facilitate teaching and learning in basic education.

When it comes to the development of Chinese corpora and resources in CL education, numerous renowned corpora (such as "Peking University Centre for Chinese Language Online Corpus, CCL Corpus" and "Beijing Language and Culture University Chinese Corpus, BLCU Chinese Corpus") have been developed. There are similarly many online resources (such as Chinese Linguistic Data Consortium, Media Language Corpus) constructed for CL education. However, these corpora and resources are mainly constructed with materials frequently engaged by first language or mother tongue users, and their findings are geared

towards the teaching of CL as a first language. Though some of these corpora have provided character lists, wordlists and grammar indices with an international standard, they merely suit foreign learners who choose to be exposed or immersed in a first language environment such as China. As affirmed by Xiao et al. (2009), CL frequency dictionaries – resources generated by these first language corpora since the 1920s – are mainly confined to China. And these corpora and resources will be less relevant or authentic for CL learners in a unique multilingual environment such as Singapore, Malaysia and other ASEAN countries, whereby the CL in these environments synthesises social, racial and cultural vibrancies for political stability.

Apart from the aforementioned first language-oriented corpora and resources for CL education, there are indeed corpora and resources built for the teaching and learning of CL as a second or foreign language, e.g. Liu (1973), Chinese State Language Commission (CSLC 2010) and Xiao et al. (2009). Two of these relevant corpora shall be referred to and compared with our corpus in the later sections. One of them is *The Graded Chinese Syllables, Characters and Words for the Application of Teaching Chinese to the Speakers of Other Languages* published by China's Ministry of Education and the China National Language and Character Working Committee in 2010. This corpus drew data from five sources, including (1) radio broadcasting, (2) audio media, (3) newspaper and the internet from 2005 to 2008, (4) current CL textbooks and teaching materials, and (5) the National Language Committee Corpus, which contained language materials from 1919 to 2002. (CSLC 2010: 2) These five corpora hold more than 40 billion characters of data. This document consists of three listings, i.e. a list of Chinese syllables, a list of 3,000 most commonly used Chinese characters and a list of 11,092 most commonly used Chinese words. All these listings comprise three grades and an extension for the third grade (for details on the distribution of words by syllables length, see CSLC 2010: 3).

The other corpus is the *Frequency Dictionary of Mandarin Chinese* edited by Xiao et al. (2009). This dictionary is based on the Lancaster Corpus of Mandarin Chinese, UCLA Written Chinese Corpus and a Chinese-English Parallel Corpus not specifically introduced. These corpora eventually yielded 73 million characters of data comprising both written and spoken language gathered from news, daily newspapers, face-to-face conversations, telephone calls, cross-talks, movie and play scripts, interviews, storytelling, public lectures, radio broadcasts, public debates etc., covering various periods between 1990 and 2006 depending on the different registers. Besides these general corpora, Tao (2015) constructed an oral corpus that consists of 54 face-to-face conversations among friends and family members, recorded in the 1980s and 2005. The data were word-segmented and tagged for parts-of-speech information, totalling to 344,141 word-tokens. Upon

analysis, Tao (2015) highlighted the saliency of the use of monosyllabic words and lexical chunks in spoken data, which suggested the importance of these features in curriculum development and delivery.

Despite the abovementioned developments in CL corpora and their findings, corpus application in CL teaching and learning in Singapore has yet to attain its full potential. For instance, there is no publicly accessible corpus for CL curriculum development and classroom teaching before the current corpora came into being. As such, curriculum developers and educators could only turn to Chinese corpora accessible via the internet (such as the "CCL Online Corpus" developed by Peking University or "Academia Sinica Balanced Corpus of Modern Chinese" developed by Academia Sinica) for reference. Singapore educators and curriculum developers have not been able to get sufficient or accurate information in these corpora, let alone applying corpus-based methods (such as DDL advocated by Smith 2011; Huang and Meng 2013) in the classroom to suit Singapore's context, whereby CL has become a second language of most Chinese families in the past 15 years.

In view of previous studies on corpus development and application and Singapore's CL teaching and learning needs, the SCCL constructed a Written Corpus and a Spoken Corpus tailored for CL education. The speciality of the corpora is ensured in three ways: Firstly, the data sampling and collection are carried out with reference to pre-data collection surveys to ensure the comprehensive coverage of media-types and student-types. Secondly, at the data processing stage, many significant and useful features (such as parts-of-speech, genre, subject and grade of texts) are annotated for educators to access information with ease. In addition, the segmentation of data has taken into consideration how the language is taught in classrooms. For this purpose, phrases such as 巴士转换站 (bus-interchange), 小贩中心 (hawker centre) are not further segmented in accordance with the linguistic conventions because such phrases are usually taught as a whole in Singapore. Lastly, tapping on the information affordance of the corpora, language teaching specific functions (such as text grading and word collocation search) are designed and developed for CL educators, so as to support their daily work (such as choice of supplementary materials, test/quiz setting etc.).

3. Data collection

Before commencing data collection, pre-data collection surveys were administered to students related to the two corpora. For the Written Corpus, 720 primary school students (from primary 1 to 6) and 640 secondary school students (from secondary 1 to 4) were recruited for a Media Engagement Survey to identify the popular newspaper and non-newspaper media engaged by primary and secondary

school students. To ensure representativeness of the surveyed students, participating students were also sampled by school-types, i.e. English mission schools, Special Assistance Programme (SAP) schools and mainstream schools, at the ratio of 1:1:2 (see Table 1 for the sample distribution among school-types and academic levels). The survey formed a basis for the scope, themes and percentage of the data collection.

Table 1. Sampling for Media Engagement Survey

Level (L)	L1	L2	L3	L4	L5	L6	Total
Primary Level:							720
• Mission Schools	30	30	30	30	30	30	
• SAP Schools	30	30	30	30	30	30	
• Mainstream Schools	60	60	60	60	60	60	
Secondary Level:							640
• Mission Schools	40	40	40	40			
• SAP Schools	40	40	40	40			
• Mainstream Schools	80	80	80	80	–	–	
Total Sample	–	–	–	–	–	–	**1360**

With the survey results, the percentage of data to be collected for each media-type was determined. But the planned percentages varied slightly due to the availability of materials for each media-type. Table 2 presents the data collected from the media.

As shown in Table 2, data from the media could be categorised into two types: newspaper and non-newspaper. The latter formed a greater percentage of the corpus compared to the former, 55.6% and 44.4% respectively, indicating that students engaged more non-newspaper media than newspaper in their daily encounters. The newspaper media included various student-specific and public newspapers from the period 2007 to 2009, whereas the non-newspaper media comprised the respective stated sources published in 2010. To ensure level appropriateness, two sub-corpora were created, i.e. a primary level and a secondary level. As shown in Table 2, newspaper media in the primary sub-corpus comprised more data from the students' newspapers (e.g. *Damuzhi*), whereas newspaper media in the secondary sub-corpus comprised more data from newspapers catered to the general public (e.g. *Lianhe Zaobao*). This difference in data composition of the sub-corpora took into consideration the genre preferences of students from different academic levels, so as to ensure level appropriateness.

As for the Spoken Corpus, two surveys – Home Language Survey and Topic-of-Interest Survey – were administered to the participating students (see Table 3)

Table 2. Data collected for the Written Corpus

	No. of articles	Character token	Percentage in corpus
Newspaper Media:			
• *Douhao 逗号* (Student-Specific)	279	292,125	11.07%
• *Damuzhi 大拇指* (Student-Specific)	300	239,272	9.07%
• *Lianhe Zaobao 联合早报*	703	442,010	16.76%
• *Lianhe Wanbao 联合晚报*	248	94,430	3.58%
• *Shin Min Daily 新明日报*	297	102,786	3.90%
Non-Newspaper Media:			
• *Comics*	12	132,876	5.04%
• *Story Books and Literature Books*	169	396,672	15.04%
• *Magazines*	285	240,397	9.11%
• *Websites*	380	143,463	5.44%
• *Assessment Books*	83	362,234	13.73%
• *Textbooks*	477	106,628	4.04%
• Others: *Flyers, Noticeboard, Brochures*	380	85,097	3.23%
Total Sample	**3,613**	**2,637,990**	**100%**

before the data collection. The Home Language Survey solicits the participating students' home language exposure and the results were presented as the Chinese Exposure Index (CEI), which identified and categorised the students in terms of home language dominance: Chinese-speaking, English-speaking or bilingual-speaking. Besides that, a Topic-of-Interest Survey was administered to the students to identify the topics they are interested in. The identified topics included *interesting/funny encounters, family members, favourite classmate, favourite teacher, favourite subject, hobbies, food, movies/cartoons, animals, favourite gifts/toys* and *weekend activities*. These topics were used to draft interview questions and design classroom activities for data collection. For the Spoken Corpus' data collection, four methods were used: free talk/one-on-one interview, picture-elicited talk, home talk and classroom talk. Tables 3 and 4 illustrate the distribution of data collected by each method.

As illustrated in Table 3, the sampling of participants for free talks/one-on-one interviews and picture-elicited talks followed the principle of stratified sampling, whereby eight primary schools were randomly selected according to school types and their distribution nationwide. These school types ensured balanced coverage of participants from different home-language backgrounds. For each school, three grades were selected: Primary 1, 3 and 6. These selected grades correspond to the

Table 3. Sampling of students for Spoken Corpus

School type	No. of sch.	Projected No. of students (actual No. of students)			Total
		Primary 1	Primary 3	Primary 6	
SAP School	2	60 (60)	60 (59)	60 (59)	180 (178)
Mission School	2	60 (56)	60 (62)	60 (61)	180 (179)
Mainstream School	4	120 (109)	120 (113)	120 (121)	360 (343)
Subtotal	**8**	**240 (225)**	**240 (234)**	**240 (241)**	**720 (700)**

three phases of CL learning outlined in the CL syllabus: Basic/Foundation I (i.e. Primary 1 and 2), Basic/Foundation II (i.e. Primary 3 and 4), and Orientation stage (i.e. Primary 5 and 6). The projected number of participants was 720, but absentees brought the finalised number down to 699.

Table 4. Sampling for Home Talk

School type	No. of Sch.	No. of students			Total	
		Primary 1	Primary 3	Primary 6	No. of students	Hours recorded
SAP School	2	6	6	6	18	18
Mission School	2	6	6	6	18	18
Mainstream School	4	12	12	12	36	36
Subtotal	**8**	**24**	**24**	**24**	**72**	**72**

Free talks/one-on-one interviews and picture-elicited talks might be efficient ways to gather spoken output, but it is not as natural as the spontaneous daily conversations. Therefore, home talks of some participating students were collected to substantiate the Spoken Corpus with spoken production occurred in daily-life scenarios. As shown in Table 4, three students of each grade from each school were sampled for the home talk data collection. Each participant was audio-recorded twice by their parents during two family activities (e.g. dinner, playtime, family outing). Each recording lasted about 30 minutes and 54 hours of data were collected. Apart from home-talk, the classroom is another environment where students have to speak Mandarin on a daily basis. Thus, two 30-minute Chinese oral classes were arranged with each of the participating schools. For each class, teachers guided students to talk about a favourite topic and the entire class was video-recorded. Table 5 shows the number of classes involved in the classroom talk segment.

Table 5. Sampling for Classroom Talk

Participants	No. of classes	Hours of recorded data
Primary 1	16	8
Primary 3	16	8
Primary 6	16	8
Total	**48**	**24**

This section describes the first stage of corpus building which consists of steps taken to ensure the speciality of the Written Corpus and Spoken Corpus, i.e. student-specific. Combining results generated by the two corpora, curriculum developers and teachers can identify the gap between students' competence at each level and their expected learning attainment for CL. Curriculum developers can hence build a more precise curriculum to bridge the gap between students' ability and the targeted competency.

4.   Data processing

The second stage of corpus building is data processing. The Lingjoin Text Mining & Semantic Parser Platform (henceforth "Lingjoin platform") was used to perform word segmentation and Part-of-Speech (PoS) tagging, and the SCCL Proofreading and Annotation System was used for proofreading the segmentation and tagging results of words. After which, the sentence-types of each sentence in the Written Corpus were manually annotated. The data processing of the Spoken Corpus followed a similar procedure except that the Spoken Corpus had to be transcribed into text before carrying out the said procedures. Every audio recording's transcript was proofread by at least two members of the research to ensure the accuracy before being consolidated into the corpus.

As shown in Figure 1, the transcripts of each child were put through multiple document checks by data-mining software to identify repetition, format discrepancy, code inconsistency etc. After that, all data were fed into Lingjoin Platform for automatic word segmentation and PoS tagging. Following this automated process, manual proofreading of segmented and annotated results of each word in the corpus was performed. Upon completion, the linguistic information (such as the word type, word token, PoS distribution of each child) was generated and tagged, and comparison of correlation was conducted to examine the relationship between home-language exposure and language output.

The two corpora underwent similar data processing procedures. It began with cleaning the collected data before consolidating it into a database, whereby the

data were processed by means of automated tagging or manual annotation. As both corpora involved raw data from different sources and in various formats, different processing tools or software were used. Some of the tools/software used in the process are Transcriber, PowerGREP, AntConc, UltraEdit, and the SCCL Proofreading and Annotation System developed by the research team. The following description illustrates how different language materials were processed for the two corpora:

– Processing of newspaper materials: newspaper materials were downloaded from Singapore Press Holdings Limited's "Newslink" database. These materials covered headline news, social issues, leisure, travel, entertainment, sports and advertisements from the period 2007 to 2009. In the process of downloading, some materials had garbage characters, inconsistency in punctuation and unnatural paragraphing. For these issues, we conducted manual proofreading and deleted some unreadable data. Simultaneously, different editing software programmes were used to ratify the punctuations and unnatural paragraphing.
– Processing of children readers, novels, comics and magazines: these data were collected based on titles suggested in the students' reading surveys conducted in 2010. The chosen materials were scanned, processed through OCR and have their format standardised. Subsequently, these extracted texts were manually proofread and consolidated into the corpus.
– Processing of audio and video recordings: generally, audio and video recordings were collected from scripted popular local TV series and movies screened in Singapore from 2009 to 2010 for the Written Corpus. While the Spoken Corpus was mainly based on students' output in interviews, classroom lessons and home activities. All recordings were transcribed in a standardised format using the Transcriber software. Transcripts were the nmanually edited and proofread at least twice before being consolidated into the respective corpora.
– Processing of data from webpages: these data mainly consisted of CL education websites frequently visited by primary and secondary students as reflected in their survey. Though the gathering of webpage data was relatively direct and easy, there were issues of differing internal codes due to the different webpage development software each site used. Therefore, the webpage data's formats and internal codes had to be standardised and manual proofreading carried out to ensure consistency with the original text.

After pre-processing the different types of data as illustrated above, the data were consolidated and processed through the Lingjoin Platform in the following order:

– First, checking text files and eliminating duplicates which may occur in the process of data consolidation.

–   Second, there were numerous special words and common word collocations in both corpora, which were not in the common wordlist of the Lingjoin platform. These special words and common collocations of words had to be identified as new words in the Lingjoin platform, and the PoS of these new words and collocations had to be defined on the platform for an accurate processing of data.

–   Third, after identifying new words and collocations, the Lingjoin platform performed automated word segmentation and PoS tagging and generated a list of words for proofreading on the SCCL Proofreading and Annotation System. With the help of concordance features on the SCCL Proofreading and Annotation System, the generated wordlist was cleaned and finalised based on word segmentation and PoS tagging standards released by the Standard Administration of China (2006).

–   Fourth, upon confirming the list of words and their respective PoS, the sentences in the corpus were annotated for sentence-types and sentence-forms on the SCCL Proofreading and Annotation System, using the combined technique of keyword annotation regular expression annotation and manual sentence-be-sentence annotation.

–   Lastly, with the proofread wordlist and annotated sentence-types, final lists and critical indices of words and sentences in the corpora were extracted using the SCCL-developed tool for data extraction and analysis. With that, the final wordlist and sentence-type listing were generated and the indices (such as word frequency, lexical diversity, lexical density, lexical coverage, sentence-type frequency, mean length of utterance etc.) were also extracted.

In summary, the Written Corpus and the Spoken Corpus construction underwent the processes of data collection, data pre-processing, data consolidation, data segmentation, annotation and proofreading, and data delivery. These processes are illustrated in Figure 1:

When proofreading the list of words in the abovementioned process, the definition of word boundaries and its PoS took reference from the "Chinese Language Part-of-Speech Annotation Guideline for Computing Technology Lab". This guideline underpins the Lingjoin Platform and is developed by the Institute of Computing Technology, Chinese Academy of Sciences. For uncertain word boundaries and PoS, the research team took reference from the "Contemporary Chinese Language Word Segmentation Specification for Information Processing" released by the Standards Administration of China (1992), and its subsequent refined versions such as "Standards of POS tags of Contemporary Chinese for CIP" released by Standards Administration of China (SAC) in 2006 (for the list of PoS taggers and its examples see Appendix A). While the proofreading of word
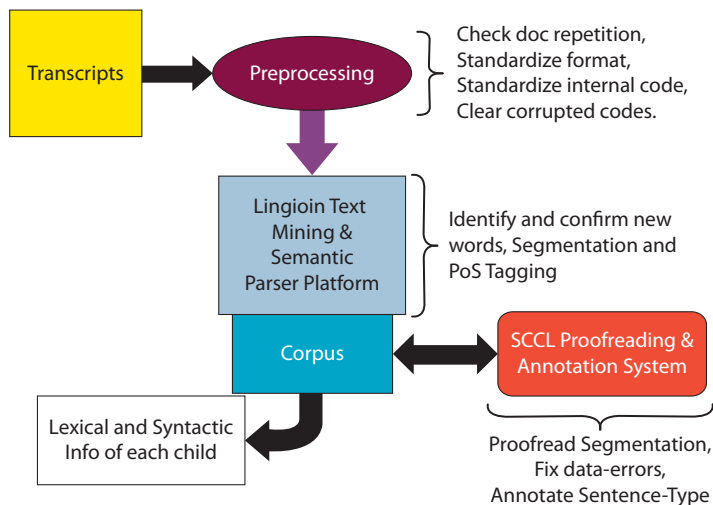
Figure 1. Continua of Mandarin competences and language exposures

segmentation abided by the aforementioned standards and guidelines, it also took into consideration the teaching needs of CL. As most teachers teach certain terms or concepts as a whole rather than word-by-word, the research team decided to forgo the convention of segmenting words to their smallest forms, such as morpheme. Instead, we preserved these units as a word or phrase in the wordlist. For example, bi-syllabic verb complement constructions, such as 拿起 (to pick up), 打开 (to open), 睡着 (to fall asleep), 走来走去 (to walk back and forth) were not further segmented. Some terms (especially those familiar to Singaporeans), such as 巴士转换站 (bus interchange), 综合诊疗所 (polyclinic) were also preserved as a unit without further segmentation.

For the annotation of sentence types, both corpora made reference to Lu (2003). In his renowned work on *Basics of Chinese Grammar*, he categorised Mandarin Chinese sentences into four types, namely, declaratives, interrogatives, exclamatory and imperatives. He also categorised Mandarin Chinese sentences into 158 patterns (inclusive of clausal relations in complex sentences) and 27 forms. All sentences in both corpora were tagged against these 189 categories using the SCCL Proofreading and Annotation Platform. While there were differences in grammar between spoken forms (see Li and Thompson 1981; Zhao 1979) and written forms, the research team decided to solely annotate the data in accordance with the grammar categorisations for written forms as syllabus and textbooks are generally in this form. (For details and examples of sentence types and sentence patterns annotated, please see Appendix B).

Upon completing the above-mentioned processes, the corpora building entered its third stage, whereby the two research teams extracted indices to illustrate

the distribution of characters, words and sentence patterns. The finalised character lists, wordlists and sentence-type lists were also generated at this stage; curriculum developers would then have a scope of words and sentence-types suitable for the use of different academic levels.

5. Results of the Written Corpus

As shown in Table 2, the Written Corpus has a data size of 2,637,990 Chinese characters. These were segmented into 1,695,214 word-tokens consisting of 53,230 word-types. In order to facilitate curriculum development, frequent word-types that covered 95% of the data in the corpus were extracted to form a common wordlist of 10,071 word-types. These words are those frequently engaged by students daily in the Singapore context. Table 6 shows the distribution of these words in terms of their PoS.

Table 6. PoS Distribution of Common Words in Written Corpus

	Parts-of-speech, PoS	No. of word-types
1	Nouns[*]	4,441
2	Verbs	3,416
3	Adjectives	1,112
4	Adverbs	500
5	Quantifiers	181
6	Pronouns	112
7	Conjunctions	106
8	Prepositions	59
9	Numerals	37
10	Particles	32
11	Onomatopoeias	30
12	Interjections	24
13	Modals	21
	Total:	**10,071**

[*] Nouns include 3,978 common nouns, 151 time words, 68 positions words, 66 locative words, 149 names of places and 29 names of organisations.

As shown above, the common words are distributed across13 PoS: nouns, verbs, adjectives, adverbs, quantifiers, pronouns, conjunctions, prepositions, numerals, particles, onomatopoeias, interjections and modals. Among these PoS, content words (i.e. nouns, verbs, adjectives, numerals, quantifiers and pronouns) forms the

majority of the corpus, totalling 9,299 words and making up 92.33% of the words generated. On the other hand, function words are scarce, totalling 772 words and making up 7.67% of the wordlist.

To understand how the coverage of this Singapore CL wordlist compares with those of international standards, such as the wordlist for International Teaching of Chinese, henceforth ITC wordlist, (State Language Commission, 2010) and the core wordlist constructed by Xiao *et al.* (2009), hereafter referred to as Lancaster Wordlist, a high frequency wordlist that covered 85% of the Singapore Written Corpus was drawn to compare with the said wordlist. The results are as follows:

Table 7. Identical words found across three wordlists

Wordlist (total no. of word-types)	No. of identical word-types with SCCL Written Corpus Wordlist (5003)	% of identical word-types
ITC Level 1& 2 Wordlist (5343)	3,513	70.22%
Lancaster Wordlist (4622)	3,382	67.60%
Comparison across three wordlists	2064	41.26%

As shown in Table 7, the SCCL Written Corpus wordlist contained 5,003 high frequency word-types covering 85% of the Written Corpus. Amongst these word-types, 3,513 are found in the ITC wordlist and 3,382 in the Lancaster wordlist. In addition, 2,063 word-types are found to be identical across the three wordlists and these can be seen as the core vocabulary to be mastered by CL learners. They should be deemed essential to the learning of Mandarin Chinese and for effective communication in the languageby learners of different origins. Therefore, these words should be incorporated into the future Singapore CL curriculum to ensure a basic mastery of Mandarin Chinese for Singapore CL learners.

Apart from identical word-types found through comparing the said wordlists, unidentical or unique word-types also serve as a reference for curriculum development.

As seen in Table 8, a comparison of the three wordlists found 1,830 unidentical word-types between the SCCL Written Corpus wordlist and the ITC Level 1& 2 Wordlist, and 1,239 unidentical word-types between the SCCL Written Corpus wordlist and the Lancaster Wordlist, amounting to roughly one-quarter and one-third of the SCCL Written Corpus wordlist respectively. As the examples illustrated, most of these unidentical words are related to politics and economics or regional differences (such as words mainly used in China). It is no surprise that the SCCL Written Corpus does not include such words as the language data collected were based on students' daily encounters and students rarely engage issues of politics and economics in CL. Furthermore, the corpus is made up of data extracted

Table 8. Unidentical words found across three wordlists

Wordlist (total no. of word-types)	No. of unidentical word-types with SCCL Written Corpus Wordlist (5003)	% of unidentical word-types	Examples of unidentical word-types
ITC Level 1& 2 Wordlist (5343)	1,830	34.25%	Words specific to politics and economics, such as 条约、经贸、盗版、集资、缴纳
Lancaster Wordlist (4622)	1,239	26.81%	Words specific to politics and economics, such as 领土、政治家、证券、会计、海关 Words specific to China context such as 开发区、自治区、党委、大队、公安局、出租车、民警

from Singapore's newspaper and non-newspaper media where these regional specific words might not be used and as a consequence are not found on the SCCL Written Corpus wordlist. Although these unidentical words are rarely seen in the Singapore context, they can still serve as extended vocabulary for advance CL learners when developing Chinese curriculum for higher academic grades.

Apart from vocabulary, the Written Corpus was found to have 118,956 sentences in total, and these sentences were annotated with 191 sentence-types and sentence patterns for the computation and identification of common sentence-types and patterns students would engage on daily basis. The distribution of sentence types is as follows:

Table 9. Distribution of sentence-types in the Written Corpus

Sentences types	Freq.	% in Corpus (118,956)	Coverage*	% of coverage ($n = 3,613$)
Declarative	88,641	74.52	3,613	100.00
Interrogative	10,611	8.92	1,498	41.46
Exclamatory	16,979	14.27	1,691	46.80
Imperative	2,725	2.29	661	18.30

* Coverage in terms of the number of documents in the corpus.

As shown in Table 9, a majority of the sentences in the Written Corpus are declarative sentences (i.e. 74.52% of the total number of sentences). For the remainder, 8.92% are interrogative sentences, 14.27% are exclamatory sentences and 2.29% are imperative sentences. As for the document coverage of Written Corpus,

declarative sentences appear in most documents (100%), whereas interrogative sentences appear in 41.46% of the documents, and exclamatory sentences and imperative sentences appear in 46.80% and 18.30% of the documents respectively.

Other than sentence-types, the individual sentences were also tagged as Simple Sentence or Complex Sentence and annotated with the respective sentence patterns and clausal relations. Table 10 shows the distribution of simple sentence patterns in the Written Corpus, whereas Table 11 shows the Top 10 complex sentence patterns engaged by student frequently.

Table 10. Distribution of simple sentence patterns in the Written Corpus

Types of simple sentence pattern	Freq.[*]	% in Corpus (n = 118,956)	Coverage[**]	% of coverage (n = 3,613)
Verbal Predicate (动词性谓语句)	91,990	77.33	3,035	84.00
Subjectless (无主句)	15,498	13.03	2,355	67.95
Adjectival Predicate (形容词性谓语句)	9,261	7.79	1,036	28.67
Reductive (简略句)	7,390	6.21	461	12.76
Single-Word (独词句)	5,188	4.36	452	12.51
"Subject-Predicate" as Predicate (主谓谓语句)	1,909	1.61	159	4.40
Nominal Predicate (名词性谓语句)	828	0.70	331	9.16

[*] Frequency in terms of occurrence in the entire corpus.
[**] Coverage in terms of the number of documents in the corpus.

As shown in Table 10, the three simple sentence patterns most frequently encountered by the students are verbal predicate sentences, subjectless sentences and adjectival predicate sentences, making up 77.33%, 13.03% and 7.79% of the total number of sentences in the corpus respectively. It is worth noting that two out of these three sentence patterns have a higher coverage of documents in the Written Corpus as compared to the rest, i.e. verbal predicate ~84.00% and subjectless ~ 67.95%. This means that these two sentence patterns are more likely to be engaged by students in most daily written materials, and hence students may face fewer difficulties acquiring them. However, other simple sentence patterns, such as single-word, "subject-predicate" as predicate sentence and nominal predicate, only cover 4.36%, 1.61% and 0.70% of the sentences in the Written Corpus respectively. Apart from reinforcing the high frequency sentence patterns to ensure the mastery of basic CL proficiency among all students, CL curriculum development should also take into consideration the low frequency simple sentence patterns as they are less likely to be acquired by students due to its low occurrence in daily written materials engaged by them. Teachers should also create opportunities to explicitly

introduce these sentence patterns so that students can understand them in order to communicate with CL speakers beyond Singapore.

Table 11. The top10 complex sentence relation patterns in the Written Corpus

Types of complex sentence relations	Types of complex relation patterns	Freq.*	% in Corpus ($n = 118,956$)	Coverage**	% of coverage ($n = 1,274$)
Consecutive Relation (承接复句)	……就……	6,583	5.53	1,665	46.08
Coordinate Relation (并列复句)	……也……	4,188	3.52	1,362	37.70
Causal Relation (因果复句)	因为……所以……	3,711	3.12	1,27	35.34
Supposition Relation (假设复句)	如果……就……	1,889	1.59	825	22.83
Adversative Relation (转折复句)	虽然……但是……	1,195	1.05	567	15.69
Purposive Relation (目的复句)	为了……	1,182	0.99	638	17.66
Coordinate Relation (并列复句)	……并……	1,165	0.98	697	19.29
Coordinate Relation (并列复句)	……而……	1,075	0.90	629	17.41
Causal Relation (因果复句)	因……而……	923	0.78	589	16.30
Adversative Relation (转折复句)	……，可（是）……	923	0.78	401	11.10

* Frequency in terms of occurrence in the entire corpus.
** Coverage in terms of the number of documents in the corpus.

Table 11 illustrates the 10 common complex sentence patterns found in the Written Corpus. Among them, six complex sentence relations are commonly used, i.e. consecutive relation complex sentences, coordinate relation complex sentences, causal relation complex sentences, supposition relation complex sentences, adversative relation complex sentences and purposive relation complex sentences. Among the common clausal relations found in the Written Corpus, the three commonly used sentence patterns are consecutive complex pattern "……就……", coordinate complex pattern "……也……" and causal complex pattern "因为……所以……". They make up 5.53%, 3.52% and 3.12% of the total number of sentences in the corpus respectively. In terms of the coverage of texts in the Written Corpus, these three sentence patterns cover 46.08%, 37.70% and 35.34% respectively. Similar to the prior interpretation of the simple sentences in the corpus, students should face

fewer difficulties in acquiring these three complex sentence patterns as they can be found in many written text documents. Curriculum developers and teachers should focus on creating ample exposure opportunities for the less frequently used complex sentence patterns, such as causal complex pattern "*因……而……*", adversative complex pattern "*……，可（是）……*" and other sentences less frequently used and covered as compared to these two sentence patterns.

6. Results of the Spoken Corpus

The Spoken Corpus consists of 1,285,094 word-tokens and 8,368 word-types. These word-types contain 7,151 common words (i.e. words that can be found in Mandarin Chinese dictionaries), 164 idioms or frequent collocated phrases and 848 proper names (i.e. names of people and places). The following table illustrates the common words' distribution in terms of PoS.

Table 12. PoS distribution of common words in the Spoken Corpus

	Parts-of-speech, PoS	No. of word-types
1	Nouns[*] (名词)	3,602
2	Verbs (动词)	2062
3	Adjectives (形容词)	726
4	Adverbs (副词)	268
5	Quantifiers (量词)	135
6	Pronouns (代词)	94
7	Conjunctions (连词)	66
8	Prepositions (介词)	39
9	Numerals (数词)	30
10	Particles (助词)	21
11	Onomatopoeias (拟声词)	49
12	Interjections (感叹词)	32
13	Modals (语气词)	27
	Total:	**7,151**[**]

[*] Nouns includes 3,384 common nouns, 109 time words, 62 positions words, 47 locative words, but excludes names of places and names of organisations.

As shown in Table 12 above, the common words are distributed over 13 PoS, i.e. nouns, verbs, adjectives, adverbs, quantifiers, pronouns, conjunctions, prepositions, numerals, particles, onomatopoeias, interjections and modals. Among these

PoS, content words (i.e. nouns, verbs, adjectives, numerals, quantifiers and pronouns) are the majority in the corpus, having a total of 6,669 words, which make up 96.19% of the words generated. Function words are rather scarce, totalling 482 words and making up about 3.81% of the wordlist.

Table 13. Distribution of words used by different levels of students

			No. of words	%
Words usedby all 3 levels		–	2,355	33.01
Words used by 2 levels	P1 & P3	162	1,533	21.42
	P1 & P6	302		
	P3 & P6	1,069		
Words used by 1 level only	P1	395	3,263	45.57
	P3	753		
	P6	2,115		
Total			**7,151**	**100.00**

Among the common words in Table 13: 2,355 words are found to be used by students of all levels (i.e. Primary 1, 3 and 6). These words can be a good reference for the core or basic vocabulary that has to be mastered at lower primary levels (i.e. Primary 1 and 2). As for the words used by both Primary 1 and 3 students only, and words solely used by Primary 1, they can be a good reference for the core vocabulary to be learnt by mid-primary levels (i.e. Primary 3 and 4). Lastly, the words used by both Primary 3 and 6, and the words solely used by Primary 3 students will be a good reference for the core vocabulary of higher primary levels (i.e. Primary 5 and 6). With these words from the Spoken Corpus and together with the 10,071 common words found in the Written Corpus, curriculum developer

Table 14. PoS distribution of words from different region

	PoS	No. of word-types	Examples
1	Adjectives (形容词)	10	大力, 搞笑,假假
2	Adverbs	6	改次, 敢敢, 好彩
3	Interjections	1	哇咾
4	Nouns	151	巴刹,大耳窿, 拜二
5	Pronouns	3	几多, 么, 做么
6	Particles	1	来的
7	Verbs	22	摆美, 电头发,淹水
	Total	**194**	

will have a critical and reliable reference on the range of words to be covered in the curriculum for the different academic levels, especially the primary levels.

Besides common words, 194 unique words used in Singapore, Malaysia, Hong Kong and Taiwan were also identified from the students' speech in the Spoken Corpus. As illustrated in Table 14, these words are mainly nouns, verbs and adjectives frequently used among Singaporeans. Some of these words are of dialect origin but pronounced in Mandarin, e.g. 假假 (to pretend), 好彩 (being lucky/fortunate) and 几多 (how much). These words usually have a Mandarin equivalent. However, there are some words that are rooted in the local culture, such as 改次 (next time), 巴刹 (wet market) and 大耳窿 (loan shark). Curriculum developers will have to further decide if these words should be included in the textbooks as they are icons of identity, or they should be avoided in the textbooks due to their unstandardised nature as compared to common words in Mandarin.

Apart from vocabulary, the Spoken Corpus was also annotated with sentence types and sentence patterns. For the annotation of sentence types and sentence patterns, the conversational turn breaks between an interviewer and a student were adopted as clear-cut boundaries of a sentence for the one-to-one interviews. Whereas the completeness, or boundary of lengthy turns in the interviews and the students' utterances in the more spontaneous classroom talks and home talks was determined by intuitive judgement (for discussion on sentence boundaries, see Goh 2007: 32–33). Upon the identification of the sentences, the Spoken Corpus was found to have 299,394 sentences in total, and these sentences were annotated with 191 sentence-types and sentence patterns for the computation and identification of common sentence-types and patterns conceived by students who contributed the data. The distribution of sentence types is as follows:

Table 15. Distribution of sentence-types in the Spoken Corpus

Sentences types	Freq.	% in Corpus (299394)	Coverage[*]	% of coverage ($n = 1,274$)
Declarative	270877	90.48%	1270	99.69%
Interrogative	20684	6.91%	1160	91.05%
Exclamatory	6594	2.20%	440	34.54%
Imperative	1239	0.41%	495	38.85%

* Coverage in terms of the number of documents in the corpus.

As shown in Table 15, a majority of the sentences in the Spoken Corpus are declarative sentences (i.e. 90.48% of the total number of sentences). For the remainder, 6.91% are interrogative sentences, 2.20% are exclamatory sentences and 0.41% are imperative sentences. As for document coverage of Spoken Corpus, declarative sentences appear in most documents (99.69%), whereas interrogative sentences

appear in 91.05% of the documents, and exclamatory sentences and imperative sentences appear in 34.54% and 38.85% of the documents respectively.

Other than sentence-types, the sentences were also tagged as Simple Sentence or Complex Sentence and annotated with the respective sentence patterns and clausal relations, as in the case of Written Corpus illustrated above. Table 16 shows the distribution of simple sentence patterns in the Spoken Corpus, whereas Table 17 shows the Top 10 complex sentence patterns commonly produced by students in the corpus.

Table 16. Distribution of simple sentence patterns in the Spoken Corpus

Types of simple sentence pattern	Freq.*	% in Corpus ($n = 118,956$)	Coverage**	% of coverage ($n = 1,274$)
Verbal Predicate (动词性谓语句)	85463	28.55%	1245	97.72%
Reductive (简略句)	68258	22.80%	1237	97.10%
Single-Word (独词句)	29117	9.73%	1164	91.37%
Adjectival Predicate (形容词性谓语句)	1286	0.43%	576	45.21%
Nominal Predicate (名词性谓语句)	191	0.06%	158	12.40%
"Subject-Predicate" as Predicate (主谓谓语句)	69	0.02%	58	4.55%
Subjectless (无主句)	31	0.01%	29	2.28%

* Frequency in terms of occurrence in the entire corpus.
** Coverage in terms of the number of documents in the corpus.

As shown in Table 16, the three simple sentence patterns most frequently used by the students are verbal predicate sentences, reductive sentences and single-word sentences, making up 28.55%, 22.80% and 9.73% of the total sentences in the corpus respectively. It is worth noting that these three sentences patterns have a high coverage of documents in the Spoken Corpus, ranging from 91.37% to 97.72%. This means that most of the students who provided the data are generally able to produce these three sentence patterns. However, simple sentences patterns, such as nominal predicate, "subject-predicate" as predicate and subjectless only covered 12.40%, 4.55% and 2.28% of the sentences in the Spoken Corpus respectively. Apart from reinforcing the high-frequency sentence patterns to ensure the mastery of basic CL proficiency among all students, CL curriculum development should also take into consideration the low-frequency simple sentence patterns. Teachers should also look into ways to expose students to these sentence patterns, so that students could understand and communicate with CL speakers beyond Singapore.

Table 17. The top10 complex sentence relation patterns in the Spoken Corpus

Types of complex sentence relations	Types of complex relation patterns	Freq.[*]	% in Corpus (n = 118,956)	Coverage[**]	% of coverage (n = 1,274)
Causal Relation (因果复句)	因为……所以……	9424	3.15%	1042	81.79%
Supposition Relation (假设复句)	如果……就……	2299	0.77%	563	44.19%
Consecutive Relation (承接复句)	……，然后……	2916	0.97%	550	43.17%
Adversative Relation (转折复句)	……，可（是）……	1353	0.45%	483	37.91%
Coordinate Relation (并列复句)	有时（候）……有时（候）……	970	0.32%	385	30.22%
Coordinate Relation (并列复句)	……也……	456	0.15%	259	20.33%
Consecutive Relation (承接复句)	……就……	623	0.21%	208	16.33%
Supposition Relation (假设复句)	……的话，……	261	0.09%	135	10.60%
Causal Relation (因果复句)	之所以……是因为……	205	0.07%	123	9.65%
Coordinate Relation (并列复句)	……还……	134	0.04%	110	8.63%

* Frequency in terms of occurrence in the entire corpus.

** Coverage in terms of the number of documents in the corpus.

From Table 17, the complex sentence relations produced within the Top 10 common complex sentence patterns are causal relation complex sentences, supposition relation complex sentences, consecutive relation complex sentences, adversative relation complex sentences, and coordinate relation complex sentences. Among these common clausal relations produced by the students, the three commonly used sentence patterns are causal complex structure "因为……所以……", consecutive complex structure "……然后……" and supposition complex structure "如果……就……". They make up 3.15%, 0.97% and 0.77% of the total sentences in the Spoken Corpus respectively. With reference to the coverage of documents in the Spoken Corpus, these three sentence patterns cover 81.79%, 44.19% and 43.17% respectively. This implies that most students can easily produce these sentence patterns and it can be expected that most students are able to master these sentence patterns easily. Therefore, the sentence patterns with low frequency and coverage as illustrated in the above table (and the sentence patterns after this Top 10) will have to be elaborated on in the curriculum and teaching process.

7. Application of the corpora

As described above, the key deliverables of the Written Corpus and the Spoken Corpus are character lists, wordlists and sentence-type lists, which can serve as references for curriculum development. However, the application of corpus is not restricted to the generation of these lists. For language education, the constructed corpus can be further developed into teaching and learning resources to support data-driven teaching and data-driven learning whereby language patterns can be synthesised using the corpus' abundant examples. For this application, the abovementioned Written Corpus has been developed into the Singapore Chinese Language Teaching Resource Platform (SCLTRP). This platform consists of six modules, namely Full-text Query, Character Query, Word Query, Sentence-Type Query, Text Grader and Teachers' Forum.

Figure 2. The word query module of the SCLTRP

Among these modules, "Character Query", "Word Query" (see Figure 2) and "Sentence-type Query" are developed to provide teachers with basic information of characters, words and sentence-types (such as character structure, word definition, frequent word collocation, frequency of sentence-type etc.) and numerous examples from the corpus. With this information, teachers can prepare for lessons on characters, words and sentence-types with ease.

The platform also consists of two core modules, i.e. the "Text Grader" (see Figure 3) and "Full-Text Query". These core modules are mainly developed to assist teachers in their lesson preparation. The "Text Grader" determines the academic grade of supplementary reading materials provided by the teachers; it also highlights the characters and words beyond the academic grade's syllabus. "Full-Text Query" provides processed full-texts that teachers can readily use for their teaching of the language or as text passages for comprehension.

Lastly, this platformalso consists of a "Teachers' Forum" for teachers' discussion on topics regarding Chinese education. On this forum, teachers canshare their experience inthe teaching of individual characters or words. Through this

Figure 3. The text grader module of the SCLTRP

forum, a community of professionals can hence be built to enhance each other's pedagogical capacity.

8. Conclusion

This paper introduces two specialised corpora developed for CL education in Singapore, in view of its changing composite and abilities of CL learners. The analysis of these two corpora provides critical information on characters, words and sentence-types for the curriculum development and teaching of CL. The wordlists generated from the two corpora were compared against wordlists by the State Language Commission (2010) and Xiao et al. (2009), and 60–70% of the words are found to be common among the compared wordlists. This indicates that the vocabulary to be learnt by Singaporean CL learners shall not deviate tremendously from CL learning worldwide, whether it is CL as first language in China or CL as second language in Europe. This common vocabulary among the three lists shall also be seen as a benchmark for basic CL proficiency and serve as a reference for instructional materials and assessments. Besides providing corpora-based information for CL education, this paper highlighted the development of the Chinese Language Teaching Resource Platform, which provides important resources and practical tools for both teachers and students.

In short, unlike the L2 learner corpora that usually collect interlanguage data for the description and theorisation of second language acquisition (SLA), the two specialised corpora have a clear pedagogical focus which is in similar tune with Zhang and Tao (2018). Zhang and Tao (2018) believed that language corpus carries three pedagogical needs, i.e. to guide curriculum development; to promote data-driven learning (DDL); and to establish the benchmarks of students' language proficiency. Among these needs, the two specialised corpora fulfilled the first and second need. Firstly, the two specialised corpora were constructed with the objectives to inform curriculum development. With the delivery of the character lists, wordlists and sentence pattern list, curriculum developers will have a better localised reference when developing their CL textbooks. Secondly, with the development and launch of the Chinese Language Teaching Resource Platform, teachers could use the ample examples in the two corpora to conduct DDL, and guide students to discover language features. Though currently the two corpora have not go into the benchmarking of students' proficiency, the processed data shall easily act as resource for such development.

Like most studies, the introduced corpora have their limitations. For the Written Corpus, some language materials face the issue of "up-to-dateness", especially news materials. The current corpus has newspaper materials up to 2009,

which were almost 10 years ago. Though common words may not deviate greatly from the wordlists, peripheral words and timely expressions have to be constantly updated in the corpus. It is also noted that the Written Corpus was not genre-balanced, and this may affect the distribution and usage of lexical items and grammatical structures as pointed out by Tao (1999). As for the Spoken Corpus, the majority of the data are from the one-on-one interview, which are deemed less natural and may face restriction in the range of vocabulary elicited. As such, it may be essential to expand the corpus with more natural data.

Acknowledgements

References

Altenberg, Bengt and Marie Tapper. 1998. "The Use of Adverbial Connectors in Advanced Swedish Learners' Written English." In *Learner English on Computer*, ed. by Sylviane Granger, 80–93. Harlow: Longman.

Braun, Sabine. 2007. "Integrating Corpus Work into Secondary Education: From data-driven learning to needs-driven corpora." *ReCALL* 19/3: 307–328.

Chinese State Language Commission (CSLC). 2010. *The Graded Chinese Syllables, Characters and Words for the Application of Teaching Chinese to the Speakers of Other Languages*. Beijing: Beijing Language and Culture University Press.

Cook, Guy. 1998. "The Uses of Reality: A reply to Ronald Cater." *ELT Journal* 52/1: 57–64.

Gavioli, Laura, and Guy Aston. 2001. "Enriching Reality: Language corpora in language pedagogy." *ELT Journal* 55/3: 238–246.

Goh, Hock Huan. 2007. *Mandarin Competence of Chinese-English Bilingual Preschoolers: A Corpus-based Analysis of Singaporean Children's Speech*. Singapore: Springer Singapore.

Granger, S., Hung, J. & Petch-Tyson, S. (Eds.) 2002. *Computer learner corpora, second language acquisition, and foreign language teaching*. Amsterdam: John Benjamins.

Hadley, Gregory. 2002. "Sensing the Winds of Change: An introduction to data-driven learning." *RELC Journal* 33/2: 99–124.

Huang, Fuju. and Deteng Meng. 2013. "Yuliaoku Shujuqudong Moshi zai Duiwaihanyu Jiaoxue zhong de Yingyong [Application of Corpus-Driven Model in Teaching Chinese as a Second Language]." *Zhongguo Jiaoyu Xinxihua* (The Chinese Journal of ICT in Education) 24:39–41.

Johns, Tim F. 1991. "Should You be Persuaded: Two Examples of Data-driven Learning." In *Classroom Concordancing*, ed. by Tim F. Johns and Philip King, 1–13. Birmingham: ELR.

Kettemann, Bernhard. 1995. "On the Use of Concordancing in ELT." *TELL & CALL* 4. 4–15.

Leech, G. 1997. "Teaching and language corpora: A convergence." In A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (Eds.), *Teaching and language corpora*, 1–23). London: Longman.

Li, Charles N. and Sandra A. Thompson. 1981. *Mandarin Chinese: a functional reference grammar*. Berkeley: University of California Press.

Lin, Jinzhan, Chunsheng Zhao, Swee Choon Ang, Hock Huan Goh, and Chee How Wong. 2015. "Ji yuyuliaoku de Xinjiapo hua wen jiaoxuezi yuan ping tai kai fa [Development of a Corpus-based Resource Platform for Chinese Language Teaching in Singapore]." *Hua wen xuekan* [Journal of Chinese Language Education], 13(1):1–15.

Liu, Eric Shen. 1973. *Frequency Dictionary of Chinese Words*. The Hague: Mouton.

Lu, Jiping. 2003. *Hanyu Yufa Jichu* [Basics of Mandarin Chinese Grammar]. Beijing: The Commercial Press.

McEnery, Tony and Richard Xiao. 2010. "What Corpora Can Offer in Language Teaching and Learning." In *Handbook of Research in Second Language Teaching and Learning*, ed. by Eli Hinkel, Vol. 2, 364–380. London & New York: Routledge.

Meunier, Fanny. 2011. "Corpus Linguistics and Second/Foreign Language Learning: Exploring multiple paths." *RevistaBrasileira de Linguística Aplicada*. 11/2, 459–477.

Meyer, Charles. 2002. *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge University Press.

MTLRC (Mother Tongue Language Review Committee). 2010. *Nurturing active learners and proficient users*. Singapore: Ministry of Education. https://www.moe.gov.sg/media/press/files/2011/mtl-review-report-2010.pdf

Römer, Ute. 2006. "Pedagogical Applications of Corpora: Some Reflections on the Current Scope and a Wish List for Future Developments." In *The Scope and Limits of Corpus Linguistics – Empiricism in the Description and Analysis of English*, ed. by Volker Gast, Special Issue: Zeitschriftfür Anglistik und Amerikanistik, 54/2, 121–134.

Smith, Simon. 2011. "Corpus-based Tasks for Learning Chinese: A Data-driven Approach." (paper presented at The Asian Conference on Technology in the Classroom Official Conference Proceedings 2011. 48–59).

Sinclair, J. (Ed.) 2004. *How to use corpora in language teaching*. Amsterdam: John Benjamins.

Standards Administration of China. 1992. *Contemporary Chinese language word segmentation specification for information processing*. (website: http://www.sac.gov.cn/was5/web/search?channelid=97779&templet=gjcxjg_detail.jsp&searchword=STANDARD_CODE=%27GB/T%2013715-1992%27&XZ=T&STANDARD_CODE=GB/T%2013715-1992, as of 2017 June)

Standards Administration of China. 2006. Standards of POS tags of Contemporary Chinese for CIP.

Tao, Hongyin. 1999. Discourse taxonomies and their grammatico-theoretical Implications. *Dangdai Yuyanxue* [Contemporary Linguistics], 1.3:15–24.

Tao, Hongyin. 2015. Profiling the Mandarin Spoken Vocabulary Based on Corpora. In Wang, William S-Y. and Chaofen Sun (eds.). *The Oxford handbook of Chinese linguistics*, 336–347. Oxford: Oxford University Press.

Xiao, Richard, Paul Rayson, and Tony McEnery. 2009. *A Frequency Dictionary of Mandarin Chinese: Core Vocabulary for Learners*. London & New York: Routledge.

Yu, Guoliang. 2009. *Yuliaokuyuyanxue de Yanjiu yu Yingyong* [Research and Application of Corpus linguistics]. Chengdu: Sichuan University Press.

Zhang, Jie & Hongyin Tao. 2018. Corpus-based research in Chinese as a second language. In Ke, Chuanren (ed) *The Routledge handbook of Chinese second language acquisition*. London: Routledge.

Zhao, Yuanruen. 1979. *Hanyu Kouyu Yufa* [Grammar of Spoken Mandarin Chinese). Beijing: The Commercial Press.

Appendix A. List of Parts-of-Speech, sub-types and examples

As mentioned in Section 4, the two corpora took reference from the Lingjoin platform for the definition of word boundaries. The segmentation of words and its PoS tagging of this platform is in accordance with the "Chinese Language Part-of-Speech Annotation Guideline for Computing Technology Lab" developed by the Computing Technology Lab of China Science Academy. The platform differentiated the following PoS and its sub-types of words:

Parts-of Speech	Sub-types	Tag	Examples
Nouns	Common Nouns	N	人；时间；学校
	Time Words	T	现在；今天；后来
	Positions words	F	上；里；之前
	Locatives	S	家里；心里；身上
Verbs		V	有；说；知道
Adjectives		A	好；大；重要
Adverbs		D	不；也；非常
Quantifiers		Q	个；次；位
Pronouns		R	我；这；自己
Conjunctions		C	和；但是；而
Prepositions		P	在；把；从
Numerals		M	一；两；几
Particles		U	的；了；着
Onomatopoeias		O	哈；呜；嘿嘿
感叹词		E	Interjections
语气词		Y	Modals

Apart from the above PoS and their sub-types, the Lingjoin platform also differentiated the following types of words:

Idioms and Chunks	L	接下来；不好意思；不可思议
Prefix	H	第；非；准
suffix	K	们；儿；者

Appendix B. List of sentence types and sentence patterns

I.   Sentence Types (句类)

Types	Example
Declarative 陈述句	亚洲人每人每年平均消耗150公斤的米。[61]
Interrogative 疑问句	新加坡有哪些著名的土生华人？[18]
Exclamatory 感叹句	甜品真好吃啊！[1]
Imperative 祈使句	时间宝贵，马上开始行动吧！[24]

II.   Sentence Patterns (句型)

Types	Example
Simple Sentence 单句	
Subject-Predicate Sentence 主谓句_(动词性谓语句)	信鸽能够感应到地球的磁场。[596]
Non-Subject-Predicate Sentence 非主谓句_(独词句)	好消息！[52]
Complex Sentence 复句	
Progressive Relation 递进复句	电梯是一种了不起的发明，不只省时，而且省力，又大众化，会提醒你楼层，还会说"小心，门关了"。[126]
Adversative Relation 转折复句	虽然情况已经好转，但是大家还是不要掉以轻心。[3]
Coordinative Relation 并列复句	我平日喜欢阅读科学类的新闻，对科学知识也很感兴趣。[26]
Conditional Relation 条件复句	患了手足口症，也不要害怕，只要按时吃药，很快就会好的。[10]
Causal Relation 因果复句	因为现代人越来越讲究健康生活，所以吃甜品的人也减少了。[40]

Types	Example
Supposition Relation 假设复句	除非你的运气很好，要不然乘电梯往往需要一种坚持的"忍耐力"。[75]
Concessive Relation 让步复句	她心肠好，即使是陌生人，也会尽力帮他们。[959]
Consecutive Relation 承接复句	奶奶在纸上画了几笔，然后要我顺着图形剪。[6]
Alternative Relation 选择复句	他不是骂我就是打我，把我的自尊心全都伤害了。[1158]
Purpose Relation 目的复句	一天，一个贫穷的男孩为了赚取学费，正挨家挨户地推销商品。[854]

Index

This volume features a discourse empirical orientation from diverse perspectives and various methodologies, in which narratives, interviews, surveys, and large-scale databases or self-created written and spoken corpora are employed and analyzed to gain a better understanding of new developments and changes in Chinese language and discourse. Authors employ updated approaches from a variety of fields, including applied linguistics, functional linguistics, corpus linguistics and sociolinguistics, to describe the structure of Chinese language and discourse and to examine its critical issues, many focusing on globalization-induced language developments and changes. With an empirically-based discourse/socio-cultural approach, this collection makes valuable contributions to research on Chinese language and discourse and serves as a sound reference for Chinese researchers and educators in diverse fields such as Chinese language and discourse, Chinese linguistics and language education, Chinese multiculturalism, and more.

John Benjamins Publishing Company