Use of Formulaic Sequences in Task-based Oral Production of Chinese

A Comparative Study of

Native and Non-native Speaker's Outputs

First Supervisor: Prof. Anwei Feng
Second Supervisor: Dr. Sue Beverton

Ho-Put Wong
MA, BBA

A thesis submitted on partial fulfillment of the requirement for the degree of Doctor of Education

School of Education
University of Durham

Abstract

Formulaic language is a long-recognized phenomenon that has inspired new pedagogical and lexicographic developments in ESL. However, it did not draw much attention until the last decade in the field of Chinese-as-a-second-language (CSL). Not surprisingly, research of formulaic sequences (FSs) in CSL has been scarce and rarely corpus-based and pedagogically focused. This study aims at filling in part of this gap.

It is widely accepted that FSs are ubiquitous and play an important role in any discourse. They are critical in acquisition, retention and production of both L1 and L2. The awareness of both learners and teachers of the importance of FSs still needs to be raised. The aim of this thesis is to investigate disparities in the use of FSs in spoken Putonghua by non-native speakers (NNSs) and native speakers (NSs) to inform learning and teaching of CSL.

This research adopts a corpus-based approach. After completing a language task, all FSs, including Task-specific Sentence Stems (i.e. function-specific utterances to carry out certain speech acts), in the transcripts of 30 NNSs and 30 NSs were identified and compared. Differences between NS and NNS data are pinpointed for drawing pedagogical inferences. It was confirmed that NS data contains higher density of FSs in general (i.e. more number of characters inside FSs as percentage of total number of characters) and greater number and varieties of TSSSs than NNSs. It was also confirm that more advanced NNSs' data is more formulaic than less advanced NNSs', and more native-like data is more formulaic than less native-like data. All the above findings attest to the significance of FSs in CSL.

With better knowledge of the formulaic nature of Chinese language, this study calls forth more lexical and task-based approaches in CSL, and might shed light on curriculum and syllabus design, teaching material development and pedagogy.

Acknowledgements

I am greatly indebted to Prof. Anwei Feng, my principal supervisor, for all his valuable guidance, patience, criticism, and encouragement throughout.

I would also like to thank Dr. Susan Beverton, Dr. Richard Remedios, Prof. Carl Bagley and administrative staff of School of Education for their professional and logistical support.

A special word of thanks is due to Dr. Weiping Wu for making the COPA NNS copus and COPA devices available, and for inspiring me to embark on this research.

I am grateful to my colleagues for their insightful advices, kind assistance in data collection, transcription, identification of FSs and error correction, and for sharing my administrative workloads. I am also grateful to Dr. Raymond Yeung and Mr. Larry Wong for their huge support in data analysis.

Most importantly, I wish to thank my wife, Caroline, my kids and my parents for their love and support, and the sacrifices they made in the last 11 years.

This study would not have been possible without the grace of my heavenly Father. Glory to Him.

Statement of Copyright

Table of Content

List of Tables, Figures and Appendices

Glossary of Acronyms

ACTFL:	American Council on the Teaching of Foreign Languages
CCL:	Center for Chinese Linguistics (Peking University)
COPA:	Computerized Oral Proficiency Assessment
CSL:	Chinese as a Second Language
FS:	formulaic sequence
LK:	Mistaken but likely non-TSSS FSs
MEU:	morpheme equivalent unit
NL-T:	Native-like and Typical utterance
NL-UT:	Native-like but Untypical utterance
NNL:	Nonnative-like utterance
NNL-C:	Nonnative-like but Comprehensible utterances.
NNL-I:	Nonnative-like and Incomplete utterances
NNS (A):	non-native participant who obtained Advanced certificates in COPA
NNS (I):	non-native participant who obtained Intermediate certificates in COPA
NNS:	non-native speaker
NS:	native speaker
TSSS:	Task-specific Sentence Stem
ULK:	Mistaken and unlikely non-TSSS FSs

Chapter I INTRODUCTION

This chapter will give a brief account of

- the definition and importance of formulaic sequences
- the major difficulties in the study of formulaic sequences
- the purposes and significance of this study

Foreign or second language teaching inevitably involves breaking down the language into teachable units. But what are the most efficient or appropriate units in teaching and learning? Traditionally single-word lexis and grammar rules have been the most salient and intuitively sound units. However, there are many occasions in which words and grammar rules alone cannot provide satisfactory answers. Below are some of the questions that once came across the author's mind in the process of learning English as a foreign language.

- Why do we say *blonde hair* but not *blonde fur, black and white* but not *white and black, If I were you* but not *If I am you* or *if I was you*, and *half past one* but not *half to two*?

- How come *state-of-the-art* is often used to talk about technology etc and doesn't seem to have anything to do with arts?

- What are the grammar rules underlying *of course, by and by* and *by and large,* and how can the meaning of the whole be derived from the meaning of the constituents?

- Why can't we put both *because* and *so* into the same sentence to make something like *Because the weather in fine, so we go out for a walk*? It sounds perfectly logical: one indicating the cause while the other the effect!

- Why do soldiers typically use *Do you read me?* and *Do you copy?* to check if their verbal messages are well received when using walkie-talkies? Why not *Do you hear me?* or *Did/Can you hear me?* which sound more logical?

- Why *avenge me* and not *avenge for me*? To the author, the former sounds like the *me* is to be punished while the latter sounds more like someone is to be punished for hurting the *me*.

- When people greet you with How do you do?, how come you don't answer and just ask them back How do you do? without expecting an answer? Isn't it a question?

Similar phenomena difficult to be explained with semantic or syntactic rules abound in Chinese as well, and as a CSL teacher, the author has heard remarks ranging from 'That's not logical!', 'That sounds a bit rude' to 'That's weird' or even 'How can Chinese be so stupid?' from frustrated adults students who tend to analyze and break down target language word strings. Listed below are but a few of their questions:

- How can *养病* (literally meaning 'to grow/foster/cultivate an illness') mean 'convalesce' and *恢复疲劳* (lit. 'to restore tiredness') mean 'take away tiredness'? Do you want to be more seriously ill and tired when you *养病* and *恢复疲劳*?

- How can *小偷儿抓住了* (lit. 'the thief has caught') and *小偷儿被抓住了* (lit. 'the thief was caught') be synonymous and mean 'the thief was caught'? One is active and the other one is passive!

- How come *好容易才买到了这本书* is synonymous with *好不容易才买到了这本书*? Both sentences mean 'I just managed to get hold of a copy of this book' while the latter has an additional *不* (= not). Doesn't the *不* carry any weight?

- Why are both *好久不见了!* (lit. 'long time don't see') and *好久没见了!* (lit. 'long time didn't see') acceptable to mean 'I haven't seen you for ages'? The former doesn't seem right grammatically!

- Why do Chinese shopkeepers acknowledge their customers with *你买点儿什么?* (= what do you want to buy?)? Isn't that too straight forward?

- How can the constituents of *他真有两下子!* (lit. 'He really has two units (of actions)') be combined to mean 'He obviously knows his stuff'?

While experienced frontline teachers can definitely manage to find or make up brilliant and (seemingly) convincing answers to the above questions, some researchers speculate that those word strings should be classified as a distinct type of linguistic units other than single-word lexis and grammar rules. These word strings might be able to serve as more sensible, practical and efficient units in language teaching and learning. The cover term of these units is Formulaic Sequences (FSs) and they are the focus of this study.

I.1 Definition and Importance of Formulaic Sequences

A formulaic sequence (FS) refers to a word string which 'is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar' (Wray 2002:9)[1]. In other words, Formulaic sequences (FSs) are multi-word expressions behaving like single-word lexis[2]. From the perspective of Universal

[1] Most of the literatures reviewed in the study have different definitions and use different terms, with different inclusions and exclusions when the researches are being conducted. However, this does not impair the value of studies on formulaic language as an undeniable phenomenon. 'Formulaic sequence' proposed by Wray (2002:9) is a rather loose and all-encompassing term and its definition is borrowed as the working definition in this study. For a slightly stricter definition see Gries (2008:6)'s 'phraseologism'.

[2] Although FSs behave like single-word lexis when stored and retrieved, many of them are analyzable and decomposable and not completely fixed lexically, syntactically and semantically (Gibbs 2007:721). For example, among the components of *it is not in his nature to …*, only 'it', 'in', 'nature' and 'to' are fixed, while 'is' can vary to 'was', 'not' can be replaced by 'hardly' or

Grammar, FSs are language-specific peripheral features rather than universal features, and are challenging to learn (Shortall 1996:38).

FSs have been labeled variously as amalgams, clichés, collocations, fixed expressions, gambits, holophrases, idioms, multiword units, non-compositional sequences and prefabricated routines etc by different researchers (Gibbs 2007:698). In Gibbs (1994), English FSs include sayings (e.g. *let the cat out of the bag*), proverbs (e.g. *a stitch in time saves nine*), phrasal verbs (e.g. *to give in*), idioms (e.g. *kick the bucket*), binomials (e.g. *spick and span*), frozen similes (e.g. *as white as snow*), phrasal compounds (e.g. *red herring*), incorporated verb idioms (e.g. *to baby-sit*) and formulaic expressions (e.g. *at first sight* and *how do you do*?). In the field of foreign language teaching, these multi-word expressions, together with single-word lexis, constitute vocabulary (Boer and Lindstromberg, 2008a:4; Moon 1998:51; Richard and Rodgers 2001:227)[3] (also called lexemes or lexical units or lexical items (Schmitt 2000:1-2)).

In spite of 'being central to language and of critical importance to the typical language learner' (Zimmerman 1997:5), vocabulary has traditionally been viewed as 'the means of exemplifying other features of the language' (Sinclair and Renouf 1988:142), thus undervalued in the field of second language acquisition through its various stages, from very traditional Grammar Translation Method to newer Communicative Approach or Natural Approach[4], with limited exceptions like Lexical Approach proposed by Michael Lewis (Zimmerman 1997:16-7). Coady (1997) comments that language teachers and scholars tend to feel that, compared with grammar, words are less challenging to learn, and can be learned naturally

---

'scarcely', and 'his' can be replaced by other possessive pronouns (Sinclair 1991:111).

[3] FSs are also defined by scholars like Lewis (1993; 1997a; 1997b) as multiword lexical items between grammar as the generative system and vocabulary as non-generative items. The 'vocabulary' used by Lewis here refers to single-word lexis.

[4] Compared with vocabulary, syntax and phonology had been prioritized as 'more serious candidates for theorizing' in Structuralism and Chomskyan school of linguistics (Richards 1976). Besides, as observed by Sinclair and Renouf (1988), it is 'exceptionally difficult to teach an organized syllabus of both grammar and lexis at the same time' (p143).

from extensive exposure, and 'teaching vocabulary is a low-level intellectual activity unworthy of their full attention' (p.274). While the above observations concern mainly single-word lexis, they hold true even more for FSs, because compared with single-word lexis, FSs have traditionally received even less attention (Boers and Lindstromberg 2008a:7).

However, as revealed in academic works in the last three decades, FSs are in fact indispensable to human languages because our language system is not exclusively based on rules (Gibbs 2007:697; Sinclair 1991; Skehan 1998)[5]. Although FSs might appear to be analyzable into smaller parts, they 'constitute single choices' and are huge in number (Sinclair 1991:110). They are memory-based (Skehan 1998:60) and tend to be context-bound (ibid:89). They have been found to constitute a substantial proportion of any English discourses (Schmitt and Carter 2004:1) and famous English corpuses (Moon 1998; Altenberg 1998). FSs are found to be better manifestations of cultural connotations more than single-word lexis (Teliya et al 1998:59). Some FSs are believed to be as tightly linked to the cognitive schemata we have formed about something as to our institutionalized cultural facts (Gerbig and Shek, 2007:319), and some are found important in helping people with Alzheimer's disease to maintain their roles in conversations (Maclagan et al 2008:185). Average mature native speakers know far more morphologically complex lexical items than single morpheme lexical items (Pawley and Syder 1983; Mel'cuk 1998). FSs are stored and reused as much as, if not more than, those word strings generated from scratch (Cowie, 1988:136), and are of equal, if not greater, significance as single lexical items[6] (Jackendoff 1995; Mel'cuk 1995).

Besides being typically stored and processed as holistic units (Schmitt and Carter 2004), FSs are found to be processed more quickly[7] (Conklin 2008; Ellis 2008:6)

[5] Sinclair (1991) advocates a dual-mode Open choice principle and Idiom principle. Skehan (1998:53-5) also suggests that language is both rule-based and exemplar-based, and learning a language involves learning both rules and exemplars.

[6] 'Single-word lexis', 'single morpheme lexical items' and 'single lexical items' etc. are adopted by different researchers to refer to the same concept. They are used interchangeably in this study.

[7] Underwood et al (2004) find that in reading tasks, the same words are processed more quickly when they are in FSs than when in non-formulaic strings (p167). On the contrary, again in reading

and spoken with better form, greater appropriateness, greater fluency[8] and more coherent intonation contour than word strings generated from scratch (Peters 1983; Nattinger and DeCarrico 1992; Wray 2002; Wood 2010a; Kuiper and Austin 1990:196; Aijmer 1996:9). They play a major part in spontaneous interactions and help speakers plan for the form and content of future utterances by freeing the processing resources during communication (Altenberg and Eeg-Olofsson 1990:2; Skehan 1998:3 & 89). Besides, knowledge of FSs is believed to help speakers of a particular language select the best expressions among all possible ones, many of which are grammatically correct but just not preferred by native speakers or highly marked[9] (Pawley and Syder 1983:192-199). FSs are also found to be important for second or foreign language learning in empirical research studies (Clark 1974; Peters 1983; Wong Fillmore 1976) and have been viewed as the core elements in developing new approaches of teaching English as a second language (Willis 1990; Nattinger and DeCarrico 1992; Lewis 1993; 1997a; 1997b)[10]. Kjellmer (1991) attributes learners' non-nativelike speech output to their poor automation of FSs (p124)[11].

I.2 Major Difficulties in the Study of Formulaic Sequences

FSs are not limited to traditionally recognized multiword units obviously operating as single units such as idioms, proverbs and sayings (Biber et al 1999). Due to their huge diversity in terms of length, structure, purpose and fixedness, the

---

tasks, for nonnative speakers, it takes more time to process an unknown formulaic word strings than a non-formulaic word string (Schmitt and Underwood 2004:186).

[8] For example, Kuiper (1996) finds that commentaries in fast sports like horse-racing contain more FSs than those in slow sports like cricket, and auctioneers' speech in high-pace auctions contain more FSs than low-pace ones.

[9] For instance, *It's twenty to six* is conventional and preferred even though *It's six less twenty*, *It's two thirds past five*, and *It's forty past five* etc. are syntactically correct (Pawley and Sider 1983:197-8). Likewise, it is ordinary and idiomatic for a lover to say *I want to marry you* but not *What is desired by me is to wed you* or *I, who am speaking, want to marry you, whom I am addressing* etc, even though they are grammatically possible (ibid:196). Levinson (2000:23) also provide some good examples.

[10] Dörnyei et al (2004) maintains that a mastery of a wide range of FSs help L2 learners get rid of their artificial textbook-like language and sound native-like (p87).

[11] Kjellmer (1991) maintains that a typical moderately fluent learner's 'building material is individual bricks rather than prefabricated sections' (p124).

development of a comprehensive definition of FSs is extraordinarily challenging and remains a major problem in the area (Schmitt and Carter 2004:2-3). Yet over fifty terms of different degrees of overlappedness, e.g. complex lexemes, formulae, lexical phrases, preassembled speech, prefabricated routines and patterns, ready-made utterances, stereotyped phrases, and unanalyzed chunks of speech, have been adopted by different researchers, in accordance with their respective purpose and focus of study (Wray 2002:9; also see Gibbs 2007 quoted above). Categorization of FSs is also lack of consensus. Over twenty taxonomies of English FSs have been proposed to give clear and organized accounts of FSs but many are internally inconsistent (Wray 2002:Chp 3) and incomprehensive. Moreover, in order to judge whether a certain word string is truly a FS, researchers have made use of native speaker intuition, frequency counts, structure or form, and phonological features, etc, but, as can be expected, each approach has its pros and cons and, when used in isolation, is often criticized to be either unreliable or unable to capture the whole picture (Wray 2002:43). As Wray (2002) puts it：

> ‘(R)esearch on formulaic language has lacked a clear and unified direction, and has been diverse in its methods and assumptions. Both within and across subfields such as child language, language pathology and applied linguistics, different terms have been used for the same thing, the same term for different things, and entirely different starting places have been taken for identifying formulaic language within data. As a result, little headway has been made in spotting larger, more general patterns, and no attempt has been made before, to compare and contrast the full range of findings and to reconcile them within a single theoretical account.’ (p4-5).

I.3 The Purposes and Significance of This Study

Formulaic language is not a unique phenomenon in English. Nattinger and DeCarrico (1992) demonstrate with ample examples that formulaic sequences ‘exist in the same abundance’, ‘perform the same conversational functions’, and ‘occur with strikingly similar characteristics and category types’ in other

languages, including Chinese (p66-68, 190-194). Mel'cuk (1995) also asserts that FSs are 'numerically predominant lexical units' in any language (p24). However, research in FS did not draw much of researchers' attention in the field of teaching or learning Chinese as a second language (CSL) until very recently, and there has been limited empirical research (Wang 2007:7-8; Wu et al 2009:2); . This study is aimed at shedding some light on the nature, identification and function of Chinese FSs, especially the interactional FSs, and their application in CSL in lexical and task-based teaching.

Words inside FSs as percentage of total words in speech samples of native speakers (NS) of English has been found higher than that of non-native speakers (NNS) (Foster 2001:85). NNSs are also found to rely on a smaller number of FSs (ibid:87). This study attempts to look at the similarities and disparities between NSs and NNSs of Chinese in their respective use of FSs, in order to deepen our understanding on how FSs facilitate production and learning of Chinese as a foreign or second language. Quantity, quality and choices of FSs employed by NSs and NNSs are compared and contrasted to yield pedagogical implications.

FSs are as important to CSL as to the teaching and learning of other languages. By demonstrating the formulaic nature of Chinese, this study advocates a more lexical and task-based communicative approach in CSL in which curriculum and syllabus design, teaching material development and pedagogy are treated more formulaically.

Chapter II LITERATURE REVIEW

This chapter will review academic works concerning

- the significance of formulaic sequences (FSs) in general
- the significance of FSs in L1 and L2 acquisition (specifically in ESL and CSL)
- some representative taxonomies of FSs
- the identification of FSs
- the FSs that can directly contribute to task-based CSL: the Interactive FSs

II.1 Research Studies on Formulaic Sequences Before 1970s

Formulaic language has long been a field of study cultivated by scholars in disciplines such as literary studies, folklore studies, social anthropology, neurology, experimental psychology and educational psychology etc (Pawley 2007:1-9). It also drew the attention of many influential linguists from early to mid 20th century. de Saussure (1916/1966) mentioned about 'cluster of signs' becoming 'a simple unit' when examining the phenomenon of agglutination (p177), Jespersen (1924) contrasted 'formulas' that can be whole sentences or groups of words with free expressions (p18), Bloomfield (1933) found 'forms' lying 'between words and phrases' (p181), Firth (1964) had a section on 'holophrase' referring to whole sentences used in actual speech (p82-3), and Hymes (1962) noticed 'linguistic routines' making up 'a vast proportion of verbal behavior' ranging from the numeral 1 to 10, ABC's and limericks to antiphonal sequences in games and ceremonies, exchanges of greetings and expressions with directive or expressive functions (p41-2) (see Wray 2002 for a detailed review). However, before 1970s, formulaic language was largely marginalized due to the prominence of Chomsky's approach to syntactic structure (Wray 2002:8; Ellis 2008:3), the Chomskian avoidance of 'engagement with what people actually say' (Wray 2002:13), the

tradition to divide language dichotomously into grammar as the generative system and vocabulary as non-generative items and neglect most of the multiword lexical items in between (Lewis 1993; 1997a; 1997b; Sinclair 2008a:407 & 2008b:xv; Gries 2008a:10-11), the dismissing of 'the relations between the generative system and the socio-linguistic habits regulating how the system is put to use' (Coulmas 1981:1), the false impression that formulaic language only constitutes a very small part of natural language and the common practice of theoretical linguists to shunt off to one side the bits that did not fit their monolithic model (Pawley 2007:11). Lewis asserts that the deficient practice to ignore FSs in analyzing a language can be paralleled with analyzing English by syllables rather than by words, which are larger and more powerful units (Lewis 1993:104).

II.2 Research Studies on Formulaic Sequences Since 1970s

With 1970s as a water shed, interest in formulaic language has vastly increased (Pawley 2007:11) and related research literature in Anglophone has been growing steadily for over three decades (Schmitt et al 2004:55), along with the following trends or changes:

- The traditional view to divide language into grammar and vocabulary received challenges from both linguistic and pedagogical point of views (Lewis 1997a:43; 1997b:255; 2000:166; Nattinger & DeCarrico 1992; Willis 1990; Hunston 2003:31; Adolphs 2008:135).

- Changed role of vocabulary:

    - On the one hand, a steadily growing amount of research on vocabulary (Schmitt and Carter 2004:11), and

    - On the other hand, the realization that '(w)ords mean things in the context of other words (Ellis 2008:1) and 'single words are not necessarily the appropriate unit for lexical description' (Gibbs 2007:698). As evidence shows that much of our language is stored in units larger than individual words, 'over-concentration on learning single words may

hinder the development of the L2 phrasal lexicon and deny the opportunities this gives for rapid retrieval and fluent connected speech in the stressful conditions of speaking and writing' (McCarthy, 1990).

- Demoted role of grammar:

  - The recognition of the lexical nature of language (Lewis 1997a:16; Skehan 1996:21), and 'faultlessly grammatical (sentences) can be rude, odd and comical' (Foster 2001:76).

  - The recognition that language is not as rule-based as we used to think. Many semantically related words behave significantly differently that 'no semantic generalization can help' the learners (Wilkins 1976:49)[12], and '(e)very word has its own grammar' (Lewis 2000:166).

  - Similarly, the realization that communicative competence is not a matter of knowing and being able to employ grammatical rules, but 'a matter of knowing a stock of partially pre-assembled patterns, formulaic frameworks, and a kit of rules, so to speak, and being able to apply the rules to make whatever adjustments are necessary according to contextual demands' (Widdowson 1989:135). Empirical studies on language use conducted in many native-speaking speech communities including law courts, auction rooms, army mess halls and academic common rooms, etc, reveal that sector-specific stocks of patterns and frameworks abound in all societies and might take years of time even for native speakers to fully master them as it involves immense learning of the non-linguistic cultural knowledge in those speech communities (Kuiper 2004:51)[13].

  - The realization that in early stages first and second language acquisition

[12] For instance, *allow, permit, approve* and *agree* all take different range of complement structures (Wilkins 1976:48), and while *in deep trouble* is socially endorsed, **in shallow trouble* is not (Lewis 1996:10).

[13] Also see Swales (1990) on FSs in different genres of English for Academic Purposes and English for Special Purposes.

learners are able to communicate in meaningful social settings with formulaic expressions before grammar rules to generate such expressions are acquired (Clark 1974; Wong-Fillmore 1976; Peters 1983).

- The recognition that in L1 discourse, words co-occur 'in ways and to a very high degree inexplicable within a syntactocentric, generative theory' (Boers and Lindstromberg 2008a:7), thanks to the overwhelming evidence provided by studies in corpus linguistics such as Sinclair (1991) and Carter and McCarthy (1997).

- The change of status of grammar from being in a 'pride of place' in traditional syllabus to being 'largely ignored' in 'a large group of notional, functional and communicative syllabuses' (Sinclair and Renouf 1988:141; Shortall 1996:31). Pedagogical grammar has been advocated to adopt a lexical approach (Little 1994).

II.3 Significance of Formulaic Sequences

As multi-word lexical phenomena between traditional poles of lexicon and syntax (Nattinger and DeCarrico, 1992:1; Lewis, 1997:43), FSs are wide spread in English native speakers' language (McCarthy, 1991:122) and in all other languages (Mel'cuk 1995:24). On the basis of substantially different purpose, scope and criteria, FSs are found to make up 4% to 5% of Oxford Hector Pilot Corpus in Moon (1998:57), 25.08 to 32.29% of native speakers' oral production of English and 16.87 to 17.23% of non-native speakers' in Foster (2001), 58.6% of spoken English and 52.3% of written English in Erman and Warren (2000), to just name a few. Altenberg (1998:102) estimates that over 80% of London-Lund Corpus of Spoken English is made up of recurrent word-combinations, though not all these recurrent word-combinations are FSs. As a whole, FSs are worthy of exploiting in language teaching (Nattinger and DeCarrico, 1992:66), and phraseological competence, i.e. the ability to use FSs properly, has been included as part of linguistic competence (Howarth 1998:161).

Sinclair (1991) asserts that, based on corpus investigation, the first mode of text interpretation is Idiom principle (p100). He also suggests that FSs, rather than words, are the primary carrier of meaning (Sinclair 2008b:409). Total number of FSs in a language is highly likely to outnumber its total number of single morpheme lexical items. Among all types of FSs that are familiar to an ordinary mature English speaker, 'lexicalized sentence stems' (sentence-length formulaic expressions) alone is estimated to be of hundreds of thousands in number (Pawley and Syder 1983:192 & 210)[14]. This type alone may easily outnumber, for example, 128,000, the estimated total number of words in academic English (Nation, 1990). Chanier et al (1993; quoted in Arnaud and Savignon 1997) report that, in French, compared to 2000 simple adverbs, there are as many as 6000 multiword adverbial expressions (3 times!); and compared with 80,000 simple nouns, there are 300,000 to 400,000 compound nouns (4-5 times!). Mel'cuk (1998:24) goes further to claim that FSs outnumber words roughly ten to one in any language and advocate that a good dictionary should include all the FSs.

Nowadays, study of FS is one of the major issues in applied linguistics in the new millennium (Schmitt et al, 2004:55)[15]. FSs have been put at the very center of language acquisition and seen as basic to the creative rule-forming processes which is preceded by a stage in which learners typically 'use a large number of unanalyzed chunks of language in certain predictable social contexts' (Nattinger and DeCarrico, 1992:xv). FSs are also believed to be beneficial for all language users in saving processing effort, as neurological evidence show that the human brain can be characterized by an imbalance between powerful memory capacity and deficient processing speed (Crick 1979:219). A huge redundant storage of multiword units of various lengths that can save processing time is believed to be 'adaptive for such an organ' (Peters 1983:86). In the field of language teaching and

[14] 'Lexicalized sentence stems' are renamed as 'productive speech formulas' in Pawley (2009:20). The number of 'productive speech formulas' is estimated to be thousands, while the number of lexically specific formulaic expressions realizing them is indefinitely large (ibid:8).

[15] There are a few linguists who do not share the same view. Krashen and Scarcella (1978) advocate that routines and patterns (their terms for memorized whole utterances and sentence frames with slots which are the most salient part of FSs) are significant parts of language teaching system, but only play a minor role (p298).

learning, the awareness that the language system is largely lexically driven has provoked the emergence of the concept of lexico-grammar (Schmitt et al 2004:55) and pattern-based models of acquisition (Ellis, 1996, 2002). Just like acceptable letter sequences (e.g. *sp* can be word-initial but not *hg*) and acceptable combinations of morphemes to form words (e.g. *un-fathom-able* but not *un-able-fathom*), we learn acceptable collocations of words (e.g. *blonde hair* but not *blonde paint*) and acceptable longer FSs based on pattern recognition through repeated exposure rather than rules (Schmitt and Carter, 2004:13-4), and the rules of language we have acquired are just the 'artifact of the pattern-based learning, rather than the underlying source of learning' (ibid). Nowadays FSs, together with single word lexis, are viewed by a significant proportion of foreign language teaching theoreticians as the key to attaining high level of proficiency (Boer and Lindstromberg, 2008a:4). They play an important role in task-based learning (Wills and Wills 1996). Schmitt and Carter (2004) summarize that with the increasing evident importance of FSs in language use, 'convincing explanation of the mechanics of their acquisition must become an essential feature of any model of language acquisition' (p14).

II.4 Formulaic Sequences in L1 and L2 Acquisition

II.4.1Formulaic Sequences in English as L1 and L2

II.4.1.1FSs in Language Development of Young Learners

FSs play a critical role in both L1 and L2 acquisition and there is a consensus that at least some learners rely on FSs when acquiring L1 (Schmitt and Carter 2004:11), and children typically start from 'formula to slot-and-frame pattern to creative construction' (Ellis 2008:5), and many such FSs are taught explicitly by caretakers without explanation (Gleason and Weintraub 1976), and taught with movements and gestures before the children could speak (Firth 1972:33). Clark (1974) observed a boy (data collected between 2;9 to 3;0) holistically recycling structurally complex utterances or parts of utterances just heard to simplify his speech production, and recycling them in combinations to express more complex ideas, seemingly without noticing or analyzing the internal structure of those

utterances[16]. She queries if language competence really comes before language performance (p8).

In her influential dissertation, Wong-Fillmore (1976) traced the acquisition of English of five Spanish-speaking subjects ranging from 5 to 7 years old in a naturalistic setting and found that they 'were able to use the new language in meaningful social settings long before there was any convincing evidence of rule learning' (p718) by relying heavily on largely fixed expressions since very early stages[17]. For example, negative fixed expressions such as *I don't wanna play* (p599) and formulaic questions such as *What does it mean* (p621) were used in very early stages, and all or part of such expressions were used creatively in a later stage to produce utterances like *I don't wanna dese one toys* (p603), and *What does it mean dese book* (p622)[18]. In other words, these children were using structurally complex expressions holistically to construct new utterances seemingly without the underlying language competence. This observation leads Wong-Fillmore to conclude that 'It may be necessary to reject the usual assumption of child language research according to which the utterances a child produces are taken as evidence on the nature of his current system of generative rules' (pvii).

Peters (1977) also observed that a 14-month-old child supposed to be on 'one-word utterance' stage producing sentence-like utterances such as *look at that*! *what's that*? and *open the door*! as holistic units, and proposes that there is a

[16] One interesting example is when the boy wanted to be carried, he produced *I carry you*, which was obviously copied from his father's prior utterance *I'll carry you* (Clark 1974:4). It seems reasonable to speculate that for the boy, *I carry you* is an unanalyzed whole that goes with a certain event or movement. The author also witnessed his 4-year-old nephew replying with *Happy birthday* to people wishing him with the same phrase. Even though the boy definitely had already acquired the two words *happy* and *birthday* per se at that time, seemingly he had not fully acquired the communicative competence to provide a proper reply and simply repeated the phrase just heard.

[17] In the beginning stage, at least 51% of their speech were formulaic (one child even relied 100% on formulaic speech), and in the end of the one year long observation, formulaic speech still constituted from 37% to 82% of their oral production (Wong-Fillmore 1976:642).

[18] These examples are chosen just to demonstrate that formulaic expressions were used as unanalyzed wholes by Wong-Fillmore's subjects. It does not mean that all formulaic expressions used in early stages are well-formed and all those in later stages are ill-formed (see Wong-Fillmore 1976 Chp 3 Section 3-4 for details).

Gestalt, i.e. holistic, approach on top of well-known Analytic approach in acquiring languages (p563-565). In Peters (1983), with more observations, she goes further to advocate that children with different personalities tend to rely on different strategies. Cruttenden (1981) also points out that use of FSs depends on whether a child prefers item-learning or system-learning style. From a diachronic perspective, Wray and Perkins (2000:19-22) declare that the relative proportion of holistic and analytic processing changes from birth to adulthood to form four distinctive stages: entirely holistic (birth to around 20 months), largely analytic (around age 2 to 8)[19], increasingly holistic (around age 8 to 18) and largely holistic with settled balance between holistic and analytic processing (from late teens on).

In the process of learning L1 grammatical rules, children segment the speech they hear into morphemes, which are 'the ultimate units of grammatical rules' (Brown 1973:390), and FSs are believed to be among those materials being segmented for constructing grammatical knowledge (Peters 1983:Chap 3). The critical point is, by the time their analytical ability is fully acquired, young learners of L1 and L2 already have stored a huge number of FSs to ease both comprehension and production (Wray 2000:481)[20]. Besides, unlike in the case of adult learners, FSs are only analyzed when needed and, if analyzed, only to the needed extend. This need-only analysis leads to a storage of unanalyzed FSs, and a multiple storage of analyzed FSs and their constituents (Wray 2002:130-2; cf. Peters 1983 and Bolinger 1976) and this contributes to the forming of native-like ability (see next section for a comparison). In other words, both possessing a huge storage of FSs and knowing what to and what not to analyze contribute to native-likeness.

II.4.1.2 FSs in Language Development of Adult Learners

Nattinger and DeCarrico (1992) suggest that FSs are ideal units of language teaching and learning for both children and adults (p27 & 32). Ellis (1994) points out that compared with all L1 learners who go through a silent period in the

[19] Wray (2002) speculates that this stage might be 'affected by beginning literacy and the analytic method of formal education' (p134)

[20] Wood (2002) maintains that analysis comes later 'partly as a result of neurological development and a resultant increase in analytic cognitive skills' (p4).

beginning, adult L2 learners typically begin to speak by making great use of FSs at the onset (p106)[21]. FSs provide adult learners shortcuts to communicate at a level far beyond their lexical and grammatical knowledge and are very critical to them from a motivational perspective (Hakuta 1976:333). Peters (1983) points out that, for mature learners, FSs can serve as a shortcutting device to save processing time and effort, so that more focus can be put on social aspects of interactions and on macrostructures of discourses (p3). She also asserts that memorizing large chunks of FSs can be useful in developing fluency, and such way of learning is consistent with communicative and notional-functional approaches emphasizing use of routines and formulas to perform speech functions (p111). Learning of FSs is also found to be important for adult native speakers in sector-specific context (Kuiper 2004), and books written for this purpose, e.g. phrase book for academic writing are not scarce[22].

Concerning segmentation of FSs, Wood (2002:5) summarizes that in adult L2 learning, FS are also broken down and analyzed and both the original formulas and the pieces and rules are retained (cf. Bolinger 1976:9; Peters 1983:89-90; Nattinger and DeCarrico 1992:12)[23]. Peters (1983) maintains that if the structural patterns in memorized FSs are extracted and put to use, they can contribute to the larger goals of free creativity (p111).

As far as case studies of adult L2 learners are concerned, Ellis (1994) claims that there are 'few case studies based on naturally occurring learner language that do not make some mention of the prevalence of formulas' (p86). Indeed, FSs are found to play an important role in some case studies of adult language learning (e.g. Schmidt 1983; Yorio 1989; Bradley 2003), though not in all of them (e.g. Shapira 1978).

In the field of ESL, FSs are foregrounded by Lewis (1993; 1997a; 1997b),

[21] According to The ACTFL proficiency guidelines, novice level (speaking) is characterized as being able to 'communicate …… primarily through the use of isolated words and phrases that have been encountered, memorized, and recalled' (http://actflproficiencyguidelines2012.org/speaking).

[22] See Peterson (1998) as an example.

[23] Tremblay and Baayen (2010)'s experiment on native speakers demonstrates that FSs are stored both as wholes and parts.

Nattinger and DeCarrico (1992) and Willis (1990) in their influential teaching syllabuses. They place great importance on native-like language usage through the learning of FSs, while explicit teaching of traditionally treasured grammar rules are given lesser weight[24], as they believe that grammar rules can be derived from analysis of FSs (Nattinger 1988:77; Willis 1990:vii; Nattinger and DeCarrico 1992:27; Wray 2000:470).

In spite of the significance of FSs in adult L2 acquisition, learning of FSs is found to be challenging for adults due to their huge number[25] and the greater memory load involved as they have much longer *signifiants* than single-word lexis (Arnaud and Savignon 1997:161), improper teaching (Irujo 1986:237; Williams 1988:51) and a serious lack of meaningful input, because native speakers tend to avoid using them when speaking to adult L2 learners, while the rich input from TV or movies are not interactive in nature (Irujo 1986:236-7). Many English for academic purposes (EAP) or English for specific purposes (ESP) textbooks are found to be ineffective in dealing with FSs (Wood 2010b:103; Chen 2010). Besides, weakly idiomatic FSs like *to have the last word* and FSs with close equivalents in L1 like *to grease someone's palm* vs. *graisser la patte à quelqu'un* tend to pass unnoticed (Arnaud and Savignon 1997:161), and successful acquisition of FSs is found to be difficult, if not impossible, without socio-cultural adaptation and integration (Dörnyei et al 2004)[26]. Yorio (1989) finds many

---

[24] There was even a debate over whether adopting Lewis' Lexical Approach means giving up teaching grammar. While asserting that 'less attention will be paid to individual words, and substantially less to traditional grammar structures' (1997b:260), Michael Lewis claims that he fully recognizes the pedagogical value of grammar rules as generative element of language and disassociate himself from the view that 'Lexis is the answer' (Lewis 1997a:14), but emphasizes that 'Grammar is particularly useful when we use novel language to talk about unusual situations while lexis is more useful to handle highly probable events fluently and effortlessly by providing prefabricated means to handle them.' (ibid:41). Yorio (1980) also emphasizes that advocating the importance of FSs does not equal abandoning the 'more traditional types of linguistic input' but to foster the use of gestalt learning strategies which might otherwise be ignored (p434).

[25] Number of FSs can be further enlarged by their exploited forms such as *Hook, Lyne and Stinker* (originally *Hook, Line and Sinker*) or *Nothing fails like failure* (originally *Nothing succeeds like success*) (Arnaud and Savignon 1997:161). Chinese examples include witty expressions like *在哪里摔倒就在哪里躺下* (= wherever you fall, just lie down there) which is derived from *在哪里摔倒就在哪里站起来* (= no matter where you fall, get up and keep it up).

[26] Dörnyei et al (2004) observe that only particularly high aptitude and motivation can compensate for the absence of socio-cultural adaption, whereas 'successful socio-cultural adaption can override

ill-formed FSs which are subject to adult learners' interlanguage rules (62-3). Bolander (1989) observe overuse while Foster (2001) observes underuse of FSs by adult learners. Milton (1998:189)[27] and Granger (1998a:155) find both overuse and underuse, and the latter also finds too much creativity within those FSs, and De Cock et al (1998) find adult learners use different FSs when compared with native speakers, or use the same FSs but with different frequency, different syntactic structure and pragmatic functions. Adult learners are also found to create their own ill-formed FSs which are then fossilized (Bolander 1989). Irujo (1993) observes that her advanced adult learner subjects' use of FSs seriously lag their overall proficiency (p207-8). Channell (1994) finds L2 learners' production 'bookish and pedantic' owing to the underuse of formulaic vagueness tags, e.g. *and so on*, despite being syntactically, phonologically and lexically correct (p21). Arnaud and Savignon (1997) also find that while professionally advanced adult learners could slightly outperform university-level native speakers in their knowledge of rare words like *hefty*, *crony* and *buxom*, their performance in FSs was significantly inferior, and speculate that adult L2 learners can reach native-like proficiency with respect to rare words but not FSs (p165-7). Concerning adult L2 learners' problems, Foster (2001) summarizes that

> Unlike first language learners, second language learners are likely to have some degree of explicit knowledge of grammar, either through their own conscious analysis, or through classroom teaching. This, coupled with a restricted bank of memorized language, may mean that they are more apt to use rules when composing language, or to overuse the stock of sequences they have memorized (p80).

Drawing on evidence of both young and adult learners, Wray summarizes that while the acquisition of FSs does facilitate acquisition, it does not help adult L2 learners attain native proficiency as it does to young learners of L1 and L2 (2000:471-2; 2002:175-6). She speculates that in contrast with the need-only pattern of analysis in L1 and L2 acquisition of young learners, adult learners cannot resist analyzing, and excessively break down the FSs they encounter into

---

below-average initial learner characteristics' (p105).

[27] In Milton (1989), some word strings that are overused or underused might not be FSs in strick sense. For example, intuitively *in this case the* among top 10 underused 4-word list does not seem like an entity (p189).

parts while the wholes are ignored and not properly retained. When the FSs are needed and reassembled with the parts retained, interlanguage rules inevitably come into play and result in errors[28] (Wary 2002:Chap 10-11; also see Cowie and Howarth 1996:91). Fitzpatrick and Wray (2006) predict most adult learners to remain victims of over-analysis by 'unwrapping the packaging in the interests of more effective learning but, in the process, losing vital information about how to put the constituent units back together' (p54). In sum, adult learners of L2 can be characterized by a small storage of FSs (some of which are ill-formed and fossilized), and too strong a tendency to break down FSs and recreate them with interlanguage rules[29].

In view of the above differences, Wray (2000) suggests that as many FSs, especially those used in real interaction, are not grammatically and semantically regular to be analyzed to infer syntactic knowledge of contemporary languages, guidance should be given when FSs are presented to adult learners, so that both analyticity and formulaicity can be accommodated (p482-4).

II.4.2 Formulaic Sequences in Chinese as a Second Language (CSL)

Though the research on FSs in CSL started much later, a common consensus concerning the importance of FSs has been formed (Wang 2007:26; Wu et al 2009:2; Su 2010:14). While hardly any studies on L1 and L2 young learners' use of FSs can be found, there are now dozens of publications on adult learners of CSL, though empirical studies are still limited (Wang 2007:8; Zhou 2009:46; Ma 2010:10).

Ding (2006:26) and Wang (2007:19) confirm that number of FSs used positively

---

[28] Wray (2004) demonstrates that errors caused by interlanguage rules will occur at the boundaries of the linguistic units perceived by the learner, i.e. the linguistic units that are broken down from FSs.

[29] Skehan (1998) also points out that NNSs tend to employ unidiomatic combinations of words created with rules.

correlates with learners' oral proficiency, and Yang (2010:41) and Ma (2010:27) observes that more advanced CSL learners used more FSs in essay composition (p26). Su (2010) finds that advanced learners have higher awareness of FSs, and learners at elementary and advanced levels tend to memorize coexist words as holistic units and imitate native speakers' expressions more than those at intermediate level who he speculates to rely more on grammar rules (p35-6).

Concerning errors in the use of FSs, Cong (2010) observes overuse and underuse of difference types of FSs (p29-30), and Zhou (2009) finds that while errors in use of collocations such as *办手续* (= to go through the formalities) and *竞争激烈* (= competition is intense) are extraordinarily serious, collocations are often neglected in teaching (p48). Liu (2012) observes that concepts that are expressed with FSs in Chinese but with single-word lexis in English, such as 造成火车出轨 (= derail a train), are extremely difficult for learners. Wang (2007:29), Yang (2010:31), Ma (2010:10) and Cong (2010:33) all call for awareness-raising in teachers and students.

Zhou (2009) also points out that CSL textbook writers' awareness of FSs is still low (p44 & 48). Ding (2006:7) and Yang (2010:33) observe that some FSs are highlighted in CSL textbooks but not in a systematic way, often with the same type of FSs listed under different headings.

Ma (2010) proposes that FSs should be graded in accordance with their frequency (p29) and Song (2009) states that both frequency and degree of difficulty should be taken into consideration when preparing a list of FSs to facilitate compilation of CSL textbooks (p49). Zhou (2009) compiled 860 FSs containing 348 most frequent words. Dictionary of utterance-length FSs have been published (Chang 1993; Liu et al 2005) but they typically contain only a few hundreds of items.

Wang (2007) proposes that FSs in CSL should be handled differently at different level: more emphasis on utterance-length FSs at elementary level, more sentence-frame FSs with slots at intermediate level (p31). She also emphasizes the role of memorization to build up a large store of FSs (p32).

When addressing the characteristics of Chinese FSs, Yang (2010) points out that FSs consisted of function words, such as *因为 ... 所以 ...* (= because…therefore…), *一旦... 便...* (= if… then…) are widely used (p33).

II.5 Classification of Formulaic Sequences

From the above review, we can tentatively conclude that FSs are very important in learning and teaching of L1 and L2. However, in order to understand what exactly FSs are, and what they include, in a more systematic way (for the working definition of this study, refer to session I.2), a review and comparison of their taxonomies seems necessary.

Categorizing FSs is notoriously difficult (Gläser 1988; Kuusi 1974). It is by no means a 'quick and tidy' procedure, as subjective judgments are inevitable and disagreement among scholars is not uncommon (Koprowski 2005:322). In classifying FSs, researchers have adopted criteria based on the various features of FSs, i.e. form (whether irregularity exists, whether variability is allowed, and whether strong word partnership prevails), function (whether tied to standardized communication situations), meaning (whether semantically transparent) or provenance (whether started off formulaic or become formulaic after some time) (Wray 2002:Chp 3), and degree of idiomaticity[30] and syntactic structure etc (Arnaud and Savignon 1997:160). However, most taxonomies fail to be internally consistent by adopting, for instance, some form-based and some function-based features that partly overlap at the same time, resulting in categories that are 'neither discrete nor comprehensive' (Hudson 1998:13 quoted in Wray 2002). Besides, some types of fixed word strings are not consistently included (Wray 2002:46).

[30] Most FSs are idiomatic to some degree, and some are more idiomatic than others (Gläser 1988). For instance, *to jump the gun* is more idiomatic than *to jump the queue* (Arnaud and Savignon 1997:161).

A few English and Chinese taxonomies are quoted below to illustrate what FSs include and how they can be categorized.

II.5.1 Taxonomies of English Formulaic Sequences

II.5.1.1 Three Representative Taxonomies of English FSs

Among over a dozen taxonomies of English FSs offered so far, Lewis (1993) and Nattinger and DeCarrico (1992) are amongst the most frequently quoted, at least in literature of Chinese FSs, while Wray's is the most comprehensive and internally consistent. They are compared and contrasted below.

Along with his Lexical Approach, Lewis (1997a; 1997b) proposes a 4-way (used to be 3-way in Lewis 1993) taxonomy to categorize FSs (in his term 'multi-word prefabricated chunks') which is simple but contains two form-based categories and two function-based categories:

> **Polywords** are typically two to four words long strings that can be found in dictionaries and behave like individual words. Examples include: *the day after tomorrow; at the weekend; all at once; by the way; on the other hand; taxi rank; record player; put off; look up; look up to; in his element* (1993)*; for example; as a result* (1997a:144)*;*
>
> **Collocations** refer to partnership or co-occurrence of individual words with varying degrees of fixedness and mutual expectancy. Examples include: *prices fell; rancid butter* (1993)*; away from home* (1997a:146)*; work hard* (v.+ adv.)*; strong possibility* (adj.+ n.)*; portrait painting* (n.+ n.)*; take the opportunity to* (v.+ n.)*; embrace the latest technology* (v.+ adj.+ n.)*; raise your blood pressure* (v.+ n.+ n.) (1997a:109)*;*
>
> **Fixed Expressions** are separate utterances. Functionally they are pragmatic in character and their purposes can be easily identified to ensure efficient productive and receptive processing. Examples include social greetings such as *Good morning* and *Happy new year*, politeness phrases such as *No thank you*, *I'm fine* and *I'll have to be going*, 'Phrase Book' language such as *Can*

you tell me the way to … please? And *I'd like a twin room for …… nights, please,* and idioms such as *You're making a mountain out of a molehill*. (Lewis 1997a:9-10)

Semi-fixed Expressions are separate utterances with slots with similar function as Fixed expressions. Examples include nearly fixed expressions permitting minimal variation like *Its/That's my fault*, spoken sentences with slots like *Could you pass me the …… please?* and *I haven't seen you for ……*, sentence heads like *What was really interesting was ……* (1997a:11), semi-fixed frames like *either…or* and *not only …but also* (1997a:144), and extended frames for formal letter or academic paper like the following:

> *There are broadly speaking two views of .... . The more traditional, usually associated with … and his/her colleagues, suggests that ...., while the more progressive view, associated with .... suggest …. . In this paper I wish to suggest a third position, which, while containing elements of the view proposed by .... also takes account of recent developments in .... which have produced evidence to suggest* .... (1997a:11)

Nattinger and DeCarrico (1992) define FSs (in their term 'lexical phrases') as conventionalized form/function composites with more idiomatically determined meaning than novel language created with syntactic rules and each with a particular discourse/pragmatic function (p1 & 36-7). By excluding FSs without such function, they offer a different categorization with better internal consistency based on both form and function. Below are their four formal categories:

Polywords are short, invariable and continuous phrases functioning like individual lexical items in 'expressing speaker qualification of the topic at hand, relating one topic to another, summarizing, shifting topics, and so on' (p38). They can both be canonical (e.g. *by the way; I'll say; at any rate; you know*) or non-canonical (e.g. *as it were; so far so good; by and large; once and for all; not on your life*).

Institutionalized Expressions are sentence-length, invariable, mostly continuous and mostly canonical phrases normally functioning as separate utterance, including 'proverbs, aphorisms, formulas for social interaction, and all of those chunks that a speaker has found efficient to store as units' (p.39), ranging from *a watched pot never boils; how do you do?; get a life; there you go; be it as it may; long time no see* to *once upon a time…and they lived happily ever after*. This type of FSs is even incorporated into 'polyword' category because they are 'more or less sentence-length polywords' (p.65).

Phrasal Constraints are canonical or non-canonical short to medium-length phrases with slots or variable lexical items that function as framework of phrases. Examples include *a day/ year/ ... ago; to tie/ wrap this up; in short/ sum/ summary; good morning/ afternoon/ ...; as far as I know/ can tell/ ...; to make a (very) long story (relatively) short; the soon/ tall/ ...er the bett/happi/ ...er.*

Sentence Builders are framework containing slots for 'parameters or arguments for expression of an entire idea' to make whole sentences, e.g. *I think that ...; not only ..., but also…; my point is that…; I am a great believer in ...; it is only in ... that ...; that reminds me of ...; the sooner ..., the better ....*

After incorporated Institutionalized Expressions as part of Polywords, Nattinger and DeCarrico further assign the above FSs into three functional categories, namely:

Social Interactions: including speech acts like summoning, nominating a topic, clarifying, checking comprehension, shifting a topic, closing, parting, questioning, requesting, refusing, complimenting, etc.

Necessary Topics: topics frequently used in daily life conversations like autobiography, language, quantity, time, location, weather, likes and dislikes, food and shopping, etc.

Discourse Devices: phrases used to connect the meaning and structure of the discourse like logical connectors, temporal connectors, spatial connectors, fluency devices, exemplifiers, relaters, qualifiers, evaluators and summarizers etc.

The table below shows the 9 combination of FSs with distinctive forms and functions (in parenthesis are the functions assigned to the preceding FSs):

Table II.1 Categorization of FSs both by form and by function, adapted from Nattinger and DeCarrico (1992:65-6)

	Social Interactions	Necessary Topics	Discourse Devices
polywords	by the way (shifting a topic) all right? (checking omprehension)	a great deal (quantity) too expensive (shopping)	in other words (exemplifier) at any rate (fluency device)
Phrasal Constraints	_____ me? (clarifying: audience) see you _____ (parting)	I am from_____ (autobiography) how much is _____? (quantity)	as far as I _____ (evaluator) as a result of _____ (logical connector)
Sentence Builders	what I mean is … (clarifying: speaker) do you know …? (nominating a topic)	what do you like to…? (likes) what time …? (time)	there's no doubt that… (evaluator) my point here is… (summarizer)

Similar to the 'extended frames' for formal letter, academic paper and company reports illustrated by Lewis (1997a:11 & 33), Nattinger and DeCarrico (1992:Chp 7) also demonstrate how sentence-length FSs can be combined to create a 'skeleton structure' of an essay for academic and business purposes, such as the following:

Opening: *It has been often asserted that … The purpose of this paper is to …, and to maintain that … The paper will show that … by comparing … and by contrasting…*

Body: .....

First paragraph *It can be said that …… lend support to the argument that …*

Second paragraph *…does not support the argument that …*

Third paragraph *Both … and … are similar in that …is unlike… with respect to …*

Closing: *In conclusion, one can generalize that…*

Nattinger and DeCarrico (1992:171-2)

In the research-based Heteromorphic Distributed Lexicon model (Wray 2002:248-251 & 263), Wray offers an all-encompassing function-based five-way division of FSs, with good internal consistency while being able to accommodate more types of FSs than the above two categorizations:

Grammatical word strings have the least autonomous meaning and low context dependency and, though quite fixed in their own form, serve as grammatical constructions of novel utterances, e.g. *in order to; on account of; out of* (as in 'get *out of* the car').

Referential word strings are referential expressions including customary collocations, polywords and phrasal constraints that constitute novel utterances, e.g. *date for your diary; major bone of contention; face the problem; take medicine; give NP to NP; highly likely; half past NUM (1-12); pull NP+GEN leg.*

Interactional (routine) word strings are sentence-length routines with interactional functions, including group chants. They typically have less creativity than novel utterances derived from Grammatical and Referential word strings, e.g. *Great to see you; Look out!; Get out of my way!; Pass the N please; I'll give you NP for it; The most important thing is; Of course; Is that a fact?*

Memorized word strings are memorized texts with even less creativity, e.g. Hamlet's soliloquy; times tables; songs; nursery rhymes; prayers.

Reflexive word strings are fundamentally different from the first four types probably due to their 'association with subcortex rather than the cortex' (ibid:256). They are fully reflexive exclamations with greatest context dependency, containing lexical units retrieved unconsciously 'as an automatic response to external or psychological stimuli' (ibid:250). They typically stand alone but can also serve as 'dummy fillers to compensate for gaps in expression', e.g. *Bloody hell!; Goodness gracious!; What the!*. This type of word strings are redundantly stored as part of interactional categories, to be used interactively (ibid:256).

It is worth noting that Wray's taxonomy overtly allows multiple representations: Wray asserts that mental lexicon is heteromorphic and language items are stored redundantly in units of different sizes and different types and when needed to be used, can be retrieved along different paths (2002:251-3, 2008:12-3)[31]. The same phrase 'watch your bag' can be generated as a novel utterance syntactically (VP + NP), as well as retrieved as fixed interactional routine (*watch your bag !*) or semi-fixed interactional routine with slot (*watch your* _____*!*) to be filled (by '*bag*'), or even produced as memorized strings for mnemonic rather than interactional purpose, such as teaching material being drilled in a language class or lines memorized in acting.

II.5.1.2 A Comparison of the Three Taxonomies of English FSs

As different authors give different examples as illustrations, and as examples are too limited in some categories，e.g. Wray (2002) only offers *in order to, on account of* and *out of* as examples of grammatical word strings, it is very difficult to match type with type. However, a few points are quite clear as shown in the table below (since Wray's taxonomy is the most comprehensive, systematic and

[31] Lewis (1997b) also allows multiple representations but it is mainly for practical reasons related to classroom teaching (p256).

internally consistent, both for English and Chinese[32], it is used as the base for comparison in this research):

Table II.2: a comparison of 3 categorizations of English FSs (with Wray 2002 as the base)

Wray (2002)	Lewis (1997)	Nattinger and DeCarrico (1992)
Grammatical word strings e.g. *in order to; on account of; out of*	Semi-fixed Expressions e.g. *either…or; not only …but also;*	Sentence Builders e.g. *not only …, but also…; the sooner …, the better …;*
		Phrasal Constraints e.g. *the ___er the er;*
Referential word strings e.g. *bone of contention; take medicine; give NP to NP; highly likely; half past NUM (1-12); pull NP+GEN leg;*	Polywords e.g. *all at once; by the way; on the other hand; in his element;*	Polywords e.g. *by the way; I'll say; you know*; *by and large; once and for all;*
		Phrasal Constraints e.g. *a day/ year/ … ago; in short/ sum; to make a (very) long story (relatively) short;*
	Polywords e.g. *the day after tomorrow; record player; put off;*	X (polyword nouns & phrasal verbs)
	Collocations e.g. *prices fell; butter rancid; work hard;*	X (collocations)
Interactional (routine) word strings e.g. *Look out!; Pass the N please; I'll give you NP for it; The most important thing is;*	Semi-fixed Expressions e.g. *Sorry to interrupt, but can I just say…; That's all very well, but……;*	Sentence Builders e.g. *I think that …; I am a great believer in …; my point is that…; it is only in … that …; that reminds me of …;*
		Phrasal Constraints e.g. *good morning/night;*
	Fixed Expressions e.g. *Just a moment, please; I'll drop you a line; The damage is already done; You are making a mountain out of a molehill.*	Polywords e.g. *so far so good;*
		Institutionalized Expressions e.g. *a watched pot never boils; how do you do?; get a life; there you go; be it as it may; long time no see; once upon a time…and they lived happily ever after*
Memorized word strings e.g. times tables; songs; nursery rhymes;	X	X
Reflexive word strings e.g. *Bloody hell!; Goodness gracious!; What the!.*	Fixed Expressions	Institutionalized Expressions

[32] While intuitively it is very easy to think of many Chinese counterparts of FSs in all five categories, only examples of the first four categories can be found in literature review on Chinese FSs (see below).

1) Like many others, both Nattinger and DeCarrico (1992) and Lewis (1993; 1997a; 1997b)'s taxonomies are obviously not comprehensive. The former does not include FSs without discourse/pragmatic functions[33], while the latter does not have a proper place to accommodate relaters like *either…or* and *not only …but also* (p144)[34]. Besides, although both Lewis and Nattinger and DeCarrico include 'extended frames' (1997a:11) or 'skeleton structure' (1992:Chp 7) for formal letter or academic paper in their taxonomies, neither of them explicitly include memorized strings of words and sentences[35] such as numeral 1 to 10, ABC's, antiphonal sequences in games and ceremonies, and limericks etc mentioned by Hymes (1962), and Hamlet's soliloquy; times tables, songs, nursery rhymes and prayers under Wray (2002)'s Memorized word strings[36].

2). Wray's taxonomy is superior not only because of its broader coverage, but also because of its higher degree of internal consistency and neat separation

[33] Unlike in Lewis (1993; 1997a; 1997b) and Wray (2002), Nattinger and DeCarrico (1992) do not treat collocations such as *work hard* and *take medicine* as FSs because they have not been assigned particular discourse/pragmatic functions and are not form-function composites (p37). Their 'polyword' category also does not include polyword nouns such as *taxi rank* and *bone of contention* and phrasal verbs such as *put off* and *look up*, which are recognized by most literature reviewed in this study as formulaic, seemingly due to the same reason. Wray's grammatical strings such as *in order to* and *out of* also seem difficult to find their place in Nattinger and DeCarrico's.

[34] Lewis (2007a) includes semi-fixed frames like *either…or* and *not only …but also* (p144) as part of Semi-fixed expressions, i.e. separate utterances with slots with pragmatic functions. It seems that these are actually relaters with discourse functions rather than pragmatic functions (Nattinger and DeCarrico 1992:64) and Lewis' taxonomy cannot accommodate them well. Wray does not use them as examples but in her taxonomy, Grammatical word strings seems to be the best place for them.

[35] It might be counter intuitive to treat these word strings, especially those longer ones like limericks and other poems, and times tables and frequently chanted prayers and religious verses, as FSs, because we can retrieve parts of them when needed as in the case of times tables, and we might forget parts of them. However, their being formulaic can be supported by reports on aphasiacs who could recite remembered long texts as holistic units but could not even repeat short sentences consisting of words in the same texts, such as the case reported by Peter Rommel in 1683 of a aphasiac woman who could not say a word besides *yes* and *and*, but could recite Lord's Prayer, Apostles's creed and some Bible verses without hesitation (Benton and Joynt 1960:113-4, 209-210). It seems that it is reasonable for a comprehensive taxonomy to include text FSs as members.

[36] As for Wray's memorized word strings, in view of their fixedness, they might be part of Lewis' polywords, but it is not clear if they can be accommodated by Nattinger and DeCarrico's.

from each other. FSs are all classified at generic functional level[37]: syntactic function, semantic function, pragmatic function, mnemonic function and emotional/physiological function of language use[38].

3). While being superior in many aspects, as Wray's taxonomy includes far more types of FSs than the others, each of Wray's five functional categories tends to contain more kinds of word strings. When practically identifying and analyzing FSs, names of categories in other taxonomies can still be integrated into Wray's. For example, Lewis' Polywords and Collocations can be formal subcategories under Wray's Referential word strings, while Lewis' Fixed and Semi-fixed expressions can be formal subcategories under Wray's Interactional word strings.

In sum, Wray (2002)'s taxonomy offers more flexibility, on top of its discreteness, to accommodate probably any formulaic word strings, and thus will be used as a 'template' taxonomy in handling Chinese FSs in this study[39].

[37] Although both taxonomies by Nattinger and DeCarrico and Wray are function-based, the meaning of 'function-based' in the former specifically refers to discourse functions such as qualifier, summarizer, topic shifter, agreement marker, fluency device (Nattinger and DeCarrico 1992:38), and pragmatic functions such as greeting, warning, objection, and denial etc (ibid:40), while the latter is a more generic term covering also the structural, mnemonic and emotional functions of language items, on top of referential and interactional ones (Wray 2002:248-251 & 263). In short, Wray's taxonomy has taken into consideration broader usage of language and is more comprehensive.

[38] This avoids the problem caused by many other taxonomies like the one proposed by Lewis (2003), in which, for example, institutionalized expressions (the whole set of pragmatic function) and collocation (a subset of semantic function) are juxtaposed. In Nattinger and DeCarrico (2002), greeting expressions like *good morning/ afternoon/ night* are only under 'phrasal constraints', but not under 'institutionalized expressions', which include *long time no see* and *how are you*? as members. Parting expressions like *see you, see you soon/ later* are under 'phrasal constraints' but *good bye* and *so long* are under 'institutionalized expressions' even though these expressions do not have fundamental difference. As shown in the table, 'phrasal constraints' might better be a subset under 'polywords' and 'institutionalized expressions', rather than an independent category at the same level, to keep categories separate, which most taxonomies cannot manage to do (Hudson 1998 quoted in Wray 2002:47).

[39] See Wong (2011) for a taxonomy of Chinese FSs based on Wray (2002)'s.

II.5.2 Taxonomies of Chinese Formulaic Sequences

II.5.2.1 Taxonomies of Three Representative Chinese FSs

Research on Chinese FSs has increased tremendously in the past few years and below are three taxonomies proposed lately:

Zhou (2007:99-100) proposes the following three categories

> **Collocations:** coexisting words, usually with the first part implying the second part, including subject-verb collocations e.g. *功能衰竭* (= physiological function failure), verb-complement collocations e.g. *瘦得皮包骨* (= as skinny as a bag of bones), modifier-modified collocations e.g. 朦胧的月光(= dim moonlight) and *可持续发展* (= sustainable development), verb-object collocations e.g. *共商国是* (= discuss state affairs), etc
>
> **Conventionalized Expressions:** including 'chengyu' idioms[40] like *公事公办* (= business is business), 'suyu' idioms[41] like *没完没了* (= without end) and high frequency fixed or semi-fixed short phrases like *撒腿就跑* (= dash off to somewhere), *吓我一大跳* (= gave me a start), gambits like 话又说回来 (= nonetheless) and interactional routines

---

[40] Chinese 'chengyu' idioms are typically composed of 4 characters and with an historical origin (The Contemporary Chinese Dictionary, 2002). The earliest 'chengyu' can be dated back to Shi Jing or Book of Songs, compiled between 1000 B.C. and 400 B.C., and many others from classical Chinese books (Zhou, 2004:227-230). Some of Chinese 'chengyu' idioms are semantically transparent but many are either archaic in wording, opaque in meaning or complicated in structure. Compared with 'suyu' (see below), they are more refined and elegant and have been favored by scholars and literati since ancient times (Zhou, 2004:224).

[41] 'Suyu' idoms include 'guanyongyu' (customary sayings), 'xiehouyu' (two-part allegorical sayings) and 'yanyu' (proverbs), etc (Zhou, 2004:333). Except for 'guanyongyu', 'suyu' are typically composed of four or more characters. Compared with 'chengyu' idioms which are widely used by people with higher education, 'suyu' idoms are normally easier to understand and widely used by ordinary people. Among 'suyu', 'Guanyongyu' are more descriptive sayings, 'xiehouyu' are two-part allegorical sayings like *皇帝的女儿 - 不愁嫁* (emperor's daughter -- need not worry that she cannot soon be wed = people or commodity in short supply), and 'yanyu' proverbs are pithy sayings that express profound precepts.

Sentence Connectors: e.g. *既不是...也不是* (= neither … nor), *要么 ... 要么* (= either … or), *除非 ... 否则* (= unless … otherwise), 宁可 ... *也不* (= would rather … than)

Qian (2008) advocates taxonomy of structure (and function) with three layers (p142):

Phrase level FSs including 'chengyu' idioms, 'xiehouyu' idioms, collocations like *春夏秋冬* (spring, summer, autumn and winter = four seasons), polywords like *阿弥陀佛* (= Buddha the Eternal life), 'guanyongyu' idioms like *吃鸭蛋* (lit. eat a duck egg = get zero scores in an exam), and gambits like *你知道* (= you know) and *我的妈呀* (my mother = my God) with syntactic functions. Some also have pragmatic functions.

Sentence level FSs including 'yanyu' idioms like *远亲不如近邻* (= close neighbors are more helpful than close relatives living afar), interactional routines like *好久不见* (= long time no see), fixed phrases like *且听下回分解* (= will be disclosed in the following chapter) used in traditional Chinese novels, catch-phrases and quotations with pragmatic functions.

Text level FSs including nursery rhymes, songs, religious texts and poems like *床前明月光,疑是地上霜.举头望明月,低头思故乡* (I saw the moonbeams play before my berth, And wondered if that can be frost on earth. I raised my head, looked at the moon, I bowed my head, thought of my home)[42]

Wu et al (2009) propose a taxonomy that highlights the fixedness of FSs (from the most fixed to the least fixed).

[42] English translation of Li Bai's well-known poem *Jing ye si* (On a Quiet Night) was by Y. B. Liang (http://blog.roodo.com/dcalfine/archives/1682094.html)

Phrasal Constraints & Sentence Builders: e.g. *对 ... 来说* (= as far as … is concerned), 从 ... *出发* (= to depart from …), 拿 ... *来说* (= take … as an example), *实话对你说吧* (= the truth is), *不管怎么说* (= in any event), *因为 ... 所以* (= because… therefore)

Fixed Chunks: including 'chengyu' idioms like *三心二意* (three heart two mind = half-hearted), 'suyu' idioms like *失败是成功之母* (= failure is the mother of success), 'xiehouyu' idioms like *黄鼠狼给鸡拜年 – 没安好心* (yellow weasel pays chicken a New Year's call = with bad intention), and 'guanyongyu' idioms like *开夜车* (drive a night train = to burn the midnight oil) and *眼中钉* (nail in one's eye = thorn in one's flesh)

Complement Chunks: e.g. *划得来* (= worth it), *了不起* (= remarkable), *来不及* (= too late), 吃得消 (= be able to endure)

Detachable Verb-Object Chunks: e.g. *洗澡* (= take a shower), *照相* (= take a picture), *握手* (= shake hands), *唱歌* (= sing a song), *起床* (rise from bad = get up), *睡觉* (sleep a sleep = to sleep), *见面* (see face = meet)

Institutionalized Expressions: e.g. *闭嘴!* (= shut up!), *最近怎么样?* (= how have you been recently?), 谢谢你 (= thank you), *我的意思是* (= I mean), *发展中国家* (= developing countries), *冉冉升起* (= (of the moon) rises slowly), *机遇与挑战* (= opportunities and challenges)

Immediate Chunks: e.g. *听不懂* (= cannot understand), *猎头公司* (= headhunter), *女生们先生们* (= ladies and gentlemen), *北京欢迎你* (= Beijing welcomes you)[43]

[43] This is the official slogan of 2008 Beijing Olympic Games.

II.5.2.2 Interim Summary of Chinese FS Taxonomies

The following points can be observed from the above taxonomies with reference to Wray (2002)'s:

1. All types of FSs in English, e.g. Polywords, Collocations, Phrasal Constrains, Sentence Builders, Interactional word strings and Memorized word strings, be they proposed by Lewis or Nattinger and DeCarrico or Wray, can find their counterparts in Chinese[44]. However, some FSs in Chinese might be structurally unique, i.e. verb-complements compounds like *瘦得皮包骨* and *划得来*, and some special verb-object compounds like *睡觉* and 游泳 (lit. to swim a swimming = to swim). These FSs might deserve more attention in teaching and learning.

2. Just like in the field of English FSs, examples given with the three Chinese taxonomies are limited and this makes comparisons difficult. However, it is reasonable to believe that there is not yet a consensus on what should be included[45].

3. There seems to be a lack of internal consistency and discreteness in all taxonomies. For example, in Zhou (2007), Collocations is form-based, while Sentence Connectors is function-based. On the other hand, Conventionalized Expressions seems to be a mixture of form-based and function-based FSs,

[44] It doesn't mean that every single FS in English has a counterpart in Chinese. Some FSs in one language might be expressed by a single-word lexis in another language. For example, the Chinese counterpart of *as a result* is *结果*.

[45] For example, while verb-complements compounds are included as FSs, Zhou (2007) only include idiomatic ones like *瘦得皮包骨* and *吓我一大跳* while Wu et al (2009) include all in their Immediate Chunks. Besides, Wray (2002)'s Memorized word strings are not constantly included. Polywords such as *电话号码* (= telephone number) and *有意义* (= meaningful) are not explicitly included in Zhou (2007)'s. Sentence connectors like *既不是...也不是* (= neither … nor) and *要么 ...要么* (= either … or) are not included in Qian (2008)'s (Even if included, they might be difficult to be categorized in Qian (2008)'s, because they are neither phrases nor sentences). And it is not clear if verb-complements compounds such as Zhou (2007)'s *瘦得皮包骨* and Wu et al (2009)'s *划得来* and *听得懂* (see below) are included in Qian (2008). It is also not clear if common verb-object compounds like *睡觉* (sleep a sleep = to sleep) and *见面* (see face = meet) in Wu et al (2009) are included, because all verb-object compounds used as examples belongs to 'guanyongyu' idioms in Qian (2008).

because it contains ‘chengyu’ idioms, ‘suyu’ idioms and high frequency fixed or semi-fixed short phrases without pragmatic functions, and gambits and interactional routines with pragmatic functions[46]. Besides, *可持续发展* (= sustainable development) is categorized as a collocation but looks more like a polyword. In Qian (2008), the boundary between phrases and sentences may not be so easy to define. While ‘Chengyu’ idioms and ‘xiehouyu’ idioms are classified as phrases, many ‘Chengyu’ idioms and most ‘xiehouyu’ idioms are structurally and functionally like sentences. Polywords like *阿弥陀佛* (= Buddha the Eternal life) and gambits like *你知道* (= you know) and *我的妈呀* (my mother = my God) are classified as at Phrase level with both syntactic and pragmatic functions. However, it seems that their functions are primarily pragmatic. In Wu et al (2009), FSs of similar nature are divided into different categories, according to their fixedness (which might be a problematic way of categorization[47]). While *发展中国家* (= developing countries) is under Institutionalized Expression, *猎头公司* (= headhunter) is under Immediate Chunk, but both are Polyword nouns. Verb-object collocation can be found in Fixed Chunks (e.g. *开夜车*), and also in Detachable Verb-Object Chunks (e.g. *洗澡*).

4. ‘Chengyu’, ‘guanyongyu’, ‘yanyu’ and ‘xiehouyu’ idioms are all included as FSs but they tend to be categorized as one or two types in the taxonomies,

[46] As a result, Zhou (2007)’s Collocations and Conventionalized Expressions are not discrete. For example, *瘦得皮包骨* (= as skinny as a bag of bones) in Collocations and *吓我一大跳* (= gave me a start) in Conventionalized Expressions are structurally similar (i.e. both are verb-complement collocations with idiosyncratic structure), functionally similar (i.e. both are idiomatic ways of describing something unusual) and as fixed as each other in form. It is not clear why one is categorized as a member of Collocations and the other one Conventionalized Expressions. According to Lewis (1997a)’s definition, both 瘦得皮包骨 and *吓我一大跳* might better be categorized as polywords.

[47] Degree of fixedness might not be easy to measure and might not be a good categorizing criterion. While *对 ...来说* (= as far as … is concerned) in the most fixed Phrasal Constraints & Sentence Builders can be varied as *对 ... 来讲, 划得来* (= worth it) in third most fixed Complement Chunks cannot. While *失败是成功之母* (= failure is the mother of success) in the second most fixed Fixed Chunks can be altered to *失败乃成功之母, 猎头公司* (= headhunter) in the least fixed Immediate Chunks cannot.

even though they might be structurally or functionally different. Categorization of these idioms might need to be reexamined.

5. While 'Guanyongyu' verb-object compounds like *开夜车* are indisputably FSs, 'Guanyongyu' nouns like *眼中钉* and *落汤鸡* are single-word lexis in dictionaries. It seems that the definition of Chinese FSs also needs to be reexamined to see what should not be included.

In sum, the Chinese FS taxonomies reviewed are neither all-embracing nor with discrete categorization. A new taxonomy is needed to make up this inadequacy.

As the principal concern of this study is formulaic sequences (FSs) in spoken Chinese/Putonghua, and the purpose is to gain a deeper understanding of the nature of them to find out how they are used by learners and native speakers of Chinese in the same specific context, thus to make useful inferences in teaching and learning Chinese as a Foreign Language (CFL), unlike morpheme equivalent units (MEUs) proposed in Wray (2008), the working definition of FS borrowed from Wray (2002) is very loose (see Section I.2), so as to include all linguistic units considered formulaic 'in any research field' (Wray 2002:9). All FSs proposed in the three Chinese taxonomies reviewed will be included in this study.

II.6 Identification of Formulaic Sequences

With the sample taxonomies reviewed in the last section, should identification of FSs be very straight forward and clear-cut? Unfortunately not! In spite of the many criteria proposed, such as Hickey (1993:32) and Gries (2008), identification of FSs has been reported very challenging, because most formulaic sequences are with very low occurrence[48] and with too many variations, plus a lack of consensus over

[48] Many indisputable FSs like *The king is dead* and *long live the King* are not frequent (Wray 2002:30), and many phrases that 'would be considered a normal part of any native speaker's repertoire' cannot be found even once in the 300-million-word The Bank of English corpus (Foster 2001:81).

many word strings that seem formulaic. Besides, the same word string can be both formulaic and non-formulaic and needs contextual and pragmatic cues to disambiguate[49] (Wray 2002:31). As a result, although many scholars have queried the reliability of intuition and proposed computer frequency counts, internal structures and phonological features etc as alternative measures, intuition has been widely used, because all other measures also have their inherent limitations and cannot absolutely get rid of the use of intuition in the whole process (Wray, 2002:chp 2). For example, while intuitively computer frequency count is very reliable and cost effective, Moon (1998), in describing how fixed expressions including idioms (FEIs: the FSs investigated in her study) in corpus can be identified, summarizes:

> Ideally, the FEIs in a corpus would be identified automatically by machine, thus removing human error or partiality from the equation. There is, however, no evidence that this is possible given the current state of the art. It is also difficult to see exactly how progress can be made. The problems arise because in so many cases FEIs are not predictable, not common, not fixed formally, and not fixed temporally (that is, they are often vogue items like slang). They are dynamic vocabulary items, whereas – at least at present – corpus processing requires givens and stability (p51).

Cowie (1998) also queries the 'insistence of some scholars involved in the computerized analysis of large corpora that frequency of co-occurrence is the only significant measure of 'conventionality' in language' (p226). Schmitt et al (2004) suggest that 'corpus data on its own is a poor indicator of whether those clusters are actually stored in the mind as whole' (p147). Moon (1998a) argues that intuition is necessary in order to identify variations or creative usage of certain FSs (p49), to reject non-FS strings containing words that co-occur but actually do not form a FS, and redefining target FS when corpus searches yield strong evidence (p50).

To alleviate the problems individual intuition may bring, Foster (2001) used professional intuition of seven native-speaking veteran linguists in the process of

[49] For example, *keep you hair on* is a FS when it means 'calm down' but is not formulaic when meaning 'don't remove your wig' (Wray 2002:31). Other FSs with freely generated counterparts include *I will talk to you later* (Kecskes 2003:5), *kick the bucket* and *Don't go away* (Kecskes 2003:109).

identification of formulaic sequences exhaustively. The process is reported to be very exhausting and time-consuming (p83-4), but seems to be a sensible move when a researcher has no better choices.

Compared with the great skepticism in Wray (2002) about the use of intuition in identification, Wray and Namba (2003, quoted in Wray 2008), Wray (2008; 2009) and Namba (2010) are much more positive in this regard. Instead of discarding or marginalizing intuition, after extensively researching data of adult native speakers as well as children, non-native speakers, and people with linguistic disabilities, Wray and Namba (2003) and Wray (2008:113-121) propose the following 11 practical and tested diagnostic criteria to confirm if a word string judged by intuition as a formulaic sequence is a morpheme equivalent unit (MEU), a term Wray coins to denote word strings that are indeed formulaic, i.e. 'processed like a morpheme, that is, without recourse to any form-meaning matching of any sub-parts it may have' (Wray 2008:12).

A. Is the word string grammatically irregular? E.g. *holier than thou; if I were you*.

B. Is the word string semantically opaque (the meaning of the whole is different from the combined meaning of the parts)? E.g. *by and by; beat about the bush*.

C. Is the word string situation-specific and/or register-specific? E.g. *many happy returns; your Majesty*.

D. Does the word string perform a pragmatic function on top of conveying the sheer meaning of the words? E.g. *I promise; I now pronounce you man and wife*.

E. Is the word string the one most commonly employed by a certain speaker when conveying a certain idea?

F. Does the speaker accompany this word string with an action, punctuation, or phonological pattern to indicate that it is a unit? Is the speaker echoing something just heard? E.g. touching wood while saying *touch wood*; making 'quotation mark' gesture in the air or speaking with a special

pitch and tone of voice while making quotes; using hyphenated phrases such as the one in 'She adopted a *do-as-you-would-be-done-by* mentality'; repeating intentionally or unintentionally something just heard.

G. Is the word string grammatically or lexically marked by the user as a unit? E.g. treating *weapons of massive destruction* as a unit and mistakenly pluralizing it at the end of it, resulting in *weapon of massive destructions*.

H. Is it highly likely that the speaker has encountered this word string before?

I. Is the word string a clear derivation of an obvious FS? E.g. *I slept like a twig* derived from *I slept like a log* or *Somewhere over the raincoat* from *Somewhere over the rainbow*.

J. Is the word string an obvious FS that has been used mistakenly? E.g. using *I am sorry* while *Excuse me* is more appropriate.

K. Is the word string too sophisticated, or not sophisticated enough, for the speaker, who is typically a learner of that language, or someone with language disorder?

When applying Wray's criteria in his research, the 12th criterion is added to the list by Namba (2010:139):

L. Is there an underlying frame (with slots to be filled) in the word string, like 'NP be-TENSE sorry to keep-TENSE you waiting' in *I am sorry to have kept you waiting*?

He also adds that among the twelve criteria, the most powerful two are semantic opaqueness (B) and pragmatic function (D) (ibid).

It is worth noting that according to the above criteria, any formulaic word strings can be included as MEUs, be it grammatically or pragmatically correct or incorrect (criterion G & J), widely accepted in a speech community or repeatedly used by only one individual (Criterion E), long-lasting or just one-off (Criterion H), used

verbatim or creatively (Criterion I). Therefore, it is important that when applying these criteria, differences in theoretical positions be acknowledged.

As shown in the table below, the 12 criteria can be categorized into four broad types of criteria: specificity of usage, irregularities, errors and deviations, and use of wholeness markers and underlying frames:

Table II.3 Summary of Wray (2008) and Namba (2010)'s Criteria for Identification of MEUs

Types of criteria	Specific criteria	Example(s)
Specificity of usage	Situation/register specific (C)	*many happy returns; your Majesty*
	Pragmatically specific (D)	*I now pronounce you man and wife*
	Person/idea specific (E)	Not provided but Barack Obama's *Yes we can* seems to be a good example
	Encountered before (H)	*Naughty boy*
Irregularities	Grammatical irregularities (A)	*holier than thou; if I were you*
	Semantic irregularities (B)	*by and by; beat about the bush*
	Developmental irregularities (K)	long texts remembered by aphasiacs reviewed in II.5.1.2
Errors & deviations	Grammatical errors (G)	*weapon of massive destructions*
	Derivation from known FS (I)	*I slept like a twig* (derivation from *I slept like a log*)
	Misuse of known FS (J)	*I am sorry* instead of *Excuse me*
Wholeness markers & underlying frames	Gesture, tone, pitch etc (F)	*Touch wood* accompanied by wood touching
	Underlying frames (L)	'*NP be-TENSE sorry to keep-TENSE you waiting*' in *I am sorry to have kept you waiting*

Although the above criteria are about MEUs rather than FSs, they can definitely serve as a useful reference in identification of the latter, which is much looser in definition because any word strings appearing to be prefabricated are included (Wray 2002:9).

In sum, a good balance of modern technology and criteria-based intuition seems to be the state-of-the-art solution in identification of FSs.

II.7 Exhaustive Identification of Formulaic Sequences

Wray (2009) points out that among the many challenges in researching FSs, 'the single most persistent and unsettling one is knowing whether or not you have identified all and only the right material in your analysis' (p28). To identify all the

FSs in a given text takes more effort than Foster (2001) who defines word strings marked by at least 5 out of 7 professional judges as formulaic, because FSs can be redundantly stored (Wray 2002:130-2; Skehan 1996:17). For example, if *I am sorry to have kept you waiting* is identified as an Interactional FS or Fixed Expression or Institutionalized Expression, *NP be-TENSE sorry to keep-TENSE you waiting*, the sentence stem proposed by Pawley and Syder (1983) underlying it is also a FS. On top of that, *I am sorry to …* is intuitively an Interactional Sentence Head or a Semi-fixed Expression, *keep-TENSE NP waiting* is seemingly a Referential FS or Phrasal Constraint with a slot[50], and *keep* and *waiting* might be a pair of collocates[51].

The exhaustive identification shown above is not just for theoretical purposes. All the five continuous and discontinuous word strings identified can be useful input in teaching and learning. As a learner of English, the author learned to use *I am sorry to have kept you waiting* and *I am sorry to keep you waiting* holistically to perform certain speech act in certain situation, and gradually realized that there is a frame like *NP be-TENSE sorry to keep-TENSE you waiting* that he could draw on to generate *we are sorry to have kept you waiting or I am so sorry to keep you waiting so long*, but not with any other verbs to replace *keep* and *wait*, and with the knowledge of *keep* and *waiting* as a pair, to avoid or self-correct sentences with errors or flaws like the following:

**I am sorry to keep you wait*
**I am sorry I keep you waiting*

---

[50] As of 22 Sept 2011, in the 100 million word online British National Corpus (BNC), there are 8 tokens of *keep me waiting* and 7 tokens of *keep him/her waiting.* Among the 19 tokens of *keep you waiting*, 9 are in non-apologetic expressions like *there are hotels which keep you waiting at reception* and *we won't keep you waiting*. That means *keep-TENSE NP waiting* is a referential FS with slot that can be used in sentences not only with apologetic functions.

[51] While *keep* tends to collocate with *waiting*, *make* tends to collocate with *wait* as in *Don't make me wait too long* (0 token of *keep me wait* and *make me waiting* can be found in British National Corpus (BNC)). In this sense, it sounds reasonable to treat *keep* and *waiting* as a pair of collocates. On the other hand, the *keep* in the sense of *keep somebody waiting* cannot be replaced by its synonyms like *hold*, *retain*, *maintain* and *preserve*, etc. This can further justify that *keep* and *waiting* are a pair of collocates.

**I am sorry for keeping you wait*

** I am sorry to make you waiting*

**I am sorry to ask you to wait*

**I am sorry I ask you to wait*

**I am sorry to let you wait*

We can imagine how difficult it is, from a pure grammatical perspective, to explain why we have to say *keep somebody waiting* but *make somebody wait*. It can also be disastrous if learners generate something like **I am sorry to have seen you crying*, **I am sorry to have forced you studying* or **I am sorry to have let him going* based on the grammatical structure of *I am sorry to have kept you waiting*. Identifying FSs along with their pragmatic functions, situation of use and unique semantic or syntactic patterns seems very critical.

In advocating the Lexical Approach, Lewis (1997b) asserts that one of the two most essential changes to the teachers' mind-set is 'a willingness to search for, identify and direct attention' toward the FSs in naturally occurring language (p269). The exhaustive identification of FSs seems to be an essential component of Lewis' approach.

In order to do exhaustive identification of FSs in Chinese, besides the FS types found in literature review, i.e. 'chengyu' and 'suyu' idioms, sentence frames, sentence connectors, poems, verb-complement compounds, verb-object compounds and fixed expressions, we will turn to Chinese 'duanyu' for more inventories.

II.7.1 Taxonomies of Chinese 'duanyu'

In the field of Chinese linguistics, there is a multi-word unit called 'duanyu' or 'cizu' (normally translated as 'phrase') which is composed of two or more words (Qi 2000:3). A 'duanyu' is normally an incomplete sentence when not being used in an authentic speech (Zhang 1989:158). The concept of 'duanyu' is quite similar to collocations and polywords in FS. A review of Chinese 'duanyu' might shed some light on the identification and categorization of FSs in this study.

Quite a few categories of Chinese 'duanyu' have been proposed by most well known Chinese linguists such as Lü Shuxiang (1982), Ding Shengshu (1979), Hu Yushu (1992), Zhang Bin (1998) and Fan Xiao (1996). Types of 'duanyu' range from three in Lü Shuxiang (1979) to 16 in Fan Xiao (1996). Table II.4 shows all the types that have been proposed and their respective proposers, plus a frequency count of the examples[52] based on a free online Chinese corpus developed by Peking University Center for Chinese Linguistics[53].

Table II.4: a summary of various categories of Chinese 'duanyu' or 'cizu'

	Types of 'duanyu'	Lü	Ding	Hu	Zhang	Fan	Examples[54]	English translation
1	Adjective-Object					✓	大着胆子 (108) 高他一头 (2)	big PRT boldness = with boldness; tall him one head = taller than him by one head
2	Adverbial-Head			✓	✓	✓	积极发展 (1732) 紧张地劳动 (2)	actively develop; concentrate PRT labour = work with concentration
3	Causative verb- Object			✓	✓		让我走 (186) 命令他们转移(0)	let me go; order them to retreat
4	Comparative				✓	✓	鲜花一样 (27) 雷鸣一般 (9)	flowers same = like flowers; thunder same = thunderous
5	Direction & Position			✓	✓	✓	开会前 (85) 山与山之间 (2)	have meeting before = before the meeting; mountain and mountain between = between

[52] All the examples are directly quoted from the works of the five linguists.

[53] CCL corpus (http://ccl.pku.edu.cn:8080/ccl_corpus/index.jsp) contains 477 million Chinese characters as of Jan 23, 2011. Although it is difficult to tell how many words are there in this corpus because, strictly speaking, neither Chinese character nor Chinese 'ci' (normally translated as 'word') can be equated with English 'word' (Zhou, 2004:53-4), based on The Contemporary Chinese Dictionary (1996), one 'ci' is on the average composed of 2.126 characters (Zhou J, 2004:50-1), so we can say that CCL corpus contains 224 million 'ci' or, for simplicity's sake, words. Though smaller than the 323-million word The Bank of English (BofE) corpus, and contains very limited spoken data, CCL is one of the most comprehensive Chinese corpus and will be consulted through out this research.

[54] Numbers in parentheses are the number of tokens of these Chinese 'duanyu' found in CCL corpus.

								mountains
6	Juxtaposed	✓	✓	✓	✓	✓	长江和黄河 (10) 数理化(133) 调查研究 (5508) 伟大而质朴 (0)	the Yangtze river and the Yellow river; maths, physics and chemistry; investigate and research; great and modest
7	Modifier-Head	✓	✓	✓	✓	✓	集体经济 (1929) 谁的钢笔 (0)	group economy = collective economy; who PRT fountain pen = whose fountain pen
8	Number & Quantifier			✓	✓	✓	三个 (35309) 一百零八条 （31）	three units = three; one hundred and eight
9	Particle ‘de’			✓	✓	✓	触目的是 (13) 乌油油的是 (0)	eye-catching PRT is = what caught my eyes is; shining black PRT is = the shining black thing is
10	Particle ‘suo’			✓	✓	✓	所读的书 (23) 所提的意见 (8)	AUX read PRT book = the books I read; AUX raise PRT suggestion = the suggestions I made
11	Prepositional		✓	✓	✓	✓	为人民 (6406) 通过锻炼 (6) 对他 (27697)	for the people; through exercise; to him
12	Appositional			✓	✓	✓	首都北京 (465) 他们俩 (1525)	capital Beijing = Beijing the capital city; they two
13	Repeated					✓	永远永远 (85) 快来快来 (2)	for ever for ever = for ever and ever; fast come fast come = come hurry hurry
14	Serial Verbs			✓	✓	✓	走过去开门 (4) 打电话通知他 (5)	walk over to open the door; make a call to inform him
15	Subject-Predicate	✓	✓	✓	✓	✓	牛是偶蹄类动物 (0) 花开 (1225)	cattle are Even-toed ungulate; flower blossoms = to blossom
16	Verb-Complement		✓	✓	✓	✓	高兴得跳起来 (21)	happy PRT jump rise = to leap with joy; sweep clean;

							打扫干净 (94) 站起来 (6786)	stand rise = stand up
17	Verb-Object		✓	✓	✓	✓	是朋友 (836) 讲故事 (707) 参观工厂 (17)	are friends; tell a story; visit a factory

The following points can be told from the above summary and some extended frequency checks:

1 Some types of 'duanyu', such as Verb-Complement and Verb-Object already exist in literature review of FSs.

2 Structurally most of the 'duanyu' examples are very close to collocations in English, e.g. *积极发展* (= actively & develop), *打扫干净* (= sweep & clean), while some are like polywords, e.g. *数理化* (= mathematics, physics and chemistry: compound noun), *大着胆子* (= with boldness: multi-word adverbial phrase).

3 Although all of the above Chinese 'duanyu' are rule-governed (i.e. formed by productive rules of Chinese language), some of them, e.g. *讲故事* (= tell a story), *花开* (= to flower) and *集体经济* (= collective economy), seem more formulaic (their constituents form stronger partnership or the meaning of the whole does not equal the sum of the constituents) than others, e.g. *参观工厂* (= visit a factory), *牛是偶蹄类动物* (cattle are Even-toed ungulate) and *谁的钢笔* (= whose fountain pen).

4 The major differences between these 'duanyu' and FS are twofold: a) FSs can be either continuous or discontinuous, but 'duanyu' are mostly continuous; b) While FSs take frequency into consideration, Chinese phrases overwhelmingly focus on syntactic structure[55].

[55] Some of the examples, e.g. *伟大而质朴* and *紧张地劳动*, are obviously quite rare, as indicated by their frequency counts in CCL corpus and by intuition, but seem to be included as examples

5 Just like in Wray's FS taxonomy, redundant storage or multiple representation (Wray 2002:251-3, 2008:12-3) can be found in the taxonomies of Chinese phrases. For example, Repeated 'duanyu' contains two Adverbial-Head 'duanyu', e.g. *快来快来* (= come hurry hurry) and Serial Verbs 'duanyu' contains two Verb-Object 'duanyu', e.g. *打电话通知他* (= make a call to inform him).

6 Although in general high frequency 'duanyu' appear to be more formulaic, there are some that are low in frequency but seemingly very formulaic, e.g. *高他一头* (= taller than him by one head)[56], *一百零八条* (= one hundred and eight)[57] and *高兴得跳起来* (= to leap with joy)[58], as well as some that are high in frequency but apparently rule-governed, e.g. *三个* (= three)，*对他* (= to him) and *是朋友* (= are friends). This implies that frequency count alone might not be a sufficient criterion for judging whether a word string is formulaic or not.

7 Not surprisingly, the longer the word string, the lower frequency can be expected, and an additional word/character may lower the frequency by 90% or more. For example, while *打电话通知他* (= make a call to inform him) has only 5 tokens in CCL corpus, *打电话通知* (= make a call to inform) (also a Serial Verb 'duanyu') has 100, and *打电话* (= make a call) (a Verb-Object 'duanyu') has 5931. On the other hand, if an

---

mainly due to their structural characteristics.

[56] Only limited adjectives e.g. *大, 小, 高, 低/矮, 早, 晚* (big, small, tall, low/short, early, late) etc and limited quantifiers e.g. *头, 年, 届, 步* (head, year, school year, step) etc can fit into this ADJECTIVE PRONOUN NUMBER QUANTIFIER structure.

[57] Among the 31 tokens of *一百零八条* in CCL corpus, 26 are serial number 'the 108th' preceding rules, regulations or constitutions, etc. The example seems not selected to mean 'the 108th', but 'the 108 heroes'. When this NUMBER QUANTIFIER sequence is followed by *好汉* (hero), they form a fixed expression denoting the 108 heroes in Water Margin (one of the Four Great Classical Novels of Chinese Literature). *一百零八条好汉* to Chinese is like Robin Hood to people in UK.

[58] Only about 10 verbs or adjectives can fit into the VERB/ADJECTIVE *得跳起来* sequence, and *高兴得跳起来* is the most frequent in CCL corpus.

additional word/character of a given sequence does not lower the frequency too much, it might imply that they frequently coexist and might be a good sign of formulaicity, e.g. *为人民* (= for the people) and *为人民服务* (= serve the people) (6406 vs. 3391 tokens), because the first part strongly implies the second part.

8 By alternating the order of the constituents of some types of 'duanyu', we can get some interesting frequency counts. For example, while *长江和黄河* (= the Yangtze River and the Yellow River) and *黄河和长江* (= the Yellow River and the Yangtze River) have the same frequency (10 tokens each), *调查研究* (= investigate and research) is far more frequent than *研究调查* (= research and investigate) (5508 versus 28). The wide gap in the latter pair might attest to the formulaic nature of *调查研究*, because *调查研究* and *研究调查* are syntactically identical and semantically very close, if not identical[59].

9 Intuitively as formulaic word strings can be found in almost every type, it seems that the above category of 'duanyu' can serve as a source of input or inspiration to create a preliminary list of plausible Chinese FSs, especially those serving as utterance-constituents, such as sentence frames in Lewis (1997) and Phrasal Constraints and Sentence Builders in Nattinger and DeCarrico (1992), polywords and collocations. For example, from *高他一头* (= taller than him by one head) (Adjective-Object 'duanyu'), we may think of sentence frame *NP 高 NP 一头* and collocation *高* & *一头*; from *鲜花一样* (= like flowers) (Comparative 'duanyu'), we may think of sentence frame *NP (好)像 NP 一样*[60] and collocation *(好)像* & *一样*; from *开会前* (= before

[59] Other examples that are found include *所见所闻* and 所闻所见 (551 vs. 16; meaning 'what is seen and heard' and 'what is heard and seen' respectively) and *不理不睬* and *不睬不理* (110 vs. 6; both meaning 'ignore or pay no attention (to somebody)').

[60] *(好)像 ... 一样* is a multiword synonym of *... 一样*.

meeting) (Direction & Position 'duanyu'), we may think of phrasal constraint *在 VP (之)前*[61]; and from *数理化* (= mathematics, physics and chemistry) and *调查研究* (= investigate and research) (Juxtaposed 'duanyu'), we may think of polyword *文史哲* (= literature, history and philosophy), *天文地理* (= astronomy and geography), *陆海空* (= the army, navy and air force)[62] and *风俗习惯* (= customs and habits), etc.

In sum, while there are some noticeable formulaic 'duanyu', there are also some word strings derived from 'duanyu' that seem formulaic. A review of studies in 'duanyu' surely increases the number of types of Chinese FSs that we can use in exhaustive identification of FSs.

II.8 FSs Important for Task-based CSL: Interactional FSs

Among the many types of FSs reviewed earlier, Wray (2002)'s Interactional FSs or their equivalents might be the most attractive to linguists advocating communicative approaches, seemingly because of their usefulness or pragmatic functions in communication. Interactional FSs can play an important role in Communicative Approaches in general and Lexical Approach and Task-based Learning of L2 in specific[63], as these approaches share the same set of principles underlying the the communicative language teaching movement from the 1980s, such as the following:

[61] *在 ... (之)前* is a multiword synonym of *... 前.*

[62] Like *数理化*, the sequence of *文史哲* and *天文地理* are always fixed, and there are far more *风俗习惯* than *习惯风俗 (598 vs. 2)*, and far more *海陆空* (247) and *陆海空* (284) than *空海陆* (2) and *海空陆* (1), while *空陆海* and *陆空海* seem unacceptable.

[63] Proponents of Lexical Approach and Task-based Learning embrace principles underlying Communicative Approaches (Lewis 1993; Willis and Willis 1996a; Nattinger and DeCarrico 1992). As Zimmerman (1997:16) observes, lexicographic research beginning in 1980s brought about a reorientation in language description which 'marked a turning point for communicative syllabus design and language teaching', and this reorientation in language description has led many to reconsider the nature of language and reevaluate the significance of vocabulary in general and FSs in specific. Richards and Rodgers (2001:223) also observe that Task-based Learning has been presented by some of its proponents as a logical development of Communicative Approaches.

- Activitives that involve real communication are essential for language learning.
- Activities in which language is used for carrying out meaningful tasks promote leraning.
- Language that is meaningful to the learner supports that learing process.

(Richards and Rodgers 2001:223)

As reviewed above, Interactional FSs are drawn from real communication, can be acquired to carry out real and meaningful tasks, and, if the tasks are carefully selected to suit learners' needs (see Section V.2.2 below), are meaningful to the learners[64].

When introducing his wellknown 'communicative competence', Dell Hymes advocates that possessing the communicative competence of a language implies possessing both knowledge and ability concerning the four aspects below:

1. Whether (and to what degree) something is formally *possible*;
2. Whether (and to what degree) something is *feasible* in virtue of the means of implementation available;
3. Whether (and to what degree) something is *appropriate* (adequate, happy, successful) in relation to a context in which it is used and evaluated;
4. Whether (and to what degree) something is in fact done, actually *performed*, and what its doing entails.

(Hymes 1979:19)

Again Interactional FSs seem to be perfect units for learning and teaching these four types of knowledge and abilities, especially the 3rd and the 4th as they are typically at speech act or utterance level.

[64] In the same vein Brumfit and Johnson (1979:3) summarize that the Communication Approaches are 'a reaction to the view of language as a set of structures' and 'a reaction towards a view of language as communication, a view in which meaning and the uses to which language is put play a central part'.

In advocating teaching language as communication, Widdowson (1978; 1979) maintians that teachers should pay more attention to the way 'sentences are used in combination to form stretches of connected discourse' (1979:49), and '(t)he key … to the analysis of discouse lies, then, in the understanding of what conditions must obtain for an utterance to count as a particular communicative act' (1979:57)[65]. Along this vein, Interactional FSs (especially those full utterances used in combination) are good candidates to be included in teaching materials because of their communicative value.

In proposing the Lexical Approach, Lewis (1997a) advocates that teachers should, instead of breaking down and analyzing whenever possible, direct learners' attention to FSs which are 'as large as possible' (p3), 'not only possible but highly likely', 'contextualized and stored in our mental lexicon as wholes' (p9) and 'immediately useful' (p34). As a result, *suspicious of people who…* and *relevant to our discussion/problem/needs* deserves more attention than *suspicious of* and *relevant to* (p9)[66]. He maintains that

> A glance at many ELT materials, particularly grammar books, shows that there is a tendency to treat all possible sentences as of equal status. While linguists may be concerned with the possible, language teaching can more usefully direct learners' attention to highly probable examples. The Lexical Approach consciously highlights certain examples as having a special status because they are Fixed, Semi-fixed or prototypical. (Lewis, 1997a:12)[67]

Interactional FSs are exactly what Lewis refers as 'Fixed, Semi-fixed or prototypical' examples with special status to be included in the teaching materials.

---

[65] In Widdowson (1979:57), he seemingly equates his 'communicative acts' with Austin and Searle's 'speech acts'.

[66] Note that *suspicious of people who…* can be turned into *NP be-TENSE suspicious of people who VP* (a Task-Specific Sentence Stem) and *relevant to our discussion/problem/needs* are Sentence Crown Frames/ Gambits.

[67] Coulmas (1981) shares a similar view by stating that 'So prevailing was the study of the undoubtedly essential property of language to allow the production of ever new sentences that the assumption that almost every sentence has an occurrence probability of close to zero was never questioned, much less put to a rigorous test' (p1).

II.8.1 Interactional FSs and Their Subsets

Interactional FSs range from the utterances or frames of utterances learners employed to communicate in meaningful social settings long before evidence of their rule acquisition can be observed (Wong-Fillmore 1976:vii), and those employed by native speakers to communicate in both daily situations and sector-specific speech communities (Kuiper 2004). Different researchers and scholars have used different labels to refer to Interactional FSs, e.g. Routines and Patterns (Krashen and Scarcella 1978), Conversational Routines (Coulmas 1981), Formulas (Ferguson 1981), Fixed and Semi-fixed Expressions (Lewis 1997a; 1997b) and Situation-Bound Utterances (Kecskes 2003)[68], and they can be found in Nattinger and DeCarrico (1992)'s various types of lexical phrases (p65-6). By nature, they all mean very much the same thing. In his Lexical Approach, Lewis prioritizes them as one of the most important[69] (1997b:260), and the focus of this study is their role in task-based teaching and learning of CSL.

Among Interactional FSs, some normally come before the main sentence (e.g. *In my opinion, ...*), some constitute the beginning of the main sentence (e.g. *I am sorry to ...*), some form the skeleton of the main sentence (e.g. *NP be-TENSE sorry to keep-TENSE you waiting*), and there are some preceded by the main sentence like question tags (e.g. *..., is that alright?*). The first type is named Gambits in Keller (1981) but is called Sentence Crown Frames in this study. The second type is named Sentence Head Frames, and the third and forth Task-Specific Sentence Stems (TSSSs) and Sentence Tag Frames.

On top of the above FSs at sentence level, Interactional FSs at discourse level are also proposed by Lewis (1997a) and Nattinger and DeCarrico (1992) under the

[68] Conversation Routines are function-bound while Situation-Bound Utterances are situation-bound, on top of function-bound, so the latter is a subset of the former. For example, *to tell you the truth* is a Conversation Routine that can be used in many situations with the same function, but *Welcome aboard* is a Situation-Bound Utterances used in particular well-framable situations (Kecskes 2003:6).

[69] The other one is Collocations.

names of ‘extended frames’ and ‘skeleton structure’ (see Section II.5.). They are called Discourse Frames in this study.

II.8.2 Characteristics of Interactional FSs

Not all Interactional FSs are syntactically or semantically idiosyncratic. On the contrary, many are grammatical and semantically regular and do not look idiomatic[70]. This even caused debate over whether FSs were correctly identified in Wong-Fillmore (1976) because some ‘free constructions’ were included (Pawley 2007:17). In spite of this, Interactional FSs still have a special status owing to their high frequency and situation-bound use (Coulmas 1981:5). Likewise, many Interactional FSs in Chinese are also grammatically and semantically regular and not neat and tidy in form (unlike vast majority of ‘chengyu’ which are of four syllabus long and some ‘suyu’ idioms which are composed of two phrases of the same length[71]), causing their special status to be easily overlooked, but they are all situation-bound and frequently used, especially in daily life (Zhang 2005:51-2; Ding 2006:10).

Coulmas (1981) also points out that many Interactional FSs are deep-rooted in culture, and the literal meaning and function (i.e. the verbal act it accomplishes) of an Interactional FS ‘are not bi-uniquely mapped on each other’ and discrepancies between Searle’s propositional content and illocutionary force can be frequently observed (p7-8)[72]. This feature is also shared by Chinese Interactional FSs (Zhang 2005:51-2; Ding 2006:10). *吃饭了没有？* (Lit. Have you eaten?), *上哪儿去？*

[70] For example, *May I take your order?* and *Can I help you?* are semantically transparent and syntactically very regular. *你吃点儿什么?* (= What are you going to eat?) and *你买点儿什么?* (= What are you going to buy?) frequently employed by waitresses and shop-keepers in China are also semantically transparent and syntactically regular.

[71] Examples of neat and tidy ‘suyu’ include *车到山前必有路, 船到桥头自然直* (= in the end things will mend), and *穷人一条心, 黄土变成金* (= unit is strength).

[72] Frequently cited examples include expressions beginning with *Can you*, *Would you* and *Could you* etc which are questions in form but requests or commands in function (Coulmas 1981:7). (*Please) help yourself* is functionally an expression to ask quests to take whatever food or drinks they want, but can be literally interpreted by learners as unpleasant ‘nobody else will help you’ (Doi 1981:13).

(Lit. Where are you going?), *哪里哪里* (Lit. Where? Where?), *没什么菜* (Lit. Very limited dishes) and *一点儿小意思* (Lit. A bit of token) are deeply culture-bound[73]. Expressions beginning with *你能不能* (= Can you) are often requests or commands rather than pure questions expecting answers beginning with *我能* (= I can) or *我不能* (= I cannot).

All the above FSs, whether being syntactically/semantically idiosyncratic, culture-bound or with discrepancies between form and function, as long as pragmatic functions can be identified, can be grouped under Wray (2002)'s Interactional FSs.

II.8.3 Identification of Interactional FSs for Task-based CSL

The next question to ask is how such FSs, especially those in spoken language, can be found. In the field of Chinese FSs, lists of FSs (not necessarily Interactional FSs) based on lists of frequently used words have been made (e.g. Zhou 2009) and dictionaries of Interactional FSs have been published (e.g. Chang 1993; Liu & Liu 2005), but they typically contains only a few hundred items and the items are not contextualized nor organized around tasks, and no information on their respective frequency or degree of usefulness is provided[74]. Textbooks with considerable task or function focuses like Kang and Lai (1990) and Shi et al (2010) have been published but were seemingly compiled based on intuition, like many other textbooks.

In the field of ESL, Willis (1990) and COBUILD team collected authentic spoken material by assigning tasks to native speakers and having their oral production during the process recorded as input of textbook and dictionary compilation. Although their textbooks focus more on words, with FSs containing certain words

[73] *吃饭了没有*? and *上哪儿去*? are common greetings like Good morning!, *哪里哪里* is a proper reply to praises, *没什么菜* is used by hosts to show their hospitality (most likely with very good food) and *一点儿小意思* is used even when presenting a valuable gift.

[74] Moon (1998) expresses dissatisfaction with some studies in English FSs because equal status is given to both rare and common FSs, and both obsolete and current ones (p47). Besides, as a result of the lack of spoken data, Interactional FSs 'had distorted frequencies, and were mainly represented in fictional dialogue' in corpuses, such as the one she used in her study (p49).

arranged around those words in their lexical syllabus (Willis 1990:15 & 54; Sinclair and Renouf 1988)[75], the way they collected data is very inspiring. As long as the tasks are properly set, this can be a perfect source of Lewis' 'highly likely', 'contextualized' and 'immediately useful' FSs (1997a)[76], with incorporated 'distributional, formal, semantic, and discoursal information' (Moon 1998:56), which are grouped under certain tasks and more ready to be used.

In order to prepare a reasonable task list for CSL, assessment guidelines provided by authoritative institutions, such as American Council on the Teaching of Foreign Languages (ACTFL)[77], and oral proficiency tests developed in the similar vein, such as Computerized Oral Proficiency Assessment[78] can be consulted, on top of student need analysis. Wu (2008) proposes a pragmatic framework with special reference to CSL in which tasks at different levels along the ACTFL proficiency guideline can be integrated. Once the task list is ready and recordings by certain number of native speakers are obtained, as in Willis (1990), lists of Interactional FSs around each task, especially the pragmatically challenging tasks (Taguchi 2007:131), can be identified, analyzed, categorized, selected, rearranged if necessary, and used in textbook compilation. This is exactly what the author is going to do in this study, except with only one task as an illustration.

Traditionally selection of FSs to be incorporated in textbooks can be based on frequency, range, availability, coverage, learnability and opportunism (Mackey 1965:176; White 1988:48-50; Koprowski 2005). This study is aiming at investigating how tasks can be a basis of selection. We will look at all the FSs identified in general, than focus on one type of Interactional FSs: the Task-Specific Sentence Stems (TSSSs), employed by NNSs and NSs in the same task and

[75] For instance, Polywords *in a way* and *by the way* are taught with the word *way* (Willis 1990:31) and Sentence Frames like *The (adjective) thing is that …*; *The (adjective) thing is to …*; *It's one thing to X and quite another to Y* are taught with the word *thing* (p39). The focus of study is on the commonest word forms, the central patterns of usage of words and the combinations the words typically form (Sinclair and Renouf 1988:148-154).

[76] Whether the FSs extracted from in this way are 'as large as possible' (Lewis 1997a:3) is another issue.

[77] http://onlinelibrary.wiley.com/doi/10.1111/j.1944-9720.1987.tb03269.x/abstract

[78] http://www.cuhk.edu.hk/clc/e_copa.htm#Chinesever

investigate their quantitative and qualitative discrepancies to inform teaching and learning of CSL, especially task-based ones.

Chapter III METHODOLOGY

In this chapter we will look at the research hypothesis and questions, and the detailed steps employed to test the hypothesis and answer the questions.

III.1 Research Questions and Hypotheses

The principal concern of this study is FSs (formulaic sequences), especially TSSSs (Task-Specific Sentence Stems), in spoken Chinese/Putonghua, and the purpose is to gain a deeper understanding of the nature of them in order to support a lexical and task-based approach to teaching Chinese as a Second or Foreign Language (CSL). This is a corpus-based empirical research in which first hand data of a language task was collected, transcribed, coded analyzed through three parallel procedures to investigate the disparities in the use of FSs by non-native speakers (NNSs) and native speakers (NSs).

To operationalize the main research aim stated above, the researcher formulated four quantitative hypotheses to be tested. The first three are of the same purpose: to test if degree of formulaicity (in terms of characters inside FSs as percentage of total characters) correlates with oral proficiency or quality of oral production.

Hypothesis 1. Given the language task, NS data is more formulaic than NNS data, i.e. there are more characters inside FSs as percentage of total characters in NS data[79].

Hypothesis 2. More advanced NNS data is more formulaic than less advanced NNS data, i.e. there are more characters inside FSs as percentage of total characters in more advanced NNS data.

[79] This is to replicate Foster (2001). Moon (1998) also computed this 'density' of FSs which she defines as 'the proportion of words in text that form part of complex lexical items' (p55).

Hypothesis 3. In NNS data, native-like utterances are more formulaic than non-native-like utterances**,** i.e. there are more characters inside FSs as percentage of total characters in native-like utterances.

Hypothesis 4. NSs employed significantly more typical TSSSs (Task-Specific Sentence Stems)[80] than NNSs.

To achieve the research aim, two further questions were asked so as to understand the nature of the FSs, especially TSSSs, in the data with a view to developing a lexical and task-based approach to teaching CSL.

Question 1. What are the FSs in general (or non-TSSS FSs) that are employed by NSs and NNSs for this particular language task and how they differ?

Question 2. What are the TSSSs that are employed by NSs and NNSs for this particular language task and how they differ?

III.1.1 Major Steps of This Study

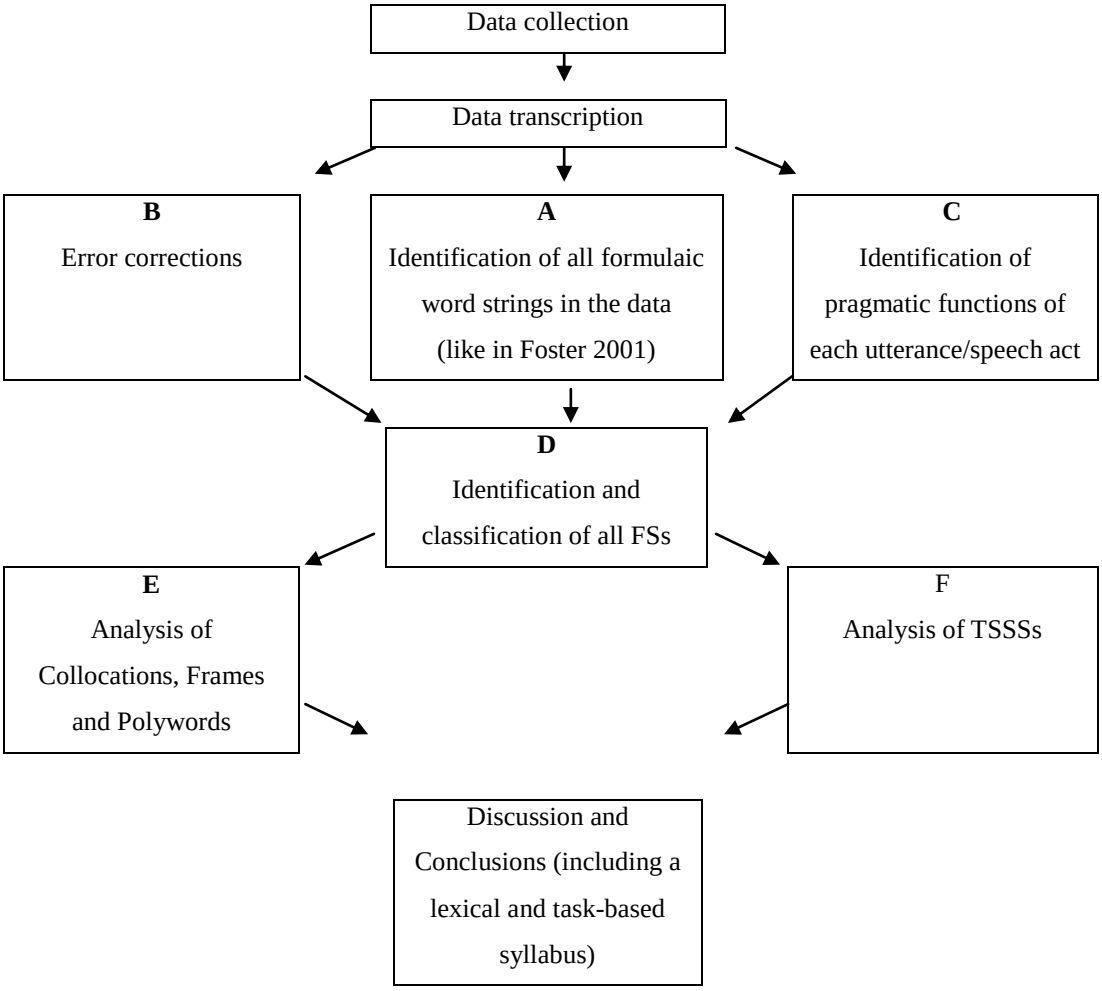
As depicted in Figure III.1, three lines of analysis were carried out in this study. After data collection and transcription, all formulaic word strings in the data were identified by a group of judges as Foster (2001) did and endorsed by Wray (2002:22-3) (Step A). On the other hand, error correction was done by another group of judges (Step B), and speech outputs were segmented by utterance, then the pragmatic functions of each utterance were identified by the author (Step C). Drawing on the result of the above three procedures, all FSs were extracted with reference to the Wray (2002)'s taxonomy of FSs, taxonomies of Chinese FSs and 'duanyu' reviewed in Chapter II[81] (Step D). The FSs were then categorized and

[80] Refer to III.4.3 for definition of TSSSs.

[81] Step A is like the process in which *have the right to*, *take another human life* and *there you go* were identified as formulaic strings in Foster (2001:83). In Step D, the constituting FSs such as *have the right to VP (a Phrasal Constraint)*, *have* and *right* (a pair of Verb-Object collocates), *take* and *life* (a pair of Verb-Object collocates) and *human life* (a Polyword noun) etc would be extracted exhaustively, while *there you go* would be classified as a TSSS.

analyzed (Step E), and, among the FSs, TSSSs (the major concern of this study) were handled separately (Step F). All the above results were than synthesized for implications to inform CSL teaching and learning.

Figure III.1 Flow chart of the research process



III.2 Data Collection

Corpus-based analysis or 'the study of language on the basis of text corpora' is the fundamental research method of corpus linguistics (Aijmer and Altenberg 1991:1). It has been placed at the center of theoretical investigation of language (Halliday 1991:41). As 'a source of systematically retrievable data' and 'a testbed for linguistic hypotheses', corpus can be analyzed to inform various academic fields like lexicography, speech recognition, speech synthesis and machine translation (Leech 1991:9 & 26) and is believed to be essential to the study of formulaic

language (Weinert 2010:2; Altenberg and Eeg-Olofsson 1990). As reviewed in the literature, huge number of corpus-based research has been done on FSs, e.g. Wong-Fillmore (1976), Moon (1998), Aijmer (1996), etc.

Traditionally, especially under the influence of Chomsky since late 1950s, linguistic research relied heavily on researchers' limited personal experience and intuition[82], and it became 'fashionable to look inwards to the mind rather to the outwards to society' (Sinclair 1991:1). Thanks to the unremitting efforts of a group of non-mainstream linguists, i.e. Randolph Quirk, Nelson Francis, Henry Kucera and Jan Svartvik, who set up large scale Survey of English Usage (SEU) Corpus, Brown Corpus and London-Lund Corpus (LLC) respectively from 1960s to 1970s, and the advancement of computational technology (Leech 1991:9), nowadays linguistic studies can be based on corpuses with much greater quantity of raw data to obtain objective evidence (Sinclair 1991:1; Leech 1998: xvi). Though cannot be matched in scale, this study also aims at extracting authentic and objective educational inferences concerning FSs from spoken Chinese corpuses, while written Chinese corpuses will also be consulted.

III.2.1 Sources of Data

The research is based on three sets of data: a learners' corpus consisted of 30 recordings of non-native speakers (NNS) of Putonghua, a compatible native speaker (NS) reference corpus also with 30 recordings, and an online large-scale corpus of written Chinese. The NNS data was taken from a corpus established by the institution the researcher works in, gathered through its Computerized Oral Proficiency Assessment (COPA) project. The NS data was collected by the researcher with similar device, to be compared and contrasted with NNS data. It should be noted that the NNS recordings were collected from authentic test situations while the NS ones were gathered by invitations of voluntary NSs. Both

[82] As Leech (1991) comments, Chomsky's 'view on the inadequacy of corpora, and the adequacy of intuition, became the orthodoxy of a succeeding generation of theoretical linguists' (p8). At the outset of his work on the Brown Corpus, Nelson Francis 'was accused by an ardent transformationlist of embarking on a useless and foolhardy enterprise' because '(t)he intuition of the native speaker, his competence, was considered the only legitimate source of grammatical knowledge.' (Francis 1992:28).

groups took the same test (see below).

III.2.2 Participants

The 30 NNS (11 male, 19 female) were learners of Chinese-as-L2 who had completed a two-year (six-term) language training program, aged from early twenties to late forties. 3 male and 8 female were native English speakers (including 4 ethnic Chinese females), 7 male and 7 female were native Japanese speakers, and the remaining 1 male and 4 female were native Korean speakers. Involvement in the data collection was voluntary. After they had passed the graduation exam of the program, they were offered the opportunity to take this COPA test free of charge and, as a reward, get an additional certificate of result describing their attainments in COPA. The 30 recordings were graded by three certified assessors based on American Council on the Teaching of Foreign Languages (ACTFL) Proficiency Guidelines. 12 of the NNSs got Advanced certificates and 18 got Intermediate certificates[83].

Table III.1 background of participants

	English-speaking	Japanese-speaking	Korean-speaking	Chinese-speaking
NNS	11	14	5	--
NS	--	--	--	30

All the 30 NS (9 male and 21 female) were undergraduate students, postgraduate students or alumni of Chinese University of Hong Kong aged from nineteen to early forties. They were all born in mainland China and use Putonghua as their mother tongue, and, as an incentive for completing the COPA test, each receive HKD50.00 from the researcher.

[83] The grades they got were based on their overall performance in the 15 tasks (see below). Therefore, their performance in the task being investigated in this study might be better or worse than the overall grades.

III.2.3 Procedure

All NNS participants attempted 5 intermediate, 7 advanced and 3 superior level Discourse Completion Tasks (DCTs) in which a conversational turn was provided[84] during the 45-minute computerized oral test and the recordings of the 10th task (speaking with tact), were transcribed and analyzed in this research. The task was at Advanced level with reference to the ACTFL Proficiency Guidelines and the scenarios are as follows:

> While studying at Beijing University, you missed an exam and had arranged with your professor, Dr. Zhang to take a make-up exam this morning. However, you forgot about the make-up and missed the exam. When you remember, you go to Dr. Zhang's office to reschedule the make-up exam. After you hear Dr. Zhang ask you what happened, **try to get your professor to allow you to take a make-up exam at another time**.

While attempting this task, participants were provided with the above written instructions and specifications in their respective native languages, shown on the computer screen, and a simultaneous recording, also in their native languages, for them to listen (about 40 seconds). Both the visual and audio aids were in the participants' mother languages to guarantee perfect understanding of the context and the requirements. Then, 15 seconds was allowed for preparation[85]. Finally, after the participants heard a male voice in Chinese saying *今天上午是怎么回事? 你怎么没来考试呢?* (= What happened this morning? How come you didn't show up for the test?), they had 55 seconds to produce their oral output as a reply to that question, to be recorded by the computer. Time allowed was shown by a countdown timer on the computer screen to remind participants of the remaining time while they prepared and answered.

---

[84] Traditionally a large part of the work on speech act expressions rely on DCTs or similar research tools (Adolphs 2008:9), and inclusion of conversational turns as prompts has been the preferred format (Bardovi-Harlig and Hartford 1992).

[85] Although the recordings collected were not strictly impromptu speech believed by Sinclair to be 'no substitute' as 'a guide to the fundamental organization of the language' (Sinclair 1991:16), the limited time for planning seems likely to ensure that the speech samples are nearly impromptu and much more valuable than quasi-speeches like film scripts and drama texts which are 'considered' language (ibid), because if this happens in reality, the student would have much longer time on the way to ponder what to say before meeting the professor.

When collecting data, due to unexpected difficulties in recruiting NS participants, 26 of them did not sit in the standardized COPA test but only attempted a shorter version handled manually with only 4 tasks, among which the above task was the 3rd in sequence, so that participants could get warmed up by attempting two other questions first. Visual and aural instructions and specifications were in Chinese, the native language of the NS participants.

III.3 Data Transcription

Recorded oral production of both NNS and NS was transcribed into Chinese characters and double-checked by two native-speaking research assistants, who were briefed with the same instructions. Fluency is an associated factor that the researcher take into account but as it is not the focus of the study, pauses were marked manually only at half-second precision level (in the transcription, a '~' indicates a pause of half a second or shorter, and '~~'stands for a pause of more than half but less than one second, and so on). The total number of Chinese characters is 3230 in NNS data and 6038 in NS data. Therefore, it is a reasonably small data set suitable for identification of FSs by intuition (Wray 2002:23) (see Appendix III.1 for a sample of NNS and NS transcript).

III.4 Three Parallel Processes for Identification of FSs

As depicted in Figure III.1, in order to analyze the use of FSs more comprehensively, the following have been done:

a. Identification of all formulaic word strings in the collected spoken data in the same way as Foster (2001), then extract all the FSs i.e. Grammatical FSs, Inferential FSs, Interactional FSs, Memorized and Reflexive FSs, if any[86] (III.4.1);

b. Correction of errors, categorization of mistakenly used FSs identified in the last step based on their type of error, and classification of all

[86] Given the nature of this task, Memorized and Reflexive FSs (Wray 2002) seem unlikely to be found in the data.

utterances by syntactic/semantic and pragmatic quality (III.4.2);

c. Identification of the pragmatic functions of the utterances and, with the result of the above two steps, identify their underlying TSSSs[87] (III.4.3).

The above processes are explained in detail below. It should be noted that because of the importance of data processing for this particular study, the processes are described thoroughly so that subtlety and complexity of the data can be fully demonstrated and appreciated.

III.4.1 Identification of all FSs:

As shown in literature review, FSs are of numerous types stored redundantly (Wray 2002; Lewis 1993, 1997a). Therefore, finding out the number of words inside FSs as percentage of total words in NNS and NS data as in Foster (2001:85) is necessary but we decided to go further. In order to have a deeper understanding of the number, types, varieties, distribution and quality of usage of FSs, this step is to identify all FSs, except for the TSSSs identified in section III.4.3 below.

Formulaic sequences are not easy to define and there is no consensus over many such word strings. Besides, many FSs have very low occurrence and there are too many variations, making it very difficult to identify them only with computer-assisted frequency count if we want to study them exhaustively. As a result, while many researchers query the reliability of applying intuition, it seems that every measure has its inherent limitations and absolutely getting rid of intuition in the whole process is not possible (Wray, 2002, chap 2). This study is not an exemption. In the following sections, intuition plays a very important role, especially in earlier steps to generate a preliminary list of FS candidates. However, several measures were taken to compensate for its weaknesses.

[87] TSSSs are one type of Interactional FSs with huge pedagogical value in task-based teaching and learning. However, due to their length, they might not cross the frequency threshold used to identify FSs in the last step and need to be handled separately.

This section involves the following five steps[88]:

III.4.1.1 Preliminary Identification by Professional Intuition

Identification of formulaic sequences exhaustively in a given text is reported to be a very time-consuming process (Foster 2001:83-4), because, unlike computer-assisted frequency check, researchers might not know exactly what are to be searched, and how many, unless there is an complete list of FSs to be based on. The problem is there is no such a list even in English, the most well researched language, needless to say Chinese. Consequently, professional intuition is exploited to generate the first set of data, so that hopefully most FSs can be identified (ibid:81).

Three native-speaking judges born and raised in mainland China or Taiwan with over 5 years (over 2500 hours) of experience in Teaching Chinese as a Second Language at tertiary level were invited to identify all the FSs in the data. In order to maximize consensus among the judgers, they were given a briefing session and a list of examples of sure Chinese FSs as found in literature review and scrutinized in earlier stages of this study[89], plus a list of instructions as below[90], before they were given a week to identify without consulting anyone else[91]:

1) Highlight formulaic word strings consisting of at least two words. Don't

---

[88] Read and Nation (2004) assert that methodological triangulation is needed in identification of FSs and at least two methods should be employed (p35).

[89] Schmitt et al (2004) come up with a list of target FSs with criteria like previous identification by other researchers, occurrence in textbooks and corpus frequency count. The list of Chinese examples was based on the same criteria.

[90] Both the examples and instructions are in Chinese characters (see Appendix III.2).

[91] Read and Nation (2004) suggest that clear criteria and high level of agreement among at least two judges working independently are needed for reliable identification (p34). The instructions are geared towards this end.

highlight single words[92].

2) If a word string is included in CSL text books as a pattern or frame, it is also a FS[93].

3) If a word string is often drilled as a unit in class, it might be a FS[94].

4) Before you highlight a Phrasal Constraint or Semi-Fixed Expression with slot(s), make sure that the Phrasal Constraint or Semi-Fixed Expression itself contains more than one word, e.g. while *请您* (Lit. beg you = I beg you / Please) in Sentence Head Frame *请您 VP* (= I beg you VP / Please VP) can be a 2-word FS, *请* (=I beg you / Please) in *请 VP* (=I beg you VP / Please VP) is not a FS, despite their semantic and pragmatic similarity.

5) Whether the use of the word string is grammatically, semantically or pragmatically correct is not an issue. Even if a sentence is all wrong or incomplete, as long as there are some strings seemingly formulaic, highlight them[95].

6) If a string is a clear derivation of a sure FS, highlight at least the original

---

[92] In this study, we follow the view that a FS should be at least two words long as proposed by Hickey (1993:32), Moon (1998:8) and Gries (2008) etc.

[93] See Wu et al (2009:3-4). Pawley (2009) also explicitly treat 'abstract (or purely syntactic) grammatical construction' as a formula, except in the notional sense (p7).

[94] This is in line with Wu et al (2009)'s Immediate Chunks (p6-7) and Schmitt et al (2004)'s 'seen as useful to students and worthwhile to teach' criterion (p56).

[95] The author noticed that in NNS data, there are obviously flawed and incomprehensible utterances like *我跟我母亲给我跟会说普通话的人来再说明一下这个情况*(= I and my mom let me and people who can speak Putonghua to come and explain this situation again). This instruction was added to remind the judges that the formulaic parts in mistaken sentences should not be overlooked. In the above example, while the first half seems difficult to comprehend, in the second half *VP 的人* (= *people who VP*) is a Phrasal Constraint that can build a noun phrase, and *会说普通话* is also a phrase in Necessary Topics on language (Nattinger and DeCarrico 1992). Another example came from NS data. The false start (i.e. the underlined part) in 我觉得~ *但并不是我对这次考试不重视* (= I think … but actually it doesn't mean I don't take this test seriously) is a Gambit (Keller 1981) or Sentence Crown Frame in this study.

parts[96].

7) Try your best to judge constantly with the same criteria and finish the judging at one go[97].

8) After the first round of judging, go over the whole transcript one more time, again at one go, preferably on another day[98].

Judges were asked to highlight all the word strings they thought to be FSs. Both NNS and NS transcripts were in Chinese characters, not Hanyu Pinyin (the official phonetic system of modern standard Chinese)[99]. Then, all the results were collated onto a master file for further identification. In Foster (2001), any word strings marked by at least 5 out of 7 judges are counted as FSs. In this study, they are counted as FSs only if marked by all three. The reliability of this study might be lower in this respect, but a few more steps were taken to compensate for this insufficiency and ensure a comparable, if not better, quality.

---

[96] For example, while *I slept like a log* is a FS, *I slept like a twig* which was derived from it is also recognized as formulaic (Wray 2008:113-121). The author decided to handle this in a more conservative way: at least the unchanged parts should be recognized as formulaic (e.g. the underlined part of *I slept like a twig*). Nevertheless, it is noticed that in some cases the derived version might also become very popular and even the changed part(s) need to be counted as well. For example, as of 8 Oct 2011, while *Nothing succeeds like success* has 6,070,000 tokens by Google search, its derived forms *Nothing fails like failure* and *Nothing fails like success* have altogether 3,940,000 tokens, reflecting their tremendous popularity, at least on the internet. Likewise, while Chinese formulaic saying 在哪里摔倒就在哪里站起来 (= no matter where you fall, get up and keep it up) has 5,350,000 tokens, its witty variation 在哪里摔倒就在哪里躺下 (= wherever you fall, just lie down there) which became popular in very recent years also has 1,930,000 tokens. Therefore, whether the amended parts should be counted as formulaic was at judges' discretion in this study.

[97] Intuitively finishing the judging at one go should be easier to exercise the same criteria. On the other hand, the data set of only 9,268 characters (NNS: 3230, NS: 6038) in this study seems to be a manageable size to be done at one go.

[98] The author actually tried to judge by himself as a pilot study and found that judging again on another day helps to reflect on the criteria.

[99] Unlike English, there are no spaces/word boundaries between words in Chinese when written in characters, but there are spaces when written in Pinyin. Pinyin transcripts were not chosen to be used in this study mainly because Chinese characters are much easier for judges to read and make judgments, and because the word boundaries and hyphenation in Pinyin transcript might help to disambiguate words from FSs and affect the validity of judgment.

III.4.1.2 Secondary Identification and Cross Checking based on Characteristics of Chinese

As mentioned in literature review, Chinese FSs share some similarities with their counterparts in English but also have their unique characteristics, as can be expected. All the word strings marked as formulaic were analyzed word by word in accordance with characteristics of Chinese language. For example, the following string in NS07 quoted in section III.4.3.2 below was marked as formulaic by all judges.

请	*您*	*再*	*给*	*我*	*一*	*个*	*机会.*	
Request	you	again	give	me	one	unit of	chance.	(literal English)
I beg you to give me one more chance.								(Edited English)

Based on his own intuition, the author agreed with the judges that this is a highly reusable formulaic utterance when begging for another chance and, as this phrase and its slight variation 请您再给我一次机会 is used by 25 NSs for 39 times, and 4 times by 4 NNSs, it is highly likely to be a fixed Interactional FS. However, besides the underlying TSSS, it might contain smaller multiword strings which might be FSs as well and worth to be singled out for further investigation. Intuition was employed to break down this phrase into the following FSs, with FSs revealed in literature review as reference:

A Sentence Head Frame or Semi-Fixed Expression:

请	*您*	*VP*	
Beg	you	VP	(literal English)
I beg you to VP / Can you please VP			(Edited English)

And a Verb-Object Frame:

NP	*给*	*NP*	*机会*	
NP	gives	NP	chance	(literal English)

NP	gives	NP	a chance	(Edited English)

And a Verb-Object Collocation:

给	*机会*	
give	chance	(literal English)
give	a chance	(Edited English)

And a Measure-word Noun collocation:

个	*机会*	
unit	chance	(literal English)
a	chance	(Edited English)

And a Sentence Core Frame:

再	*VP*	*一*	*个*	*NP*	
again	VP	one	unit	NP	(literal English)
VP	one more	NP			(Edited English)

In this step, all the above five phrases were extracted from the utterance '*请您再给我一个机会*' and were coded as five FS candidates.

Then for each FS candidate found, both NNS and NS transcripts were searched electronically to make sure that all such strings were coded in the same way. This process, though very time-consuming, significantly lifted the internal consistency and, we believe, can largely compensate for the short comings of the last step (only three judges) because there are some strings e.g. *下一次* (= next time) that were, for some reason, marked by three judges in some places as formulaic but only by two in some other places. In this second step, as a rule, as long as a string was once marked by all judges as formulaic in one place, adjustments were made so that the same string was counted as formulaic in all other places. By doing this, we will 'include too much in the

first instance, rather than too little, on the assumption that it is better to examine and discard something than to overlook it' (Wray 2008:4), and definitely increase consistency of judging[100].

Another advantage is worth mentioning here. Purely corpus-derived high frequency word strings might be psycholinguistically invalid, i.e. unlikely to be stored in the mind as wholes (and might not be sensible to be drilled as a unit in teaching) and need to be screened with professional intuition. Examples include, but not limited to, *in addition to the* in Biber et al (1999), *I think this*, *I think I*, *the lecture and* and *I have the* that featured within top 10 in 3-word sequences in some NNSs' data in Adolphs and Durow (2004:117) and *I see what you* and *what I want to* that featured within top 20 in 4-word sequences in NS data in Schmitt et al. (2004:130)[101]. The FS candidates identified in this step tend to be sensible formulaic word strings that can intuitively be viewed as a unit. As a result, even though continuous three-character strings *给我一* (= give me one) or *我一次* (= I one time) seem very formulaic by frequency count, they were not listed as FS candidates for further analysis. On the other hand, discontinuous but sensible collocations like 给 and 机会 (= give & chance) and 再 and 一次 (= more/again & once) were not missed out.

[100] Foster (2001:83) gives a sample marked transcripts as follows (each pair of brackets indicates that the word string was marked as formulaic by one judge, and only those bracketed by 5 judges or above were treated as formulaic):

> ((((((it doesn't matter))))))(((((what the circumstances))))), (((((she didn't have the right to))))) (((((take his life))))). If she was that er emotionally (((((((you know))))))) er distressed, then she should have- ((((((I don't know))))))) (got out of the situation). (((((It's difficult to say))))) when you are not (((((in the situation))))) but (((((((at the end of the day))))))) she did (((((take another human life))))). (((((((There you go.)))))))

It can be noted that many word strings were bracketed 5 times, i.e. not considered as formulaic by two of the seven judges. It seems likely that, with human as judges, the same word strings might be bracketed 5 times in one place but only 4 times in another place. In other words, the same word strings might sometimes be included and sometimes be discarded as FSs in the same study. Inconsistency like this was avoided by this step in this study.

[101] Schmitt et al (2004) make a distinction between these 'recurrent clusters' (which are 'solely corpus-based' and might not be stored holistically in the mind) and FSs, and maintain that the former term 'carries no psychological assumptions' (p128).

All FS candidates found were then coded as Collocations, Frames or Polywords according to the following definitions (see IV.3 for examples):

Collocations: 'the occurrence of two or more words within a short space of each other in a text (Sinclair 1991:170). These words co-exist frequently and might be continuous or discontinuous. Words in the Collocation string are mostly of different parts of speech, and each component still carries its own meaning[102]. Categories are mostly derived from Chinese 'duanyu' as reviewed in Chapter II, plus a few other combinations.

Frames: continuous or discontinuous word strings with slots. Some are sentence frames and others phrase/'duanyu' frames.

Polywords: continuous word strings including 'chengyu' idioms[103] and 'suyu'[104] sayings that function like

[102] We adopt Sinclair's division of idioms and collocations in this study. Although the line between them are not clear, co-occurrences of words are idioms if they are interpreted as 'giving a single unit of meaning', and are collocations if the individual words keep some meaning of their own (Sinclair 1991:172).

[103] Irrespective of the fact that they have quite similar grammatical functions as other Chinese words and are structurally very diverse, Chinese 'chengyu' idioms are traditionally treated as one special type, seemingly due to their fixedness and uniqueness in form (mostly with four characters), their historical origin and their flexibility in use, and most importantly their high frequencies. While well-known 'by hook and by crook' and 'kick the bucket' have zero token in 18 million word Oxford Hector Pilot Corpus (OHPC), and 'by hook and by crook' has barely over 50 tokens in the 323 million word The Bank of English (BofE) corpus (Moon, 1998a:60), 'chengyu' idioms in Chinese are mostly of very high frequency. For example, *千方百计* and *想方设法*, the Chinese near equivalence of 'by crook and by hook', have 3678 and 1457 tokens respectively (as of 23 Dec. 2010) in the 477 million character CCL corpus (as 85% of Chinese words are consisted of 2 to 12 characters, it is pretty safe to say that the 477-million character CCL corpus is smaller than BofE corpus in terms of words). In this study, they are categorized as Polywords, and further subcategorized as adjectives, adverbs and verbs etc, according to their major grammatical functions alone, as illustrated in CCL corpus and dictionaries.

[104] Like 'chengyu' idioms, 'suyu' saying also have pretty high frequencies in corpus. For instance, while its counterpart 'when the cat's away, the mice will play' has zero token in OHPC, *山中无老虎猴子称大王* and its varieties have 12 tokens in CCL.

one single word or a fixed statement[105] (Some component words might not carry substantial meaning).

III.4.1.3 Tertiary Check with Authoritative Dictionaries

The Standard Dictionary of Contemporary Chinese (2004), The Contemporary Chinese Dictionary (2002) and Yingyong Hanyu Cidian (2000) were consulted to make sure that all FS candidates identified so far are multiword strings, but not words. This step served to screen out some fake FSs，i.e. words that were marked by judges mistakenly as FSs, such as *突然间* (= suddenly / all of a sudden)[106] and *这个*[107] (= this / this one).

III.4.1.4 Frequency Count as the Last Quality Assurance

Free online corpus is used to make sure that the FS candidates to be further investigated are real FSs at least quantitatively, even if not qualitatively. Frequency count of each FS candidate was carried out by consulting the online corpus developed by Peking University Center for Chinese Linguistic (CCL) with 477 million Chinese characters. As an arbitrary and expedient decision, any FS candidates with 5 tokens or above were defined as FSs in this study.

---

[105] The concepts of idioms, sayings, proverbs, similes and metaphor etc (Moon, 1998) in English might not have perfect and clear-cut correspondence in Chinese. And the counterpart of an English saying might be a Chinese idiom, and an English simile might have two counterparts in Chinese, one being an idiom and the other a metaphor. For the sake of simplicity, such Chinese expressions are divided into two groups only ('chengyu' idiom and 'suyu' sayings) in this study and they should not be strictly equated with idioms or sayings in English.

[106] In Chinese, *突然* (= suddenly / all of a sudden) is a word and *突然之间* (= suddenly / all of a sudden) is a FS. This might be the reason why *突然间*, which is structurally in between, was viewed mistakenly as a FS.

[107] *这个* means both 'this' and 'this one', and it seems to be structurally composed of *这* (= this) and *个* (a measure word/counter). This might be the source of confusion.

III.4.1.5 Labeling Mistakenly Used FSs

When identifying FSs through the previous steps, quality of word strings, i.e. whether a string is used syntactically/semantically or pragmatically mistakenly, was not taken into consideration (see III.4.1.1). As long as a string seemed formulaic, it was labeled as such. However, after syntactic/semantic and pragmatic errors had been identified in section III.4.2 (see below), FSs identified in section III.4.1.1-4 were revisited and those used mistakenly were labeled as acceptable FSs and unacceptable FSs, and the former was further divided into Likely Choices and Unlikely Choices (see Table III.2 below).

After all the above five steps, a list of scrutinized FSs will be available for further investigation. That being said, it does not mean all the items in the list are indisputable. They are qualified FS in this study only. Whether their identity as FSs or MEUs (Wray, 2008) can be recognized is still subject to the judgment and further procedures of other researcher(s). On the other hand, among those screened out by the frequency check, some might still be very formulaic, but just too rare.

III.4.2 Errors Identification and Classification of FSs by Quality of Use

The reason why mistakes need to be identified in a research focusing on FSs is that when the data transcription had been done, the researcher tried to identify FSs by himself as a pilot study and encountered a serious problem: syntactic, semantic and pragmatic mistakes abound in both NS and NNS data, especially in the latter, as can be expected, and FSs abound in mistakes. If FSs in mistakes are ignored, density of FS in NNS data will be disproportionately lowered and blur the picture, plus the details of the FSs contained in mistakes will be missing. But if FSs in mistakes are included and handled alike, they will blur the picture from another direction. So the final decision was that the FSs in mistakes should be labled as well, and handled differently in analysis (see Section III.4.2.2 below). In order to decide which FSs are in mistakes, mistakes needed to be identified first, and then

categorized in accordance with their nature, i.e. syntactic/semantic[108] or pragmatic, and seriousness, i.e. likely to be used at the same place with some correction or unlikely to be used, to make more useful inferences. And in order to identify mistakes, we need to draw on professional CSL teachers' intuition.

Four major steps were involved in this process.

III.4.2.1 Error Identification and Error Correction

Two other native-speaking judges born and raised in mainland China with over 5 years (over 2500 hours) of experience in Teaching Chinese as a Second Language at tertiary level were invited to identify mistakes for error analysis. They were given a month to finish their work, long enough for them to make corrections on the transcripts.

Both judgers were given a briefing session and a sample transcript which had been corrected for them to follow. A list of instructions was also given as below, before they started to work on the 60 transcripts, without consulting anyone else and without knowing the focus of this study[109]:

1) Make corrections in the way you normally do on assignments and tests of you students.
2) Cross out the mistaken parts and write down corrections next to them.
3) Correct strictly all mistakes of any kind in both NS and NNS data, be they

[108] For practical reasons, in this research, syntactic and semantic errors are handled as one type, as opposed to pragmatic errors. Syntactic and semantic errors including wrong word order, mistaken parts of speech, mismatched word pairs, unclear expressions, etc, were very familiar to the judges as those were the errors they identified and corrected everyday. On the other hand, the latter included the possible but not probable expressions (Lewis 1997) that are correctly generated by grammatical rules but just not preferred by ordinary native Chinese speakers in certain context or under any circumstances, thus need to be replaced by something substantially different.

[109] See Appendix III.3 for the Chinese version distributed to judges.

syntactically, semantically or pragmatically problematic[110].

4) If the whole phrase or even compound sentence sounds are pragmatically not appropriate, replace it with a whole new phrase or sentence that normal native Chinese people would most likely use under that situation[111], along the original line of thinking.

5) The materials to be corrected are transcripts of spontaneous oral test, so false starts such as the two in the first half of 'I would like to er I am very er may I ask you a question' and involuntary redundancies such as the 'I' and 'don't' in 'I I I don't er don't know' and all the 'ah', 'uh' and 'er' can be tolerated and do not need to be corrected.

6) Try your best to correct constantly with the same criteria.

7) Try to finish all the corrections in a few consecutive days.

8) After the first round of correction, go over the whole transcript one more time, again in a few consecutive days.

Then, all the results were collated onto one file for further investigation. As long as one of the two judges made corrections, the corrected part was viewed as problematic and was collated to the master document for categorization of errors.

III.4.2.2 Categorization of FS error types based on corrections

Errors in FSs were first divided into two categories: syntactic/semantic errors

[110] The judges were orally reminded that the corrected version should be of textbook standard to be used in CSL class and no mistakes should be tolerated.

[111] The judges were reminded that mistakes abound also in NS data, because the speech output was not well-planned and almost spontaneous (only 15 seconds was allowed for preparation before recording), and mistakes were to be corrected alike. Pragmatically inappropriate utterances in NS data also need to be replaced with utterances most Chinese would accept and can be used in textbooks as model samples.

and pragmatic errors. For instance, in the following sentence to request for one more chance, while the NNS intended to say *Can I take this test now*, the verb *拿* (= to get/take) is not the proper one that goes with a test[112], and *考这个试* (the native way to express *take this test*) was not used. Consequently the Collocation *拿* and *考试* was labeled syntactically/semantically erroneous and not acceptable[113]:

Error 3.1

(X) *可 不 可以 现在 拿 这个 考试*

Can not can now get this test (Literal English)

Can I get/take away this test (paper) now (Edited English with mistakes)

In the following utterance to give an account of what happened in that morning, the NNS seemingly wanted to express that she overslept and missed the exam (*我睡过头了，错过了这次考试*)[114], but merged the two sentences and deleted something in the middle and resulted in the mistaken Collocation *睡过* and *考试* which was also labeled syntactically/semantically erroneous and not acceptable:

Error 3.2

(X) *我 睡 过 了 你 的 考试*

I sleep pass PRT you PRT test (Literal English)

I overslept your test (Edited English with mistake)

In the following utterance to testify her innocence, the NNS missed out a *是* (= verb to be) in the FS *不是故意的*, constituting a syntactic error. However,

[112] Besides *拿考试*, large number of seemingly fossilized *考考试* (literarily meaning 'take a test') were also found in this research.

[113] The native way of expressing *Can I get this exam paper now?* is *我现在能不能拿这个考试卷?*

[114] The phrases *睡过头了* and *错过了考试* were learnt in the 2nd and 3rd term of the 2-year 6-term Chinese program the NNS participants took.

*故意的*as a FS can be used in the same utterance when *是* is added back, it is classified as a mistaken but Likely Choice (see Table III.2 below):

Error 3.3

(X) *不* *故意* *的.*

Not deliberately PRT (Literal English)

I not that on purpose (VERB is missing) (Edited English with mistake)

The next example is syntactically correct and semantically comprehensible but involves the erroneous use of *有约会* (= have an appointment).

Error 3.4

(X) *我* *知道* *我们* *今天早上* *有* *一* *个* *约会.*

I know we this morning had one unit appointment. (Lit. Eng.)

I know that we had a date/appointment this morning. (Edited Eng.)

While *有约会* undoubtedly means 'have an appointment', it is usually used to talk about a date between lovers or an appointment between friends but unlikely to be between a professor and a student for a makeup test[115]. Consequently the Collocation 有约会 was categorized as a mistaken but Unlikely Choice (see Table III.2 below).

The most challenging task in the above process is the great difficulties in judging whether a mistaken FS functioning as a part of an utterance is pragmatically or semantically/syntactically wrong, because the utterance might be problematic in both respects. After careful consideration, the following simplistic categorization is adopted, as shown in Table III.2:

[115] CCL online corpus was consulted and it was found that the first 10 tokens of '*约会*' are all about dating a lover, either literally or metaphorically, and among the first 10 tokens of '*NP 跟 NP 约会*', 6 are definitely between lovers and 4 are with friends. The above utterance was replaced by a judge with *我知道我们本来今天早上约好了要考试* (= I know we had an appointment to do the makeup test this morning).

Table III.2 Categorization of problematic FSs

Acceptability	**Acceptable FS**		**Unacceptable FS**
Usage	**Likely Choice (reused by judges or likely to be used here by the author's intuition)**	**Unlikely Choice (not reused by judges and seems unlikely to be used here by the author's intuition)**	**N.A.**
examples	*故意的* (as in Error 3.3 *不故意的*);	*有约会* (as in Error 3.4 *我知道我们今天早上有一个约会*)	*拿考试* and *睡过考试* (as in Error 3.1-2; *non-existing VERB OBJECT collocation*

Problematic FSs were divided into two types: the Acceptable and the Unacceptable, and those in the former type were further divided into the Likely Choice group and the Unlikely Choice group. While the Likely Choice group can be used in the original utterance to express similar ideas after some corrections, as indicated by the judges' notes or by the author's native intuition and knowledge as a CSL teacher, the latter seems difficult or even impossible to be reused, unless employed to deliberately express something inappropriate for the occasion.

III.4.2.3 Categorization of Utterance based on Corrections

While the last section deals with components of utterances, this section focuses on utterances as a whole. Based on the corrections judges made, all utterances were labeled as one of the following 3 types of word strings (not necessarily neat sentences): Nonnative-like (NNL) utterances, Native-like but Untypical (NL-UT) utterances, and Native-like and Typical (NL-T) utterances[116].

III.4.2.3.1 Nonnative-like (NNL) Utterances

Utterances under this category can be further divided into Nonnative-like but Comprehensible (NNL-C) utterances and Nonnative-like and Incomplete (NNL-I) utterances.

[116] In this study, when categorizing utterances into NNL, NL-UT and NL-T, native-likeness equals being grammatically correct and semantically comprehensible, and only NL-T are those utterance that are typically used by native speakers in tackling the language task at issue.

A Nonnative-like but Comprehensible (NNL-C) is an utterance with obvious mistakes such as serious lexical mistakes, applying L1 grammatical rules and using L2 grammatical rules too creatively or in nonnative ways, but the meaning or intention can still be adequately conveyed. For example, in order to plead for another chance, one NNS provided the following foreign-like but grammatically largely correct utterance which is comprehensible but unlikely to be used by native-speaking Chinese probably under any circumstances:

Error 3.5 (NNS01)

(X) *可不可以 再一次 约 好 这个 考试 的 时间*

Can cannot once again fix properly this test PRT time (Lit. Eng.)

Can (we) once again arrange properly the exam time or not? (Edited Eng. with mistakes)

On the other hand, a Nonnative-like and Incomplete (NNL-I) utterances is an utterance with inappropriate syntactic structure and/or insufficient and/or disordered information and the message and function need to be unscrambled from the context. For example, when begging for another chance, another NNS produced the following utterance seemingly to show that he is desperate and willing to accept discounted marks in exchange for another chance:

Error 3.6 (NNS29)

(X) 你 差 我 的 扣分 我 也 不能 说.

You bad I PRT deduct points I also cannot say. (Lit. Eng.)

You bad my deduct points I also can't say. (Edited Eng. with mistakes)

III.4.2.3.2 Native-like but Untypical (NL-UT) Utterances

Most utterances under this category are syntactically correct and semantically comprehensive, but were not accepted by one or both judges in this study, seemingly because they are not typical utterances in handling the language task at issue (i.e. not suitable as exemplary speech acts in textbook). Some other utterances under this category are pragmatically inappropriate (likely to

be used by native speakers in circumstances other than the task involved in the research). The examples below belong to this type.

The following utterance was deployed by an NNS to explain how he forgot to take the test. After saying *我生病了，而且每天都要考试* (= I am sick and there are tests everyday), he added

Error 3.7

(X)	*我*	*就*	*忘记*	*了*	*嘛*	
	I	then	forget	PRT	PRT	(Literal English)
	Then I just forgot (you know)					(Edited English)

The whole utterance is syntactically perfect and semantically clear, but it is normally used to emphasize that the speaker is not to blame, or what had happened is reasonable, mainly because of the *嘛* in the end[117] which turns the whole sentence pragmatically inappropriate, or at least untypical to be used in a textbook as input or model, for the language task at issue[118].

The next utterance is used by a NS, as part of his request for another chance to retake the missed exam, to convince the professor that he should be forgiven. After *能不能再给我一次机会呢*? (= Can you give me one more chance?), he added

Error 3.8 (in NS07)

(X)	*因为*	*每*	*个*	*人*	*都 有*	*错误*	*的*	*时候.*
	Because	every unit		people	all have mistaken	PRT	time	(Literal English)

[117] *嘛* is a Chinese particle articulated with a low tone and used in the end of a sentence to indicate that something speaks for itself as in *这也不能怪他，头一回做嘛* (= He's not to blame. After all, it was the first time he did it) (The Contemporary Chinese Dictionary 2002:1292).

[118] The above utterance was replaced by a judge with *所以我今天早上忘了考试* (= That's why I forgot the test this morning).

Because everybody makes mistakes (Edited English)

The reason why this utterance was not accepted by judges seems understandable. If a student just missed the makeup test especially arranged for him/her and the third chance is still wanted, this utterance may demonstrate that the student has little regret and might make the professor unhappy[119].

But how could a well-educated Chinese use inappropriate words like this? It seems that the difference in the motives of NS and NNS participants is worth mentioning here. Unlike NNSs, when the NSs participated in the study, they did not aim for a certificate or other academic rewards. Given the negligible financial reward of HKD 50, there was the possibility that some NSs participated for fun and were not as serious and cautious as NNSs who knew that whatever they said was going to affect their certificate. Therefore, on the one hand NS data was put to the test with the same criteria in this study, with pragmatically inappropriate utterance classified as such. On the other hand, we might need to interpret NS mistakes with care because it might not reflect the reality.

If we take *因为每个人都有错误的时候* (= because everybody makes mistakes) as an example, we can imagine that while the utterance might be irritating to the professor, it is not impossible to be used in real life (for instance, if the student knows that the professor gave many more chances to his favorite students who made even worse mistakes, or if the professor himself made many serious mistakes before). However, in a CFL textbook for general purposes, extreme-case utterances might not be preferred by judges who are also experienced teachers.

On top of the above, there are also some utterances that sound native-like but

[119] This can be reflected in the judges' corrections. One judge replaced the sentence with *我真的是很对不起您* (= I am really sorry for this), and the other simply deleted it and provided no replacement, even though they were asked to do so, seemingly because it is not appropriate to defend in this vein.

not well organized probably due to the pressure in spontaneous oral production[120]. For example, after asking for another chance, NS21 produced the following utterance to reprove his wrongdoing, but with obviously too many false starts that makes the whole utterance unfocused and vague, and consequently a merely acceptable utterance. This type of Native-like but Untypical (NL-UT) utterances are not ideal to be included in teaching materials as exemplary speech acts, though might be good for listening comprehension exercises:

我知道这个可能~可能就是说本来~第一次考试我也是因故缺考然后~安排了~嗯~也安排了补考~因为我很~就实在是很不好~

I know this maybe~maybe you know originally~first test I also somehow missed and then~arranged~uh~also arranged makeup test~cauz I very~and really is very not good (Lit. Eng.)

I know this maybe ~ maybe you know in the first place~somehow I missed the first test and then~it was arranged ~ uh ~ the makeup was also arranged ~cauz I was very~well it's really bad (Edited Eng.)

III.4.2.3.3 Native-like and Typical (NL-T) Utterances

Finally we have the typical native-like utterances with no or very minor mistakes. Most of NL-T utterances are native-like utterances that can be used by almost any participants as part of their production and, with very minor modification such as deleting false starts and redundant words, can serve as exemplary speech acts in CSL teaching materials. Below is an example produced by a NS (underlined is the part that can be deleted to make the utterance more like a model speech act of requesting in textbooks):

请求 你 给 我~ 再 给 我 一次 机会 补考

Beg you give me ~ again give me one chance make up the exam. (Lit. Eng.)

I beg you to give me ~ give me another chance to do the make-up exam. (Edited Eng.)

[120] In a way these utterances can demonstrate that the speech samples collected in the study are quite impromptu (see Section III.2.3).

III.4.3 Identification of Utterance Functions and Underlying TSSSs

In order to direct learners' attention to word strings which are 'as large as possible', 'not only possible but highly likely', and 'immediately useful' (Lewis 1997a), and to efficiently facilitate task-based CSL, it appeared certain to the researcher that one step further needs to be taken, as most of the FSs identified in III.4.2 are not ready for carrying out speech acts involved in the language task at issue. In order to enable a learner to complete a language task, Interactional FSs at speech act level need to be extracted from corpuses and provided as input or exemplars.

In the field of corpus linguistics, large-scale corpuses can provide huge number of Interactional FSs which have been actually used. For example, Aijmer's Conversational Routines in English (1996) analyzed Interactional FSs with pragmatic functions like thanking, apologies, requests and offers, and those with discourse-organizing function, i.e. discourse markers or conversation gambits, identified in London-Lund Corpus of Spoken English. Research in this regards helps to display macro pictures of how certain types of Interactional FSs are actually used[121]. However, research for pedagogical purposes might need to take another path, because with dozens or hundreds of thanking FSs identified in corpuses, L2 learners might not know which one to use under certain circumstances for certain language tasks, and do not know what the other Interactional FSs they can employ to use with the thanking FSs they choose, as a formation or discourse, to complete the task. For pedagogical purposes, the researcher saw the need to coin the term Task-specific Sentence Stems (TSSSs) to refer to Interactional FSs that target on a specific language task.

[121] For instance, in Aijmer (1996) chapter 2, strategies, continuation patterns, grammatical aspects, prosody and fixedness, distribution over different texts, sentence stems and their extensions, pragmatic functions, discourse function (e.g. as closing signals), pragmatic frames of Thanking FSs are analyzed in depth.

III.4.3.1 Definition of TSSSs With Reference to Related Terms

As the concept TSSS overlaps with FS, Interactional FS and 'lexicalized sentence stems', it might be helpful to distinguish them before we go further.

A TSSS is defined as the core content-bearing elements of an utterance-level multi-word speech act that carry the main message. It can serve as an exemplary utterance or utterance-frame in language teaching. It is a fuzzy-edged concept borrowed from 'lexicalized sentence stems' in Pawley and Syder (1983), overlapping with FSs as revealed in literature review, coined as a handy term when studying the exact wording used to realize the speech acts[122] in this study.

III. 4.3.1.1 TSSSs versus FSs

It should be noted that while TSSSs are extracted from speech acts, many TSSSs and their embedded speech acts might not be counted as FSs if frequencies are taken into account[123]. However, in view of the value of these low-frequency speech acts and their TSSSs in second language teaching and learning, they are all analyzed in this study as plausible FSs, as long as they are nativelike and suitable to be included in a CFL textbook[124].

Besides the frequency and nativelikeness issue, there are two more differences between TSSSs and FSs:

1. TSSSs are utterances or utterance frames, while FSs can be constituents of

[122] In this study, we borrow Pawley (2009)'s (who in turn follows Austin (1962) and Searle (1969)) definition of 'speech acts' which broadly refers to utterances performing prototypical discourse functions other than referring and predicating. These speech act utterances are normally formulaic expressions bound to particular discourse contexts and particular discourse functions (p6).

[123] It can reasonably be stipulated that actually the vast majority of speech acts and their TSSSs are of very low frequencies. As shown in Section II.7.1, an additional syllable/character can lower the frequency of a word string by 90%, and most speech acts found in this study (and probably in most authentic speech events) are unlikely to have 5 occurrences in the mini spoken corpus being investigated (see Section III.4.1.4 above).

[124] According to some researchers, Wray (2008 Chapter 9) for instance, mistaken word strings produced by learners can also be FSs. However, the core content-bearing elements in them are not counted as TSSSs in this study because they cannot serve as an exemplary utterance-frame in language teaching.

utterances (i.e. Grammatical and Referential FSs), and can also be at utterance level (i.e. Interactional and Reflexive FSs) and text or discourse level (i.e. Memorized FSs).

2. All TSSSs have a clear pragmatic functional orientation, but not all FSs do. TSSSs only make sense when we are talking about a certain specific speech act. In this sense, TSSSs are part of Interactional FSs.

III. 4.3.1.2 TSSSs versus Interactional FSs

In terms of functions, Interactional FSs can either be very broad (e.g. Sentence Heads) or very specific, while TSSSs tend to be more specific. For example, Sentence Heads like *Would you (mind) …?* or *May I …?* (Nattinger and DeCarrico 1992:62) signify broader 'requesting' functions, while TSSSs are used to signify more specific or concrete ones such as 'requesting for a birthday gift', 'requesting for an opportunity' or 'requesting for something to be done', etc, with *Would you mind buying a NP for NP's birthday?*, *Would you kindly give NP another chance to VP?* or *May I have my NP ready as soon as possible?*. As a result, theoretically number of TSSSs can be expanded indefinitely (Pawley 2009:8).

III. 4.3.1.3 TSSSs versus Lexicalized Sentence Stems

TSSSs are the closest to the lexicalized sentence stems coined by Pawley and Syder (1983:208-215). Neither TSSSs nor lexicalized sentence stems are productive grammatical rules, as they contain '(lexical) elements which are not inserted by means of lexical rules' (Aijmer 1996:22). However, while lexicalized sentence stems are all utterance frames with slots[125], TSSSs include both utterance frames and highly stereotyped utterances that can hardly be altered, e.g. *Merry Christmas* (Ferguson 1981:25)[126]. Moreover, TSSSs can be more tasks-specific[127].

---

[125] For instance, the lexicalized sentence stem of *Mr. X is sorry to keep you waiting all the time* is *'NP be-TENSE sorry to keep-TENSE you waiting'. Mr. X, is* and *keep* are inflections of the lexicalized sentence stem, while *all the time* is its extension (Pawley and Syder 1983:210).

[126] Very fixed speech acts like *How do you do?* can be viewed as a TSSS without slots to be filled.

[127] While *'NP be-TENSE sorry to keep-TENSE you waiting'* is a lexicalized sentence stem, it can also be a TSSS under circumstances when the 'NP' can be realized with more than one

III.4.3.2 Identification of TSSSs

As most of the TSSSs cannot pass the frequency threshold because they are long and not uniform, they were extracted with separate procedures as described below.

III.4.3.2.1 Identification of Utterance Functions

Firstly, the functions[128] of all utterances (e.g. 'alerting' and 'apology', etc.) were identified according to the roles they play in the discourses, with reference to categories of communicative function in van Ek et al (1975:11-12) and Wilkins (1976:p41-54), and Austin's classes of utterances (1962:151-164). The shortest output of NSs (NS05) which was divided into 10 utterances is pasted below as an example:

1. (overt apology NS05a) *啊教授~不好意思*(= Er professor ~ I am sorry)
2. (excusing-detail NS05a) *我今天早上因为~家里发生了一点突发事件~嗯~* (= This morning I ~ because something urgent happened in my home ~ er ~)
3. (excusing-detail NS05b) *我的外公突然生病了~* (= My grandpa suddenly got sick)
4. (excusing-detail NS05c) *然后我要送他到医院去~* (= Then I had to take him to hospital ~)
5. (excusing-forgot NS05a) *所以~~嗯忙起来就突然间忘记了~补考的事情~嗯~* (= So ~~ er was so occupied and suddenly forgot ~ about the makeup test ~ er ~)
6. (requesting for another chance NS05a) *请求你给我~再给我一次机会补考~* (= I beg you to give me ~ give me another chance to do the makeup ~)
7. (asking for forgiveness NS05a) *嗯对你带来的不便~嗯希望您能原谅~* (= Er for the inconvenience caused ~ er I hope you can forgive me)

---

interchangeable options (for instance, when both 'I' and 'We' can fit). The following are similar TSSSs to be used in other tasks when the subjects and objects are more specific:

'I be-TENSE sorry to keep-TENSE you waiting'

'We be-TENSE sorry to keep-TENSE them waiting'

'Mr. X be-TENSE sorry to keep-TENSE him waiting'

'We be-TENSE sorry to keep-TENSE her waiting' etc

[128] Functions in this study refers mainly to the illocutionary force discussed in Austin (1962) and Searle (1969, 1979). See Adophs (2008:22-3) for a discussion of the notion 'function'.

8. (overt apology NS05b) *真是不好意思~~* (= I am so sorry ~~)

9. (vowing NS05a) *我向您保证~下一次的补考我一定不会忘记的~嗯~* (= I pledge ~ to show up in the test next time)

10. (requesting for another chance NS05b) *请求您再给我一次机会吧*(= I beg you to give me another chance)

There are some long utterances containing two or three small utterances denoting different functions (the functions might be of the same type or different types) but share the same sentence head, like in the example below:

> 看你能不能就说~能原谅我这次又又缺考~嗯~然后再给我安排一次~嗯~补考的机会(= I would like to see if you know ~ if you can forgive me for missing the test again again ~ er ~ and then give me another chance ~ er ~ to take the test)

In such cases, the small utterances were singled out and the sentence head was duplicated so that both small utterances denote a discrete and complete function or speech act for further analysis:

> (asking for forgivenessNS21a) 看你能不能就说~能原谅我这次又又缺考~嗯~ (= I would like to see if you know ~ if you can forgive me for missing the test again again ~ er ~)

> (requesting for new appointment NS21b) 看你能不能就说...然后再给我安排一次~嗯~补考的机会 (= I would like to see if you know ~ if you can … and then give me another chance ~ er ~ to take the test)

Then the utterances of the same function, i.e. a list of possible realization of a certain speech act, were grouped together, and then similar groups of utterances were grouped under a bigger category for further analysis.

III.4.3.2.2 Identification of Underlying TSSSs of Each Utterance

From the conversational routine *I am sorry to have kept you waiting* Pawley and Syder (1983) elicited the underlying lexicalized sentence stem *NP be-TENSE sorry to keep-TENSE you waiting*. Likewise, from the following utterance in NS 07:

请　　您　再　给　我　一　个　　机会.

Request　you　again give　me　one　unit of　chance.　(literal English)

I beg you to give me one more chance.　(Edited English)

we got the following TSSS:

请　　NP　再　给　NP　一　个　　机会.

Request　NP　again give　NP　one　unit　chance.　(literal English)

I beg NP to give NP one more chance.　(Edited English)

III.4.3.2.3 Identification of Underlying TSSSs of Each Type of Utterance

Together with other utterances of similar form and function (such as the 10th utterance of NS05 quoted above in **III.4.3.2.1**), the TSSS can be modified as below with broader coverage.

请（求）　NP　再　给　NP　一个/次　机会　　（吧）.

Request　NP　again give　NP　one unit　chance　(PRT).　(literal English)

I beg NP to give NP one more chance.　(Edited English)

III.5 Summary of the Chapter

After the above three parallel processes, we got a list of frequency-checked Collocation, Frame and Polyword FSs, with specifications on whether they were used correctly, used mistakenly but can be used in the original utterance to express similar ideas after some corrections, or difficult or even impossible to be reused, unless employed to deliberately express something inappropriate for the occasion. We also obtained a list of non-frequency-checked TSSSs derived from Native-like and Typical (NL-T) utterances. These two groups of FSs will facilitate the data analysis in later chapters which form the basis for the development of the lexical and task-based approach to teaching CSL.

Chapter IV DATA ANALYSIS

This chapter presents the FSs identified in this research, a contrastive analysis of FSs used by NNSs and NSs, and the analysis directed against the four quantitative hypotheses in the beginning of Chapter III.

Some general findings for an overview of the collected NNS and NS data will be presented firstly, followed by a section on non-TSSS FSs. Task-Specific Sentence Stems (TSSSs) will be handled separately in the last section of this chapter because of their special status in task-based learning (TBL).

IV.1 General Quantity and Quality of NNS and NS Oral Production

NNSs and NSs produced 3230 and 6038 syllables (or Chinese characters) respectively, resulting in a ratio of 1：1.9. If syllables without concrete meaning such as *ah*, *oh* and *en* etc. are deleted, it is 2696 versus 5796 characters, resulting in a ratio of 1:2.15. Therefore, it is quite reasonable to say that meaningful production by NNSs is about half of NSs. Higher speed or fluency of NSs can account for the majority of the difference, and the fact that while NNSs spoke 46 seconds on the average, NSs spoke 50 seconds. T-test result in Table IV.1 below shows that NNS production is significantly lower than NS ($p < 0.001$), with an average of only 107.7 syllables, 92 fewer than NS.

Table IV.1: t-test result comparing average number of syllables/characters produced by NNS and NS

Gp 1	Gp 2	Avg1	Avg2	Std 1	Std 2	n1	n2	CL (%)	Z-value	T-Stat	Implication
NNS	NS	107.7	201.3	46.9	39.5	30	30	99.9%	3.2368	8.2244	NS > NNS

As shown in Section III.4.2.3, all utterances were classified as Native-like and Typical (NL-T), Native-like but Untypical (NL-UT) and Non-native-like (NNL). Table IV.2 below shows the percentages of the above three types of utterances in NNS and NS data and their t-test comparisons. The following observations can be made:

1. As can be expected, NNS produced significantly less NL-T ($p < 0.001$), but more NL-UT ($p < 0.05$) and NNL ($p < 0.001$) utterances. However, almost half (47.5%) of NNS production is NL-T and more than a quarter (26.8%) is NL-UT, demonstrating that three fourth of the NNS production is largely grammatically correct.

2. Even 16.1% of NS utterances are NL-UT. This might be a result of the cognitive complexity and communication stress (Skehan 1998) involved in this pragmatically complex language task being studied. It might also be speculated that when speaking spontaneously, people are not always skillful with words, and some people are less skillful with words than others.

Table IV.2: t-test result comparing average number of characters produced by NNS and NS classified as NL-T, NL-UT and NNL word strings

Gp 1	Gp 2	Avg1	Avg2	Std 1	Std 2	n1	n2	CL (%)	Z-value	T-Stat	Implication
NL-T in NNS	NL-T in NS	47.5%	83.8%	0.24	0.21	30	30	99.9%	3.2368	6.2129	NS > NNS
NL-UT in NNS	NL-UT in NS	26.8%	16.1%	0.18	0.21	30	30	95.0%	1.6716	2.0901	NNS > NS
NNL in NNS	NNL in NS	25.8%	0.0%	0.19	0.00	30	30	99.9%	3.2368	7.1525	NNS > NS

IV.2 Collocations, Frames and Polywords (or non-TSSS FSs[129])

This section is to tackle the 1st question asked in the beginning of Chapter III:

[129] Non-TSSS FSs include Collocations, Frames and Polywords, Non-TSSS FSs were exhaustively identified while not all TSSSs are extracted (see Section VI.3 for details). Besides, as TSSSs typically contain many words, most of them are unlikely to pass the frequency threshold to be qualified as FSs (see Section III.4.1.4). This is part of the reasons why only numbers of non-TSSS FSs are compared in this section.

What are the FSs in general (or non-TSSS FSs) that are employed by NSs and NNSs for this particular language task and how they differ?

As specified in Section III.4.1.2, the FSs in general can be divided into Collocations, Frames and Polywords in terms of forms. In this research, they are under the expedient cover term 'non-TSSS FSs', for convenience sake, to be distinguished from TSSSs in the next section.

IV.2.1 Non-TSSS FSs Identified in NNS and NS data

Non-TSSS FSs range from formulaic utterance fragments (Wong Fillmore 1976:718) to conjunctive FSs connecting utterances. Functionally, they can be Grammatical, Referential or Interactional FSs[130] (see Section II.5.1).

Pedagogically, in production exercises for instance, while TSSSs can serve as ready-to-use exemplar speech acts in, say, task-based learning (TBL) to deal with a task, non-TSSS FSs analyzed below are constituents of TSSSs and may serve as multiword ingredients in creating novel utterances or discourses, as well as serving as inputs for form-focused exercises in TBL (Willis and Willis 1996a).

Lewis (1997a) asserts that chunking (i.e. the ability to discern the constituents of a text) is central to effective communication and efficient acquisition (p58). In his Lexical Approach, Lewis (1997a) stresses the importance of raising the awareness that language is composed of various types of FSs (p45), the importance of noticing and understanding the FSs in a text (p55)[131], and the importance of efficiently recording and using the FSs (p53-4)[132]. The findings below might serve as illustrations of the results of a chunking operation as described in Section III.4.1.

[130] No Memorized and Reflexive FSs are found in this study.

[131] Also see Lewis 2000b:158-163.

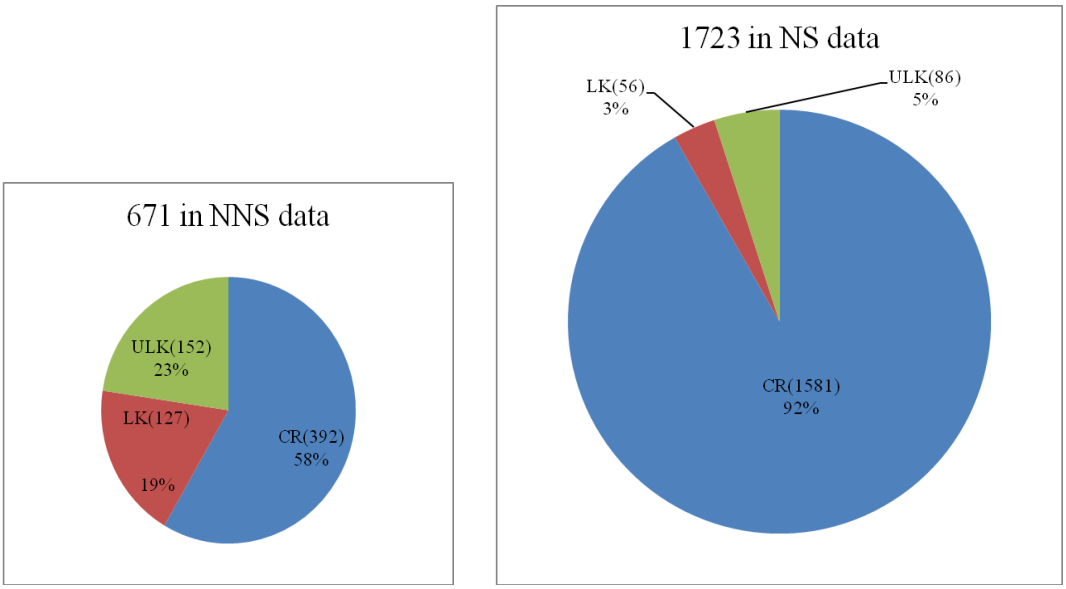
[132] Awareness-raising or noticing of FSs, among other language elements, is also advocated by task-based learning proponents (Richard and Rodgers 2001:236). Also see Schmidt (1990:145), Wills J. (1996:58) and Willis and Willis (1996b:68).

Below we will present what the non-TSSS FSs employed by NNSs and NSs for this particular language task are, how they differ and how they can infer teaching and learning of CSL.

IV.2.1.1 General Findings

As reported in Section IV.1, NNS and NS data contain 3230 and 6038 characters (ratio: 1:1.9). On the other hand, the number of non-TSSS FSs extracted from the NSS and NS data are 671 vs. 1723 (ratio: 1:2.57), indicating a lower density of FSs in NNS data. Two other important ratios denoting the quality of use can be derived from Figure IV.1a below (or the last line of the Appendix IV.1). They are the overall percentage of mistakenly used non-TSSS FSs (sum of LK and ULK[133] in the figures below): 42% in NNS data (= 19% + 23%) and 8% in NS data (= 3% + 5%).

Figure IV.1 (data from Appendix IV.3.1) Overall quality of non-TSSS FSs

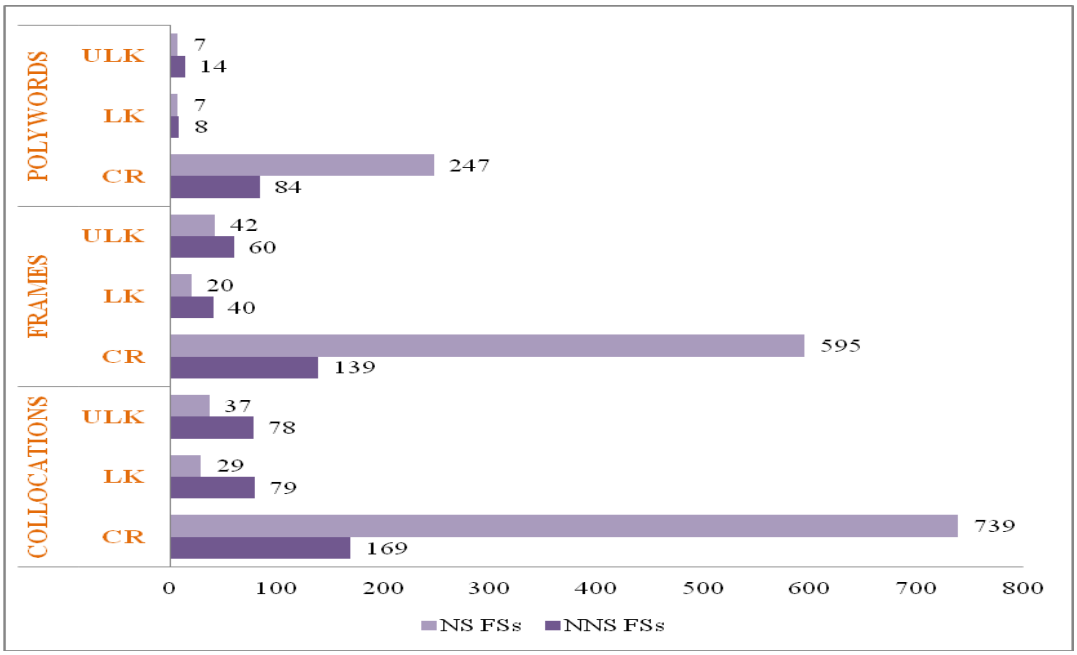


When broken down by three formal types, i.e. Collocations, Frames and Polywords, the quality of non-TSSS FSs in NNS and NS, i.e. if they are ULK

[133] LK and ULK are genuine FSs used incorrectly. They stand for Likely Choice (FSs mistakenly used but are likely to be used at the same place with some correction) and Unlikely Choice (FSs mistakenly used and are unlikely to be used at the same place). On the other hand, CR stands for 'correctly used'. See Section III.4.2.2 Table 3.2 for examples.

(mistaken and unlikely), LK (mistaken but likely) or CR (used correctly) are as follows.

Figure IV.2 (data from Appendix IV.1) 3 formal types of non-TSSS FSs & their quality



The following can be observed from above:

1 While NNSs used less FSs and made more mistakes (i.e. ULK + LK), the proportions of CR, LK and ULK across Collocations, Frames and Polywords are quite similar in both NNS and NS data: CR being the majority, followed by ULK, then by LK. More research is needed to confirm if this is a reflection of the NNSs' high proficiency as the result of a 2-year intensive training.

2 Nearly half of FSs identified are Collocations, followed by Frames and then Polywords. Again, this holds true for both NNS and NS data. This might to a certain extend reflect the natural composition of spoken language, though, again, more research with wider range of language sample is needed before any valid conclusions are drawn.

3 Among the three formal types of non-TSSSs FSs, Polywords have the least quantity but the best quality: 84 or 79.25% of NNS Polywords and 247 or 94.64% of NS Polywords are used correctly (compared with 139 or 51.84%

and 169 or 58.16% of NNS Frames and Collocations, and 595 or 91.80% and 739 or 90.56% of NS Frames and Collocations). It seems appropriate to speculate that it is because Polywords such *不好意思* (= sorry), *真的* (= really), *今天早上* (= this morning) and *加班加点* (= work overtime) etc are conceptually and functionally (if not formally, phonologically and syntactically) simple and are more fixed linguistic entities than Collocations and Frames. It also seems appropriate to infer that Collocations and Frames are pedagogically more challenging.

The following three pie-charts show the disparities in NNS and NS choice of Collocations, Frames and Polywords.

Figure IV.3 (data from Appendix IV.2.30) NNS and NS Choice of Collocations

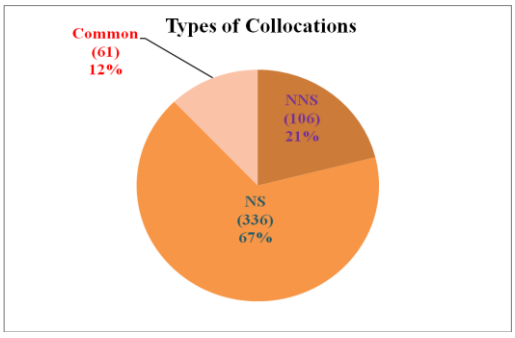


Figure IV.4 (data from Appendix IV.3.15) NNS and NS Choice Frames

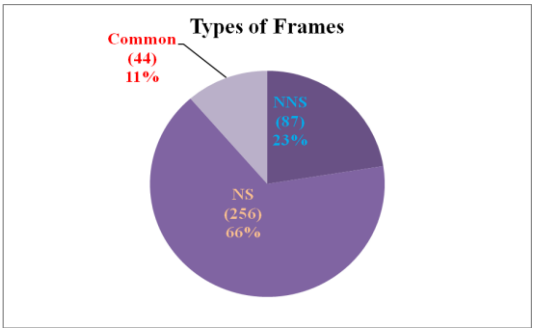
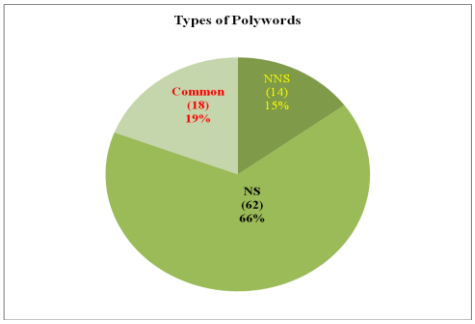


Figure IV.5 (data from Appendix IV.4.7) NNS and NS Choice of Polywords



As shown above, there are 503 (= 61 + 106 + 336), 387 (= 44 + 87 + 256) and 94 (= 18 + 14 +62) distinctive forms[134] of Collocations, Frames and Polywords identified in the collected data. NNSs used less than half of NSs', and only 11% -19% of the FSs were used by both groups, indicating a huge disparity in their choices. It should be noted that this disparity is only indicative, because even two groups of NSs might have considerable disparity as well.

Nonetheless, it seems reasonable to speculate that the disparity between NNSs' and NSs' choice of non-TSSS FSs is mainly due to the different expressions/TSSSs they employed to realize similar speech acts. The details can be found in TSSS analysis but the totally different non-TSSS FSs extracted from a typical NNS and a typical NS utterance below with similar function to realize similar speech act can illustrate the difference in choices, as well as difference in number of non-TSSS FSs:

From the NNS *可不可以改天再考试*, we extracted 3 FSs:

1) *可不可以 VP* (Sentence Head Frame)
2) *改天再* (Noun + Adv Collocation)
3) *考试* (Verb + Obj Collocation)

From the NNS *能不能再另外安排一个时间让我参加考试*, we extracted 6 totally different FSs:

1) *能不能 VP* (Sentence Head Frame)
2) *再另外* (Redundant words Collocation)
3) *安排时间* (Verb + Obj Collocation)
4) *个 + 时间* (Measure word + Noun Collocation)
5) *让我* (Causative verb + Pronoun Collocation)

[134] For example, 考试 and 参加考试 are two distinctive forms of Verb + Obj Collocations. Frequency of each distinctive form is not dealt with in the section.

6) *参加考试* (Verb + Obj Collocation)

IV.2.1.2 Collocations identified

Collocations identified in the collected data include the following 28 formal subtypes (see Appendix IV.2.1-28 for detailed lists and NNS vs. NS quantitative analysis of each):

1	Adverb + (adjective/verb) + adverb	*一直都* (always); *不太* (not very); *不是很* (not very); *忽然就* (suddenly); *很早就* (very early on)
2	Adverb + (adverb) + verb/adjective	*一定不会* (definitely won't); *从来没有* (never happened); *并不是* (actually not); *努力学习* (study hard); *完全忘了* (totally forgot)
3	Causative verb + pronoun	*给我* (allow me to); *求您* (ask you to); *请求您* (beg you to); *麻烦您*(beg you to)
4	Conjunction + adverb + (verb)	*那就* (then); *所以就* (therefore); *然后就是* (and then be)
5	Fillers	*真的; 就是; 就说; 也是* (you know/ well …)
6	Juxtaposed nouns	*白天和晚上* (day and night); *学习和工作* (study and work); *妈妈和爸爸* (mom and dad)
7	Measure word + noun	*个人* (certain number of people); *件事* (certain number of matter); *次机会* (certain number of chance)
8	Modifier + noun as head word	*充分的准备* (sufficient preparation); *这一门学科* (this course); *最后一次机会* (the last chance)
9	Noun + adverb	*一早就* (long time ago); *以后才* (as late as after); *改天再* (sometime later); *结果就* (as a result)
10	Noun + verb	*一早起来* (after getting up early)
11	Place + direction	*心上* (in one's mind); *家里* (at home); *门外* (out of the door)
12	Preposition + noun	*在一块儿* (to be together); *因故* (because of some reasons)
13	Redundant words	*我自己* (I myself); *再另外* (again); *统统都* (all)
14	Repeated words	*好多好多* (many many); *非常非常* (very very); *清清楚楚* (very clear)
15	Subject + predicate	*头疼* (head ache); *病严重* (illness serious); *情况特殊* (case special)
16	Verb + Adv	*不知道怎么就* (don't know why but); *忙起来就* (so busy that)
17	Verb + complement of degree	*忘得一干二净* (completely forgot); *起得晚* (got up late)
18	Verb + complement of direction	*下来* (come down); *起来* (get up); *过去* (go over); *带来* (bring); *醒来* (wake up)
19	Verb + complement of movement	*安排一下* (make some arrangement); *考一次* (take the exam once); *想一想* (think it over)
20	Verb + complement of potential	*毕不了业* (cannot manage to graduate); *记得住* (be able to remember); *说不出来* (cannot tell)
21	Verb + complement of potential (fixed)	*对不起* (sorry); *来不及* (cannot meet the timeline)
22	Verb + complement of result	*考上* (manage to pass an entrance exam); *找到* (manage to find); *记错* (remember mistakenly)

23	Verb + complement of time	*想很长时间* (ponder for a long time); *准备很长时间* (prepare for a long time)
24	Verb + number (+ measure word)	*有一个* (there is a); *有一些* (there are several); *有好多好多* (there are many many)
25	Verb + object (2 syllables)	*开车* (to drive); *生气* (get mad); *撒谎* (tell a lie)
26	Verb + object (3 syllables)	*下功夫* (put effort); *犯错误* (make mistake); *选修课* (take a course); *处理事* (take care of a business)
27	Verb + object (4 syllables)	*把握机会* (seize the opportunity); *有高血压* (suffer from high blood pressure); *对不起您* (feel sorry to you)
28	Verb + verb	*赶回来* (rush back); *等着我去做* (waiting for me to do); *可以接受* (can accept)

IV.2.1.2.1 Quality and quantity of Collocations

A comparison of NNS and NS Collocations is shown in Figure IV.6 below (note the differences in the scales as 28 subtypes are divided into 4 separate charts).

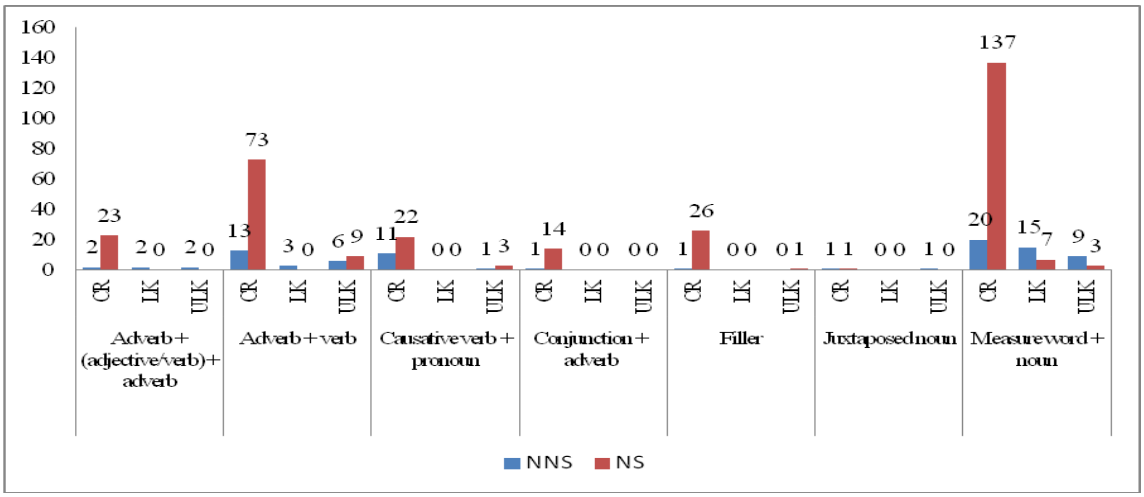
Two observations can be made with regard to the quantity and quality of Collocations used by NNSs and NSs:

a. NNSs' underuse and overuse of FSs with reference to NS production

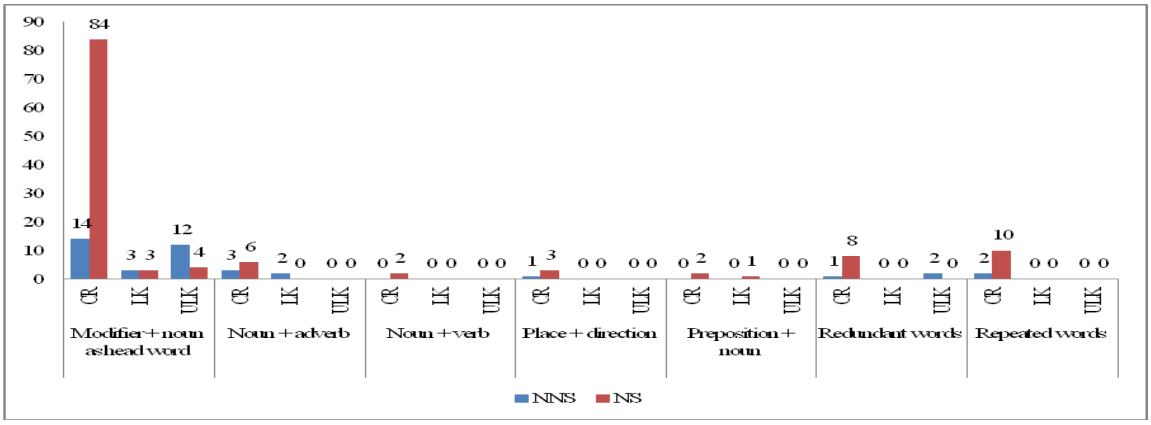
While NSs have all 28 formal subtypes, NNSs do not have 'Noun + Verb', 'Preposition + Noun' and 'Verb + Adverb' collocations, but NSs only produced 2 to 3 FSs in these subtypes and it seems difficult to draw any conclusions based on these limited data. On the other hand, while NNSs produced far less as a whole, they proportionally produced more in 'Juxtaposed nouns' (2:1), 'Verb + Complement of potential' (2:2) and 'Verb + Object (2 syllables)' (49:41). However, while the first two subtypes had too limited data for meaningful analysis, the last one should not be interpreted as 'NNSs outperformed NSs', because the ratio is 51:93 and 26:81 in 'Verb + Object (3 syllables)' and 'Verb + Object (4 syllables)' subtypes, indicating that NNSs produced disproportionally more, or overused, shorter 'Verb + Object' FSs and underused longer ones. For example, NNSs were found to rely on *考试* but had too few *参加考试* when expressing the notion 'to take a test' (see Appendix IV.2.25-27).

Figure IV.6 (data from Appendix IV.2.29) 28 formal subtypes of Collocations

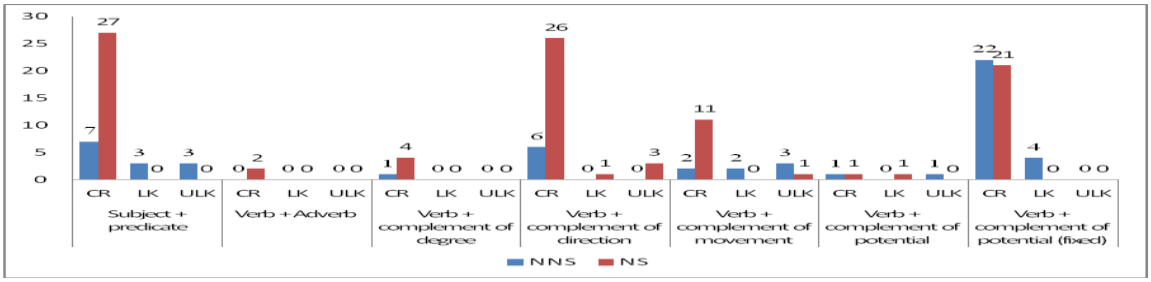
a. Subtype 1-7



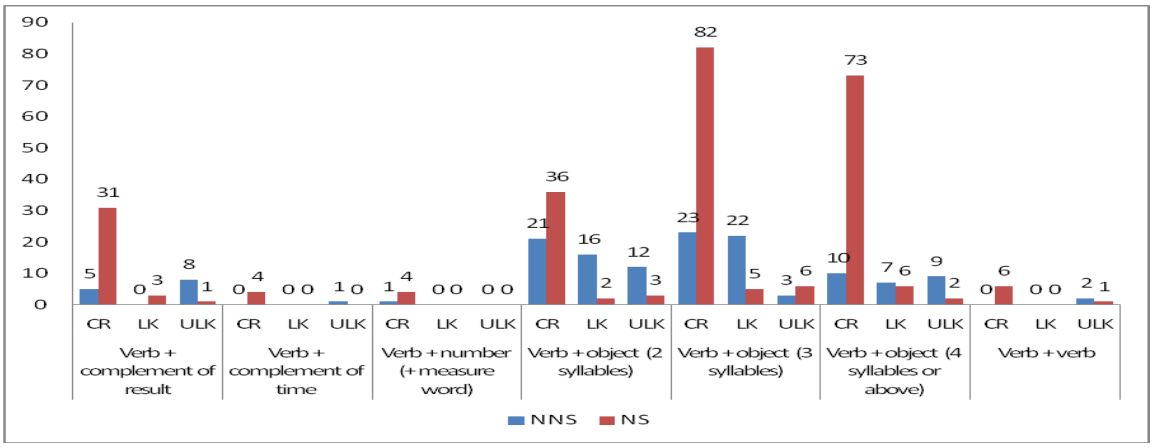
b. Subtype 8-14



c. Subtype 15-21



d. Subtype 22-28



Overuse can be found when we had a closer look at the detailed lists in appendixes. In 'Adverb + (Adjective/Verb) + Adverb', for instance, while NSs' 23 FSs are of 16 different types and distributed more evenly, each with 1 to 3 tokens, 4 out of the 6 NNSs' FSs are *不太* (see Appendix IV.2.1). Other examples of overuse include *让我* (Causative verb + Pronoun), *很大的麻烦* (Modifier + Noun), *对不起* (Verb + Complement of potential), *约好* (Verb + Complement of result), *帮忙*, *有考试* and *没有借口* (Verb + Object).

On the other hand, NNS underuse can also be found in many subtypes, especially those with very low 'NNS versus NS' ratios, such as 'Conjunction + Adv' (1 versus 14) and 'Filler' (1 versus 25). Examples include 然后就(Conjunction + Adverb), *就是(说)* (Filler), *一定会* (Adverb + Verb), *(请)求您* (Causative verb + Pronoun), *门+课* (Measure word + Noun), *博士您* (Redundant words), *缺考*, *有原因* *and* *安排补考* (Verb + Object).

As a whole, NNS production can be characterized as overuse and underuse of non-TSSS FSs because plausible cases can be found at least in subtype 1, 3, 4, 5, 9, 13, 14, 19, 21, 22 and 27 (see respective Appendixes for details).

b. Items seldom highlighted or covered in traditional textbooks

Among the FSs identified, some are covered in most, if not all, CSL textbooks as products of grammatical calculation. For example, *记得住* together with *记不住* and *记得住记不住* can be found under grammar point 'Complement of Potential'. Some appear in textbooks as more fixed patterns or expressions, such as *并不是...(而是)* and *家里*. However, some are frequently used but seldom highlighted as a unit, e.g. *很早就，改天再，统统都*. Some might seem grammatical irregular and have to give way to more grammatical and unauthentic ones, e.g. *不是很*[135]. Some others, e.g. *就是（说）* (as hesitant fillers) and *好多好多*, are seldom covered because they are very colloquial and not chosen (or probably not even noticed!) by textbook authors, as most dialogues in textbooks are not based on authentic spoken data.

The effect of being highlighted in the textbooks, plus sufficient practice of course, can be far-reaching. One interesting example is *改天再* mentioned above. It was used 4 times by 3 NNSs but not by NSs at all[136]. Although, in the textbooks, *改天再* was not introduced as a unit, it was embedded in useful phrases like *改天再聊吧* (Li 1995:340). The frequent coexistence of its two constituents was reinforced by classroom practice with the sentence frame *NP 改天再 VP*[137]. Although *改天再* was only used by NNSs and not preferred by NSs[138] and can thus be viewed as an overuse case, it can nonetheless illustrate the importance of awareness-raising of FSs in teaching materials and classroom instructions. It might be safe to say that

---

[135] For example, in Kungfu I (2002), *不很多* was used as a reply to the question *我们班的同学多不多* in a dialogue. The more common and colloquial *不是很多* was not introduced.

[136] Syntactically and semantically the 4 *改天再* by NNSs were almost used perfectly: 1) *我的要求就是改天再考*; 2) *可不可以改天再考*; 3) *可不可以改天再安排时间让我补考呢*; 4) *能不能改天再考试吗*? These utterances are all speech acts used to request for another appointment. NSs typically preferred *能不能再另外安排一个时间让我参加考试，您可以给我特别再安排一次补考吗* or *能不能再给我安排一次补考*, etc. (see IV.3.1.3.3 below).

[137] As a matter of fact, *改天* is never followed by words other than *再* in the textbooks NNSs used, nor in the data collected in this study.

[138] The NNS utterances containing *改天再*, though accepted by the judges in this study, are pragmatically not very appropriate. As revealed in CCL example, it is almost always used to mean something has to be postponed to give way to something urgent or due to lack of necessary conditions, e.g. *我还要赶下午的飞机...，吃饭的事就只好改天再说了；...不在，改天再交吧*. This might explain why it was not employed by any NSs.

at least to the 3 NNS participants in this study, *改天再* has been inputted and can be outputted as a unit, or a FS.

IV.2.1.3 Frames Identified

Frames identified in the collected data include the following 13 types (see Appendix IV.3.1-13 for detailed lists and NNS vs. NS quantitative analysis of each):

1	Adverb + verb frame	*一个人在 NP* (alone in NP); *一直在 NP* (all the time in NP)
2	Adverbial frame	*NP 跟 NP 一起* (NP (does something) together with NP); *NP 跟 NP 都* (both NP and NP); *每 NP 都* (every NP); *整个 NP 都* (the whole NP)
3	Conjunctive frame	*不光 VP, 而且 VP* (not only VP, but also VP); *由于 VP, 所以 VP* (because VP, therefore VP); *如果 VP 的话, 就 VP* (if VP, then VP); *我知道 VP, 但是 VP* (I know that VP, but VP)
4	Noun frame	*NP 和 NP 的决赛* (final match between NP and NP); *VP 的时间* (time for VP); *VP 的经验* (experience of VP); *VP 的错误* (mistakes in VP)
5	Particle + particle frame	*VP 啊 VP 啊 VP* （VP and VP and VP）
6	Position and direction frame	*从 NP 中* (from the centre of NP); *在 NP 上* (on top of NP); *在 NP 面前* (in front of NP)
7	Preposition + verb frame	*NP 对 NP 重要* (NP is important to NP); *NP 向 NP 道歉* (NP makes apologies to NP); *NP 比 NP 少* (NP is less than NP); *NP 把 NP 给错过了* (NP missed NP); *NP 跟 NP 一样* (NP is the same as NP); *NP 给 NP 打电话* (NP makes a call to NP); NP 给 NP 添麻烦 (NP brings troubles to NP); NP 为 NP 做准备 (NP prepares for NP)
8	Sentence Core frame	把 NP 再 VP 一下 (VP a little bit of NP); 已经 VP 了 (have already VP); 不 VP 了 (no longer VP); 不会再 VP 了 (will not VP any more); 不是故意 VP 的 (did not VP deliberately); 再 VP 一个 NP (VP NP one more time); 是 VP 的好机会 (is a good opportunity to VP)
9	Sentence Crown frame[139]	对 NP 来讲 (As far as NP is concerned); 不知道为什么 (Don't know why but); 如果 VP 的话 (If VP), 如果可以的话 (If possible), 你知道 (You know that); 我想 (I think)
10	Sentence Head frame	不知道可以不可以 VP (I don't know if it is

[139] Sentence Crown frame are named 'gambits' in Keller (1981). They are markers/signals facilitating conversational discourses (Keller 1981:94; Aijmer 1996:2) and typically come before Sentence Head frames and other sentence constituents.

		possible to VP); 你可以不可以 VP (Can you VP); 希望你可以 VP (I hope you can VP); 我也不知道为什么 VP (Neither do I know why VP); 我一定 VP (I definitely will VP); 我只是想 VP (I just want to VP); 请你 VP (Please VP); 是不是能 VP (Is it possible that VP)
11	Sentence Tag frame	VP, 可以吗 (Is it okay to VP); VP, 行不行 (Is it okay to VP)
12	Verb + object frame	NP 帮 NP 忙 (NP gives NP a hand); 安排 VP 的机会(to arrange an opportunity to VP); 是 NP 的问题 (it is a problem of NP); 养成 Adj 的习惯 (to cultivate a Adj habit)
13	Verb + Verb frame	*听见 NP 响* (heard the sound of NP); *NP 到 NP 去* (NP goes to NP); *NP 送 NP 回 NP* (NP escorts NP back to NP); *NP 带 NP 去 NP* (NP takes NP to NP); *有 NP 需要 VP* (there is NP that need to be VP); *记得要 VP* (I remember that I need to)

IV.2.1.3.1 Quantity and Quality of Frames

A comparison of NNS and NS Frames is shown in Figure IV.7 below (note the differences in the scales). Altogether there are 13 formal subtypes of Frames identified in the data.

Two observations can be made with regard to the quantity and quality of Frames used by NNSs and NSs:

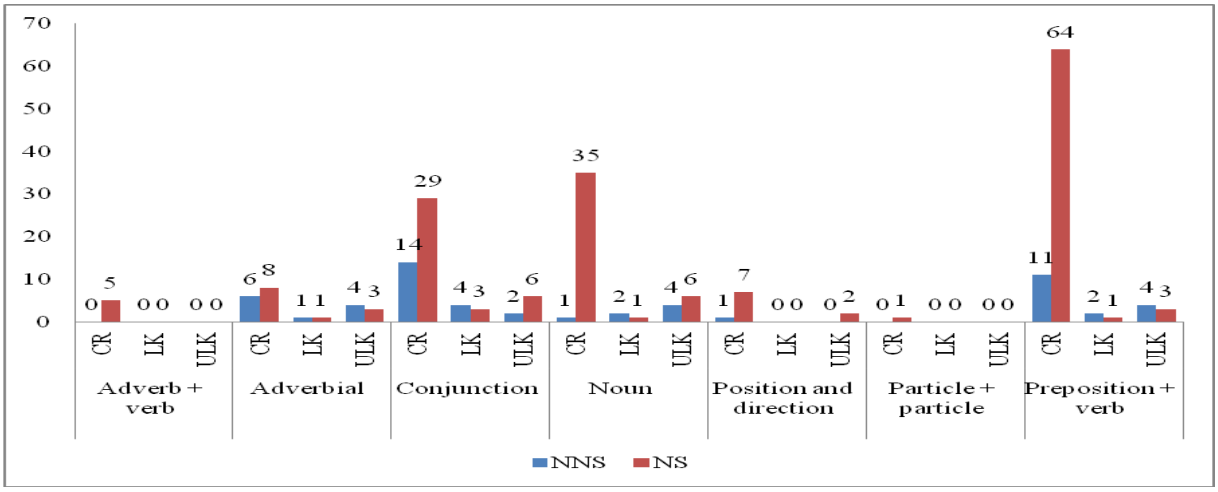
a. NNSs' underuse and overuse of FSs with reference to NS production

While NSs do not have 'Sentence Tags' (3:0), NNSs do not have 'Adv + Verb' (0:5) and 'Particle + Particle' (0:1). However，it seems difficult to draw any conclusions based on these limited data.

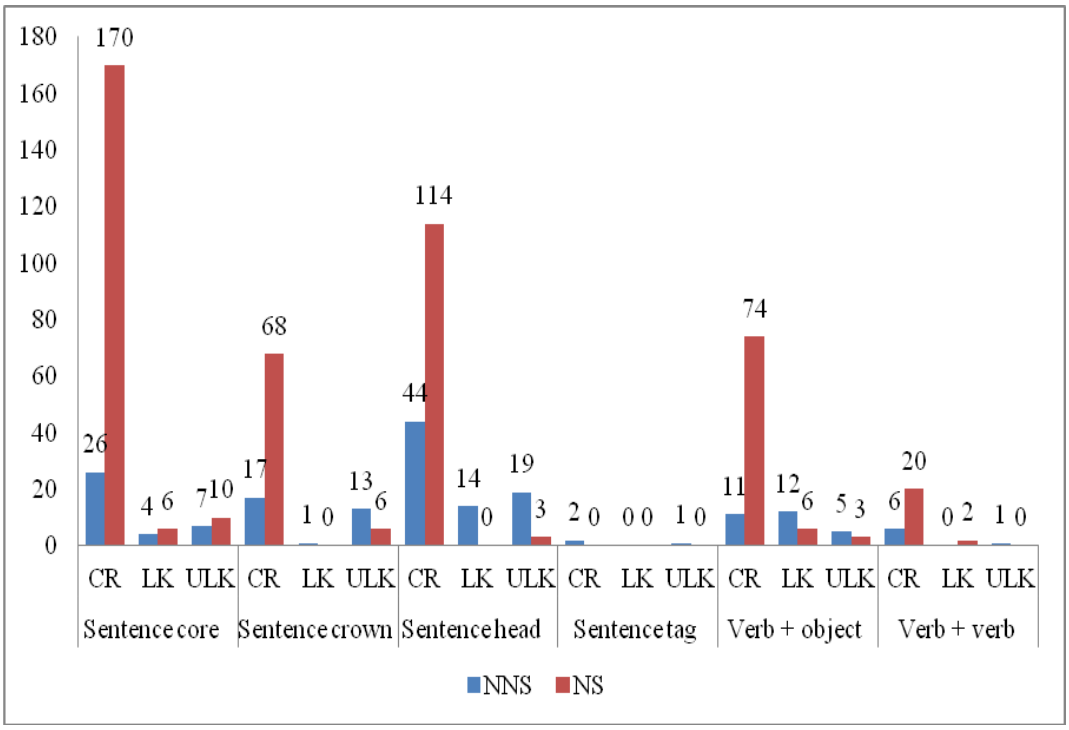
While NNSs produced far less as a whole, they proportionally produced quite a lot in 'Adverbial' (11:12), but seemed to overuse *NP 跟NP 一起, 每NP 都*and *什么NP 都*. NNS overuse can also be observed in 'Conjunction' (e.g. *因为VP 所以 VP*), 'Preposition + Verb' (e.g. *NP 跟NP 说话*), 'Sentence Crown' (e.g. *VP 的时候*) 'Sentence Head' (e.g. *可不可以 VP, 请您 VP, 我想 VP, 我要 VP*). (see Appendix IV.3.2-10 for details)

Figure IV.7 (data from Appendix IV.3.14) 13 formal subtypes of Frames

a. Subtype 1-7



b. Subtype 8-13



On the other hand, NNSs had obvious underuse in 'Noun' (e.g. *VP 的机会* and *VP 的事情*), 'Position and Direction' (e.g. *在NP 上*), 'Preposition + Verb' (e.g. *NP 对NP 重要*, *NP 给NP 安排* and all those with *把*) and 'Sentence Core' (e.g. *把NP (给) VP 了*, *给VP 了*, *会VP 的* etc), 'Sentence Head' (e.g. *希望您能 VP*), 'Sentence Crown' (*无论如何*). (see Appendix IV.3.4-10 for details)

As a whole, again, NNS production can be characterized as overuse and underuse of non-TSSS FSs in most subtypes of Frames.

b. Items seldom highlighted or covered in traditional textbooks

Just like Collocations, some Frames, e.g. *从NP到NP* and *NP跟NP一样*, are overtly and constantly covered in CSL textbooks as patterns. However, some are frequently used but might not be highlighted as a unit, e.g.*不会VP的, 刚刚VP完, 有NP需要VP，整个NP都, 一直到NP才* etc. Some others, e.g. *比较Adj一点儿, VP这样的问题, NP等等的东西, 把NP给VP掉了, 看您能不能VP, 我不知道您是不是VP, NP听见NP响* and *有NP等着NP去VP* etc are seldom included in textbooks because they are very colloquial and difficult to be found or noticed without the aid of corpus analysis based on spoken data.

IV.2.1.4 Polywords Identified

Polywords identified in the collected data include the following 5 types (see Appendix IV.4.1-5 for detailed lists and NNS vs. NS quantitative analysis of each):

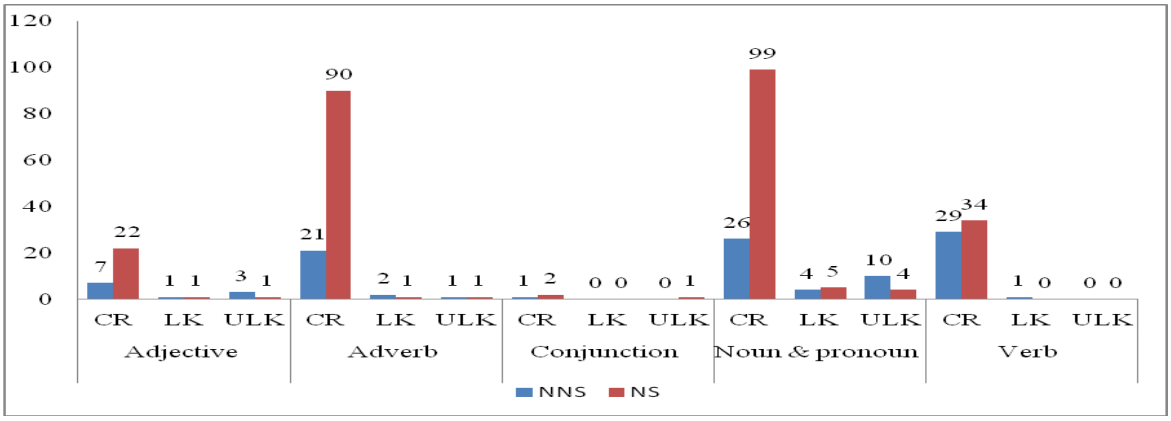
1	Adjectives	*不好意思* (embarrassed); *有礼貌* (polite); *有意义* (meaningful) ; *糊里糊涂* (muddled); *不(太)舒服* (feel ill); *绝佳* (extremely good); *一干二净* (extremely clean); *无意识* (without consciousness)
2	Adverbs	*一个人* (alone); *主要是* (mainly); *实际上* (actually); *非常的* (very); *诚心诚意* (sincerely) ; *实在是* (indeed) ; *真的是* (really) ; *真是* (really) ; *真的* (really) ; *故意的* (on purpose)
3	Conjunctions	*或者是* (or); *所以说* (therefore); *还有* (besides)
4	Nouns & pronouns[140]	*电话号码* (phone number); *另外一个* (another); *这几天* (these few days); *最后一个* (last one); *今天早上* (this morning); *什么时候* (what time); *这(一)次* (this time); *一大半*(most of); *一开始* (at the very beginning)
5	Verbs	*死了* (died); *加班加点* (work overtime); *忘记了* (forgot); *忘了* (forgot) ; *实话实说* (tell the truth)

[140] Nouns and pronouns are grouped together because of some borderline cases such as *何时* (= when) which is categorized as a noun in The Standard Dictionary of Contemporary Chinese (2004) but a pronoun in Yingyong Hanyu Cidian (2000).

IV.2.1.4.1 Quantity and Quality of Polywords

A comparison of NNS and NS Polywords is shown in the chart below. There are 5 subtypes of Polywords in which NNSs produced less in every one of them. However, NNSs used almost the same number of Polyword verbs[141]. This is due to the fact that NNSs used same amount of *忘记了* (forgot) and *忘了* (forgot) with NSs when telling the professor that they forgot about the test, which seems to be a necessary speech act in the case (28 tokens in each group; See Appendix IV.4.5). The other observation is that NNSs Polyword verbs were almost used correctly, seemingly because they were embedded in structurally simple utterances like *我忘了(今天的)考试.*

Figure IV.8 (data from Appendix IV.4.6) Overall distribution of Polywords



Two more observations can be made with regard to the quantity and quality of Polywords used by NNSs and NSs:

a. NNSs' underuse and overuse of FSs with reference to NS production

NNS overuse can be observed in 'Adjective' (e.g. *糊里糊涂*), 'Adverb' (e.g. *真的*) and 'Noun & Pronoun' (e.g. *别的时候*), but underuse is far more

[141] Most Polyword verbs in the collected data have unitary internal structure: combinations of a verb and particle *了*. The verbs that can fit into this category are seldom used along, especially in spoken language. For example, *忘* is almost always followed by *了* or *掉* in CCL corpus. *Don't forget us* should be translated as *别忘了我们* or *别把我们忘了*. This is why these word combinations were treated as Polyword verbs in this study.

serious in Polywords, especially in ‘Adverb’ and ‘Noun’.

Among the 24 Polyword ‘Adverbs’, NSs produced all but NNSs only produced 4. Half of the 24 Polyword ‘Adverbs’ identified have a *是* in the end, which does not carry substantial meaning (e.g. *的确也是*, *反正是*, *其实是*, *其实也是*, *已经是*, *应该是*, *真的是*, *主要是* etc), NNSs used none of them (Appendix IV.4.2).

Let’s take a closer look at the use of 3 synonyms meaning ‘really’ in Polyword ‘Adverbs’. *真的* was used 18 and 26 times by NNSs and NSs respectively, while the other two with 是, i.e. *真的是* and *真是*, were used 18 and 2 times but solely by NSs. Both *真的* and *真是* were covered in the course materials NNSs used, but 真的 was introduced earlier and with higher frequency, and NNSs only relied on it. On the other hand, 真的是 was never introduced and it is not surprising that it was not used at all, even though it was very frequent in NS data.

There is also a pair of synonyms in Polyword ‘Noun and Pronoun’: *这次* and *这一次*. They both mean ‘this time’ but can come in front of a noun to form noun phrases like *这次考试* or *这一次考试*. *这次* and *这一次* were used by NSs 21 and 7 times, while the latter was used once by a NNS (Appendix IV.4.4). A deeper search of the raw data found that NNSs relied almost exclusively on *这个*, which is not a FS, to form noun phrases like *这个考试*. The reason behind NNSs’ preference seems to be the same as *真的*: *这个* was introduced first and used far more frequently in class[142], resulting in NNSs’ heavy reliance on it[143].

As a whole, again, NNS production can be characterized as overuse and underuse of non-TSSS FSs in most subtypes of Polywords.

---

[142] *这个* can precede far more nouns than *这(一)次*.

[143] It can also be speculated that it is because *这次考试* or *这一次考试* (= this time test) are semantically incompatible with NNSs’ first languages.

b. Items seldom highlighted or covered in traditional textbooks

Among the Polyword FSs identified, many, e.g. the Polyword 'Adverbs' with *是* mentioned above, Polyword 'Conjunctions', i.e. *或者是, 所以说, 甚至是*, and Polyword 'Nouns and Pronouns' like *一大半* and *一开始* are seldom covered in traditional textbooks. Again, it can be speculated that it is because they are colloquial and not noticed by textbook authors, who tend to make up perfect dialogues when compiling textbooks.

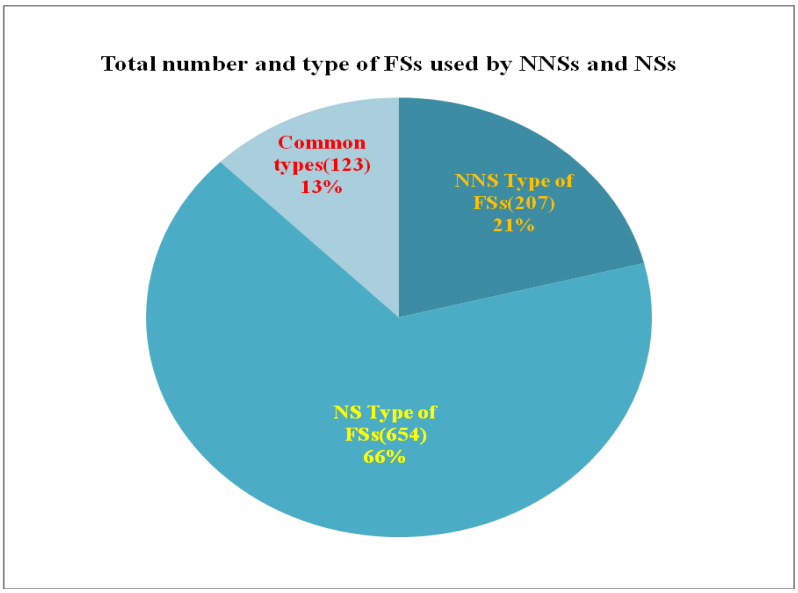
IV.2.1.5 Interim Summary

As shown in Table IV.3 and Figure IV.9, altogether 2393 non-TSSSs FSs have been identified in this study. They are of 984 type/choices under 3 formal categories and 46 formal subcategories. Only a minority of them were used by both NNS and NS participants.

Table IV.3 Total number and type of FSs used by NNSs and NSs

	NNS no. of FSs	**NS no. of FSs**	**Total no. of FSs**	**NNS Type of FSs**	**NS Type of FSs**	**Type of FSs used by both**	**Total Types of FSs**
COLLOCATIONS	326	805	1131	167	397	61	503
FRAMES	239	657	895	131	300	44	387
POLYWORDS	106	261	367	32	80	18	94
Total	671	1723	2393	330	777	123	984

Figure IV.9 Total number and type of FSs used by NNSs and NSs



IV.2.1.5.1 Quantification of Disparities between NNSs and NSs

In order to draw an easy-to-understand fuller picture, results are further quantified as below, as though we are marking students' performance at school.

If all a CR (correct) receives one point, a LK (likely FS) receives half point and a ULK (unlikely FS) receives no point, then we can get Table IV.4 (results converted to a hundred-mark system).

As shown in Table IV.4, while NS scored 93.3 (far from perfect but understandable, as NS participants were not reading a well-versed essay but speaking almost spontaneously), NNS scored a much lower 68.1. Of course the calculations of the marks are internally consistent but externally arbitrary (say. if LKs are given more points and if ULK can also receive some, the disparities will be narrowed, and vice versa). However, this is no doubt a good indication of the differences between the interlanguage of NNSs who had studied Putonghua for only 2 years and the fully developed native language of well-educated NSs.

Table IV.4 Quantification of NNS and NS FSs

	NNS FSs				NS FSs			
	NNS Total	CR (correct)	LK (likely)	ULK (unlikely)	NS Total	CR (correct)	LK (likely)	ULK (unlikely)
COLLOCATIONS	326	169	79	78	805	739	29	37
Points	**25.9**	**20.99**	**4.91**	**0**	**93.6**	**91.8***	**1.8****	**0*****
Adjusted points#	**64.1**	**52**	**12.1**	**0**				
FRAMES	239	139	40	60	657	595	20	42
Points	**24.24**	**21.19**	**3.05**	**0**	**92.07**	**90.55**	**1.52**	**0**
Adjusted points#	**67.3**	**58.9**	**8.4**	**0**				
POLYWORDS	106	84	8	14	261	247	7	7
Points	**33.71**	**32.18**	**1.53**	**0**	**95.98**	**94.64**	**1.34**	**0**
Adjusted points#	**83**	**79.3**	**3.7**	**0**				
Total FSs	671	392	127	152	1723	1581	56	86
Total Points	**26.57**	**22.85**	**3.72**	**0**	**93.3**	**91.7**	**1.6**	**0**
Adjusted total points#	**68.1**	**59.6**	**8.5**	**0**				

* 100 ÷ 805 x 739 = 91.8 (All correct FSs received full marks)

** 100 ÷ 805 x 29 x 0.5 = 1.8 (All likely FSs received half marks)

*** 100 ÷ 805 x 37 x 0 = 0 (All unlikely FSs received no marks)

All NNS points are adjusted in accordance with the differences in the number of each category.

As mentioned in Section III.2.2, 18 of the 30 NNS participants (60%) only obtained Intermediate grades while the task being studied is at Advanced level,

and even for the other 12 who received Advanced grades, according to ACTFL (American Council on the Teaching of Foreign Languages) Proficiency Guidelines, their proficiency level is only equivalent to a 2 in DLPT (Defense Language Proficiency Test)[144] scale in which 0 stands for novice and 5 stands for nativelike proficiency (Intermediate grades are equivalent to 1 or 1+ in DLPT scale). The huge disparity in Table IV.4 seems to be a reasonable reflection of their disparity in oral proficiency.

On a related issue, although the COPA grades NNS participants got are overall ones based on all the 15 tasks (see III.1.3 for details) they attempted, and their oral productions were graded based on whether the 15 language tasks were successfully completed rather than based on the use of FSs, it seems reasonable to argue that the quality of non-TSSS FSs is a good reflection of their language proficiency (cf. IV.2.2 below).

IV.2.1.5.2 Overall Disparities between NNSs and NSs

As revealed by the NNS vs. NS quantitative analysis above (including Appendix IV.2.1-28, IV.3.1-13 and IV.4.1-5), compared with NS ones, major characteristics of NNS non-TSSS FSs include:

1. Significantly smaller quantity and lower density (671 vs. 1724);

2. Fewer varieties and different choices, indicating a plausible underuse and overuse[145] (cf. Milton 1998:189)

    - Fewer choices in general (330 vs. 777) and in most of the categories;

[144] DLPT is a battery of foreign language tests produced by the US Defense Language Institute and used by the US Department of Defense.

[145] As a substantial part of NNS and NS production (30.89% and 18.88%) is about the very diverse details of how they missed the test, the small percentage of FSs shared by both group might not be completely attributable to NNSs' underuse and overuse of certain FSs. However, at least the much higher proportion of ULK (unlikely) FSs is a good indicator of misuse, which, from a different angle, might indicates underuse of the FSs that should have been used.

- Mainly relying on a small number of FSs;

- Limited number in common with NS data

3. Far lower quality, i.e. with significantly higher percentage of errors (LKs + ULKs) in general (41.58% vs. 8.24%) and in most subcategories.

Combined with the result of inspecting the course materials used by the NNSs, we may tentatively conclude that this inferiority might be partly due to the following:

1. a lack of input and awareness-raising in the course materials and pedagogical interventions;

2. a grammar-centred rather than lexical and task-based approach adopted in the language program;

3. an under-representation of colloquial data in textbook compilation;

4. avoidance strategies adopted by the learners, i.e. relying excessively on the safe and familiar items;

Although the 4th point seems to be the learners' responsibility, it can logically be, to a certain degree, rectified by changes introduced in the first 3 aspects.

IV.2.2 Testing of Hypotheses 1-3

This section presents quantitative findings to test the first three hypotheses concerning density of non-TSSS FSs in NNS and NS data, and compares them with related quantitative findings in other studies when applicable.

As specified in the first three hypotheses in the beginning of Chapter III, we are interested in knowing if degree of formulaicity (in terms of characters inside FSs as percentage of total characters) correlates with oral proficiency or quality of oral

production, i.e. if NS data is more formulaic than NNS data, if advanced NNS data is more formulaic than less advanced NNS data, and, in NNS data, if native-like utterances are more formulaic than non-native-like utterances.

IV.2.2.1 Quantitative Results Pertaining to Hypothesis 1

As shown in Table IV.5 below, in NNS and NS data, on the average 60.4% and 71.9% of syllables/characters are inside FSs, and the difference is significant ($p < 0.001$). As a result, Hypothesis 1 (NS data is more formulaic than NNS data, i.e. there are more characters inside FSs as percentage of total characters in NS data) can be proved.

Compared with the studies on English FSs reviewed in Section II.3, the above percentages are quite high. This might firstly be due to the fact that spoken data contain higher density of FSs[146], secondly due to the nature of the task at issue in which a very high concentration of Interactional FSs is needed to apologize, excuse, request and win favorable impressions, and thirdly because FSs are exhaustively identified in this study (see Section III.4.1)[147].

Table IV.5: t-test result comparing percentage of syllables/characters inside FSs in NNS and NS data

Gp 1	Gp 2	Avg1	Avg2	Std 1	Std 2	n1	n2	CL (%)	Z-value	T-Stat	Implication
percentage of characters inside FSs in NNS	percentage of characters inside FSs in NS	60.4%	71.9%	0.12	0.07	30	30	99.9%	3.2368	4.4313	NS > NNS

Two related t-tests are presented below to prove the 1st Hypothesis from slightly

[146] Biber et al (1999), Brazil (1995), Erman and Warren (2000) and Leech (2000) also find higher density of FSs in spoken data.

[147] For instance, in Moon (1998) with far lower percentage of words inside FSs, many FSs included in this study are excluded (p2-3, 50), and there is a lack of spoken data and the FSs 'functioning as greetings, valedictions, and other speech acts had distorted frequencies' (p48-49).

different angles.

Firstly exact numbers of non-TSSS FSs used by each participant were compared. Table IV.6 tabulates the result of a t-test demonstrating that on the average a NNS produced 22.37 non-TSSS FSs within the time limit, significantly less than 57.43 by NS ($p < 0.001$).

Table IV.6: t-test result (comparing average number of non-TSSS FSs in NNS and NS data)

Gp 1	Gp 2	Avg1	Avg2	Std 1	Std 2	n1	n2	CL (%)	Z-value	T-Stat	Implication
number of FSs in NNS	number of FSs in NS	22.37	57.43	11.34	11.28	30	30	99.9%	3.2368	11.8070	NS > NNS

Secondly, number of non-TSSS FSs per character in NNS and NS data was also compared to confirm Hypothesis 1 from another perspective (Table IV.7). For every character in NNS data, there is 0.21 FSs (or 21 FSs per 100 characters), significantly lower than 0.29 (or 29 FSs per 100 characters) in NS data ($p < 0.001$).

Table IV.7: t-test result comparing number of FSs per character

Gp 1	Gp 2	Avg1	Avg2	Std 1	Std 2	n1	n2	CL (%)	Z-value	T-Stat	Implication
Number of FSs per character in NNS	Number of FSs per character in NS	0.21	0.29	0.05	0.04	30	30	99.9%	3.2368	6.5730	NS > NNS

Table IV.8 below provides more details on the non-TSSS FSs identified in this study. While the number of characters produced by NSs is about 2 times of NNSs (3230 : 6038), the total number of non-TSSS FSs is 2.57 times of NNSs' (1723 : 671), indicating a far lower density of FSs of this kind in NNS data, echoing the t-test results presented above. NS data have more FSs of this kind in all three subcategories (i.e. Collocations, Frames and Polywords).

Table IV.8: Density of 3 types of non-TSSS FSs in NNS and NS data

	NNS FSs	NS FSs
COLLOCATIONS	326	805
No. per 100 characters*	10.1	13.3
FRAMES	239	657
No. per 100 characters	7.4	10.9
POLYWORDS	106	261
No. per 100 characters	3.3	4.3
Total	671	1723
No. per 100 characters	20.8	28.6

* 'No. per 100 characters' = 'number of FSs' / 'number of total character produced' x 100

It should be noted that all the non-TSSS FSs identified are included in the table above, irrespective of their quality of use. As specified in III.4.2, FSs that really exist but used incorrectly were also included in this study. In III.4.2.2 (Table III.2), mistakenly used FSs are divided into LK (mistaken but LIKELY to be used in the same place with different patterns or collocation etc.) and ULK (mistaken and UNLIKELY to be used in the same place). Two facts can be told from Table IV.9 below:

1 On the average NNSs produced far fewer non-TSSS FSs but far more mistaken non-TSSS FSs.

2 Both NNSs and NSs have more ULKs than LKs, though might not be with significant difference.

Table IV.9: total number of mistaken non-TSSS FSs in NNS and NS data

	Total no. of non-TSSS FSs	Mistaken but likely non-TSSS FSs (LK)	Mistaken and unlikely non-TSSS FSs (ULK)	Total mistaken non-TSSS FSs

NNS	671 (100%)	127 (18.93%*)	152 (22.65%*)	279 (41.58%*)
NS	1723 (100%)	56 (3.25%*)	86 (4.99%*)	142 (8.24%*)

* calculated by dividing total number of LK & ULK by total number of non-TSSS FSs of all participants.

Table IV.10 shows the average number of mistaken non-TSSS FSs of each participant. Note the slight differences between the percentages in this and the last table.

Table IV.10: average mistaken non-TSSS FSs in NNS and NS data

	Average no. of non-TSSS FSs	Average Mistaken but likely FSs (LK)	Average Mistaken and unlikely FSs (ULK)	Average Total mistaken FSs
NNS	22.37	21.9%	22.5%	44.3%
NS	57.43	2.98%	4.45%	7.45%

Results of t-tests run to see if NNS and NS participants have significant difference in their quality of FSs are presented in Table IV.11. The 1st and 2nd lines show that on the average, NNS participants produced significantly more LK ($p < 0.001$) and ULK FSs ($p < 0.001$). 3rd line shows that it holds true when LK and ULK were combined ($p < 0.001$).

Table IV.11: t-test result (comparing percentage of FSs with defects in NNS and NS data)

Gp 1	**Gp 2**	**Avg1**	**Avg2**	**Std 1**	**Std 2**	**n1**	**n2**	**CL (%)**	**Z-value**	**T-Stat**	**Implication**
LK in NNS	LK in NS	21.9%	3.0%	0.189	0.055	30	30	99.9%	3.2368	5.1569	NNS > NS
ULK in NNS	ULK in NS	22.5%	4.5%	0.146	0.09	30	30	99.9%	3.2368	5.6442	NNS > NS
LK & ULK in NNS	LK & ULK in NS	44.3%	7.5%	0.247	0.14	30	30	99.9%	3.2368	6.9956	NNS > NS

As an interim summary, NS data contains significantly more characters inside FSs,

and contains more non-TSSS FSs given the same number of characters, thereby confirming that NS data is more formulaic in nature. Furthermore, as NNS data contains far more FSs that were not used correctly, if only correctly-used FSs count, the difference between NNS and NS data in terms of degree of formulaicity will be further widened.

IV.2.2.2 Quantitative Results Pertaining to Hypothesis 2

While Hypotheses 1 is about the differences between NNSs and NSs, Hypothesis 2 is about the differences between NNSs with better proficiency and those with lower proficiency, i.e. differences between those who got Advanced certificates (NNS (A)) and whose who got Intermediate certificates (NNS (I)) in COPA test (see Section III.2.2). Four important t-tests run above to prove Hypothesis 1 were rerun to generate the results in Table IV.12 below.

Table IV.12: Four t-tests to compare differences between Advanced NNS and Intermediate NNS

Gp 1	Gp 2	Avg1	Avg2	Std 1	Std 2	n1	n2	CL (%)	Z-value	T-Stat	Implication
number of characters in NNS(A)	number of characters in NNS(I)	137.9	87.5	39.37	39.73	30	30	99.9%	3.2368	5.5222	NNS(A) > NNS(I)
percentage of characters inside FSs in NNS(A)	percentage of characters inside FSs in NNS(I)	61.9%	59.5%	0.10	0.13	12	18	95.0%	1.7011	2.8529	NNS(A) > NNS(I)
number of FSs in NNS(A)	number of FSs in NNS(I)	30.25	17.11	11.35	7.92	12	18	99.9%	3.4082	3.4324	NNS(A) > NNS(I)
number of FSs per character in NNS(A)	number of FSs per character in NNS(I)	0.22	0.20	0.04	0.06	12	18	99.0%	2.4671	2.7890	NNS(A) > NNS(I)

The following points can be drawn from the above results:

1. Advanced NNSs produced significantly more number of characters (137.9 vs. 87.5 in the 1st line; $p < 0.001$). In other words, they had greater quantity of

output and probably higher fluency[148].

2. Percentage of characters inside FSs in NNS (A) is only 2.4% higher than NNS (I) (61.9% vs. 59.5% in the 2nd line). Nonetheless, the difference is statistically significant ($p < 0.05$). Density of FSs seems to be a good indicator of proficiency because it shows the differences between NSs and NNSs (see Section IV.2.2.1), as well as between NNSs (A) and NNSs (I), although the disparity is not as great statistically in the latter case. Hypothesis 2 (Data of more advanced NNSs contains significantly higher density of FSs than less advanced NNS) can thus be proved.

3. As in Section IV.2.2.1, comparison of average number of non-TSSS FSs in the 3rd line and average number of non-TSSS FSs per character in the 4th line can serve to further prove the Hypothesis 2 from different perspectives**.**

Two related and interesting t-test results are presented in the two tables below.

Table IV.13: t-test to compare difference in number of characters in grammatical errors between Advanced NNSs and Intermediate NNSs

Gp 1	Gp 2	Avg1	Avg2	Std 1	Std 2	n1	n2	CL (%)	Z-value	T-Stat	Implication
Percentage of characters in grammatical errors in NNS(A)	Percentage of characters in grammatical errors in NNS(I)	10.9%	11.0%	0.05	0.08	12	18	95.0%	1.7011	1.5761	Not Significant

NNS(A) and NNS(I)'s average number of characters in grammatical errors are almost identical (10.9% vs. 11%)[149], and statistically insignificant.

[148] Fluency is not the focus of this study and was not scientifically measured. Higher fluency was inferred from greater quantity of output in the same given time (see Section III.2.3).

[149] It should be noted that seriousness of errors was not taken into consideration in this study.

Table IV.14: t-test to compare difference in number of characters in NL-T between Advanced NNSs and Intermediate NNSs

Gp 1	Gp 2	Avg1	Avg2	Std 1	Std 2	n1	n2	CL (%)	Z-value	T-Stat	Implication
Percentage of characters in NL-T in NNS(A)	Percentage of characters in NL-T in NNS(I)	50.1%	45.8%	0.25	0.24	12	18	95.0%	1.7011	1.7094	NNS(A) > NNS(I)

NNS(A) have significantly more characters in Native-like and Typical utterances (NL-T) than NNS(I) ($p < 0.05$).

When combined with the previous t-tests, it seems reasonable to argue that what makes NNSs(A) superior than NNSs(I) is not grammatical accuracy, but amount of production (or fluency), density of FSs and number of characters in Native-like and Typical utterances (NL-T).

IV.2.2.3 Quantitative Results Pertaining to Hypothesis 3

As specified in Section III.4.2.3, all utterances were labeled as one of the following 3 types of word strings based on the corrections judges made: Nonnative-like (NNL) utterances, Native-like but Untypical (NL-UT) utterances, and Native-like and Typical (NL-T) utterances. Hypothesis 3 aims at discovering the relationship between degree of formulaicity and native-likeness.

Firstly, we will look at the NS output. Although there are some grammatical errors in NS data, their quantity is small, negligible and not enough to turn an utterance into Nonnative-like utterances (NNL). As a result, there are only Native-like but Untypical (NL-UT) utterances, and Native-like and Typical (NL-T) utterances in NS data. The table below shows that although there is an average difference of 5%, percentages of characters inside FSs in NS NL-T and NL-UT are not significantly different even at 95% Confidence Level. In other words, NL-T and NL-UT in NS data are as formulaic as each other.

Table IV.15: t-tests to compare density of FSs of NL-T and NL-UT in NS data

Gp 1	Gp 2	Avg1	Avg2	Std 1	Std 2	n1	n2	CL (%)	Z-value	T-Stat	Implication
percentage of characters inside FSs in NL-T (NS)	percentage of characters inside FSs in NL-UT (NS)	70.1%	65.1%	0.07	0.13	30	15	95.0%	1.6811	1.3146	Not Significant

When we turn to NNS data, similar result evolved. Percentage of characters inside FSs in NL-T is 2.2% higher than NL-UT, and they are also not significantly different even at 95% Confidence Level. Again, NL-T and NL-UT in NNS data are as formulaic as each other.

Table IV.16: t-tests to compare density of FSs of NL-T and NL-UT in NNS data

Gp 1	Gp 2	Avg1	Avg2	Std 1	Std 2	n1	n2	CL (%)	Z-value	T-Stat	Implication
percentage of characters inside FSs in NL-T (NNS)	percentage of characters inside FSs in NL-UT (NNS)	59.6%	57.4%	0.12	0.21	29	28	95.0%	1.6730	0.4829	Not Significant

Based on the above t-tests, plus analysis in Section IV.2.2.1 above, we may conclude that while NNS and NS output are significantly different in density of FSs or degree of formulaicity, their respective NL-T and NL-UT are not.

When we go further to compare native-like utterances (NL-T and NL-UT) with non-native-like ones (NNL) in NNS data, the results are consistent, as shown in Table IV.17 below. Both NL-T and NL-UT are more formulaic than NNL ($p < 0.01$). The 3rd line shows the result when NL-T and NL-UT are combined, indicating a significant difference between native-like and non-native-like utterances ($p < 0.01$). As a result, Hypothesis 3 **(**in NNS data, native-like utterances are more formulaic than non-native-like utterances) can be proved.

Table IV.17: t-tests to compare density of FSs of NL-T, NL-UT and NNL in NNS data

Gp 1	Gp 2	Avg1	Avg2	Std 1	Std 2	n1	n2	CL (%)	Z-value	T-Stat	Implication
percentage of characters inside FSs in NL-T (NNS)	percentage of characters inside FSs in NNL (NNS)	59.6%	44.6%	0.12	0.22	30	25	99.0%	2.3988	2.9318	NL-T > NNL
percentage of characters inside FSs in NL-UT (NNS)	percentage of characters inside FSs in NNL (NNS)	57.4%	44.6%	0.21	0.22	29	25	99.0%	2.4002	2.9215	NL-T > NNL
percentage of characters inside FSs in NL-T (NNS) & NL-UT (NNS)	percentage of characters inside FSs in NNL (NNS)	60.1%	44.6%	0.10	0.22	30	25	99.0%	2.3988	3.1352	NL > NNL

IV.2.2.4 Interim Summary

1 On the average, NS data is more formulaic than NNS data. This finding is consistent with other researchers' work on other languages, such as Foster (2001), Schmitt (2004) and Forsberg and Fant (2010).

2 Data of more advanced NNSs (as reflected by their oral proficiency attainments) contains significantly higher density of FSs, greater amount of production within the same time limit, and more Native-like and Typical utterances (NL-T) than less advanced NNS. These are consistent with findings of Ding (2006:26), Wang (2007:19), Ma (2010:27) and Yang (2010). On the contrary, insignificant difference in grammatical accuracy between advanced and less advanced NNSs was observed. By investigating the grammatical accuracy, and oral and written test scores of Chinese English majors, Ting and Qi (2005 quoted in Wray & Fitzpatrick 2008) also find the number of FSs to be a better predictor of the quality of output than grammatical accuracy.

3 In NNS data, native-like utterances are more formulaic than non-native-like utterances.

4 Percentage of non-TSSS FSs used mistakenly in NNS data is significantly higher than that in NS data. This further widens the disparity between the two

groups.

Based on the above, the correlation between density of FSs and level of proficiency in the task being investigated can thus be established.

IV.3 Task-specific Sentence Stems (TSSSs)

This section is to tackle the 2nd question in the beginning of Chapter III:

> What are the TSSSs that are employed by NSs and NNSs for this particular language task and how they differ?

Searle (1979:178) asserts that while the purpose of language is communication, the actual unit to realize human communication is the speech act. The main purpose of this section is to understand the utterance-level linguistic means participants used to carry out their speech acts as captured in their oral production[150]. The quantity and quality of NNS and NS utterances are compared, and the task-specific sentence stems (TSSSs) mainly extracted from utterances classified as native-like and typical (NT-T) in Section III.3.3 are highlighted (some of them were frequently used, and some were only used by a few but still highly recommendable to learners based on experienced native-speaking CSL teachers' judgements). The deficiencies of NNS production are occasional addressed, with reference to the CSL program they had completed before trying this language task being researched, to make pedagogical inferences.

The language task being studied in this research is a pragmatically challenging task which is approximately what Goffman (1971:109) called a 'remedial work', except with an indispensable 'request' component to get an additional chance for a makeup exam. In terms of difficulty, it is similar with the PDR-high tasks[151] in

---

[150] Wray (2002) maintains that '(s)ome messages are much more common than others, and so it is a ratio of message to message-expression that will best help us to understand how some expressions of a given message are favored over others' and calls for research in this area (p31). This study might in part address this.

[151] PDR stands for interlocutors' power difference (P), social distance (D) and the degree of imposition (R) (Taguchi 2007:114). For other attempts to characterize task difficulty, see Candlin

Taguchi (2007) and Forsberg and Fant (2010) in which the speakers faced someone with more power and considerable social distance, and 'the request made by less powerful party implies a high degree of imposition' (Forsberg and Fant 2010:56)[152]. Therefore, in spite of being a daily life speech event, huge disparities in the performance of NNSs and NSs can be expected due to the complications involved.

The spoken task was chosen also because higher concentration of FSs can be expected[153], because of the huge number of Interactional FSs contained in the data, and because the many types of Interactional FSs contained range from simple and straightforward apologies and favor-asking to more tactful sympathy-winning and self-reproving etc[154]. Although the speech outputs are paragraph-length, most Interactional FSs can also be used independently in conversations.

IV.3.1 TSSSs identified in NNS and NS data

In this research, we grouped the TSSSs with specific functions under a few broad functions (e.g. 'overt apologies', 'self-reproving' and 'showing regret' are all specific functions under broad 'Apologizing' function) and how they contribute to the completion of the language task[155]. In other words, a broad function contains a few specific functions, and a specific function might contain a few forms[156] of TSSSs (see below for examples).

As mentioned in Section III.3.2 and III.3.3, based on corrections, all utterances

---

(1987) and Skehan (1998).

[152] As mentioned in **III.2.2**, the NNSs and the NSs participants attempted the language task with quite different motives and it is not clear if the NSs felt the same degree of imposition as NNSs.

[153] Aijmer (1996:7) and Ellis (2008:5) summarize that FSs are more frequent in spoken language.

[154] As a matter of fact, many tactful expressions were not thoroughly practiced or even not covered in the two-year language course the NNS participants took. This may contribute to the huge disparities between NS and NNS output.

[155] Whether there are such things as broad functions at all and whether certain specific functions should be under a certain broad function is debatable and might largely depend on research and pedagogical needs.

[156] A form of TSSS means a frame of utterance like *请（求）NP 再给 NP 一个/次机会（吧）* in III.4.3.3 or *NP be-TENSE sorry to keep-TENSE you waiting* in Pawley and Syder 1983 (p210).

were classified as Nonnative-like (NNL), Native-like but Untypical (NL-UT), and Native-like and Typical (NL-T), and utterances were categorized according to their functions. Altogether 629 (252 by NNSs and 377 by NSs) utterances have been identified and grouped into 27 specific function types, before being further grouped under 5 broad function types as shown in Table IV.18. We will find out all the forms of TSSSs in each specific function type below.

A few facts can be told from Table IV.18, and some inferences can be derived from it:

1. While NSs' production covers all 27 specific function types of utterances, NNSs' only covers 21 of them, but in each of the 6 types not covered by NNSs, NSs only produced 1 or 2 utterances, indicating that they are not very common among native speakers and might not be very critical. It seems that NNSs have a very similar coverage as NSs in terms of function types, when tackling the language task at issue.

Table IV.18 Broad and Specific Function Types of utterances

Broad function types	Functions of each type in discourse	Specific Function types	NNS Number of characters	NNS Number of utterances	NS Number of characters	NS Number of utterances
Apologizing	To apologize, ask for forgiveness, show one's regret etc.	apologizing-asking for forgiveness	7	1	102	9
		apologizing-overt apologies	243	35	465	49
		apologizing-self-reproving	113	12	458	21
		apologizing-showing regret	0	0	45	2
Discourse devices	To signal the starting or ending of a discourse and to sustain the smoothness of speech.	discourse device-alerting	0	0	12	1
		discourse device-closing NNS	2	1	7	3
		discourse device-fluency device by questioning	8	1	8	2
Excusing	To explain why the make-up test was missed.	excusing-detail	988	71	1140	85
		excusing-exonerating	91	7	267	15
		excusing-forgot	280	24	260	20
		excusing-gloss over	14	1	212	12
		excusing-testifying	94	6	450	24
Requesting	To beg for another chance.	requesting-another chance	321	22	812	46
		requesting-general	78	6	124	11
		requesting-new appointment	424	26	167	10
		requesting-showing eagerness/ readiness	84	7	236	7
		requesting-vowing	108	6	245	12
		requesting-winning sympathy	178	10	446	21
Winning	To demonstrate	winning favorable	14	1	19	1

favorable impression	the speaker's integrity.	impression-complimenting				
		winning favorable impression-expressing gratitude	17	2	157	8
		winning favorable impression-showing consideration	116	9	174	8
		winning favorable impression-showing honesty	35	3	157	7
		winning favorable impression-showing modesty	0	0	29	2
		winning favorable impression-showing submission	15	1	46	1
Total			3230	252	6038	377

2. As a whole, NSs produced about 50% more utterances (377 versus 252). However, NNSs have more utterances in ‘excusing-forgot’, ‘requesting-new appointment’, ‘winning favorable impression-showing consideration’ and ‘requesting-showing eagerness’. On the contrary, NSs have disproportionally more utterances in ‘apologizing-asking for forgiveness’, ‘excusing-gloss over’ and ‘excusing-testifying’. It seems that these types worth a closer investigation, because there might be cultural implications on top of linguistic ones.

3. Among the 27 specific types, the distribution is far from even. For both NNS and NS data, the mostly produced 9 types (one third) account for 504 utterances out of 629 (80.1%). Seemingly some function types are more common and critical than others under this context.

4. Among the specific function types in each broad type, the distribution is also very uneven. For example, under broad function type ‘Apologizing’, the top 2 (‘apologizing-overt apologies’ and ‘apologizing-self-reproving’) out of the 6 specific types account for 90.7% of total utterances. This might suggest that some specific types are more core than others in apologizing.

5. Similar to point 4, distribution across broad function types are not even either. While some seem more central, some are peripheral under this context. For instance, while ‘Excusing’ contains 265 utterances (42.1%), ‘Winning favorable impression’ only contains 59 utterances (9.3%) and ‘Discourse devices’ only 8 (1.2%).

In the following sections, each of the broad types, and each of the specific types are analyzed, one by one, except for those with too few occurrences, and 'excusing-details', which are narrative in nature and not typical speech acts (see IV.3.1.2 below). For each specific type, besides some examples as illustrations, two more kinds of information are provided: number of utterances classified as Nonnative-like utterances (NNL), Native-like but Untypical utterances (NL-UT), and Native-like and Typical utterances (NL-T) and number of users, so that it can be seen clearer how good (syntactically/semantically and pragmatically) the utterances are, and how 'popular' a certain function type is, i.e. whether used by only a few participants or by many. The TSSSs are also highlighted.

IV.3.1.1 Analysis of Utterances in Broad Function 'Apologizing'

Table VI.19 demonstrates that participants use lots of utterances to express apologies overtly, to reprove of themselves and to ask for forgiveness. Our analysis will concentrate on these.

We can also tell from the last line of Table VI.19 that the quality of utterances is quite good. Vast majority of NNS production and almost all NS production are satisfactory, i.e. belong to NL-T (Native-like and Typical utterances).

Table VI.19 Specific Function Types in Broad Type (Apologizing)

		NNS				NS			
Specific function types	Examples (might not be the whole utterance)	No. of NL-UT utter-ances	No. of NNL utter-ances	No. of NL-T utter-ances	No. of users	No. of NL-UT utter-ances	No. of NNL utter-ances	No. of NL-T utter-ances	No. of users
asking for forgiveness/ understanding	我请您原谅我. Please forgive me. (Edited Eng.)			1	1			9	9
overt apologies	真对不起. I am so sorry. (Edited Eng.)		4	31	25			49	24
self-reproving	这是我的错. This is my fault. (Edited Eng.)	6	2	4	9	5		16	15
showing regret	我想起来的时候也非常懊悔. I regret so much whenever I think about this. (Edited Eng.)							2	2
Total		6/ 12.5%	6/ 12.5%	36/ 75%		5/ 6.2%	0/ 0%	76/ 93.8%	

IV.3.1.1.1 Analysis of 'Asking for Forgiveness/ Understanding' Utterances

As shown in the 3rd line of Table IV.19, all the utterances in this group are of very good quality. They are all classified as Native-like and Typical utterances (NL-T), and are free of errors. However, the great disparities in number of utterances and number of users (1 NNS versus 9 NSs) seem worthy of deeper investigation.

Table IV.20 Task-specific sentence stems (TSSS) in 'asking for forgiveness/understanding'

	NNS freq.	TSSSs (with sample extensions)	NS freq.
①	0	*NP1*[157] *希望 NP2 能 (够) (再) 原谅 NP1 (一次).* NP1 hope NP2 can again forgive NP1 once. (Lit. Eng.) NP1 hope NP2 can forgive NP1 (one more time). (Edited Eng.)	3
②	1	*请 (NP) 原谅 (NP).* Ask (NP) forgive (NP). (Lit. Eng.) Please forgive NP. (Edited Eng.)	3
③	0	*NP1 请求 NP2 原谅.* NP1 beg NP2 to forgive. (Lit. Eng.) Please forgive NP1. (Edited Eng.)	1
④	0	*看 NP1 能不能 原谅 NP2 再次 缺考.* See NP1 can not can forgive NP2 again miss the test (Lit. Eng.) NP2 would like to see if NP1 could kindly forgive NP2 for missing the test again. (Edited Eng.)	1
⑤	0	*希望 NP 能够 体谅.* Hope NP can understand. (Lit. Eng.) Hope NP1 can understand (NP2's situation). (Edited Eng.)	1

As tabulated in above table, NSs produced 9 utterances with 5 forms of TSSSs, all with *希望, 请, 请求* or *看* as the first verb representing the speakers' action, and with *原谅 or 体谅* as the second verb representing the action expected from the listeners. The TSSS used by NNS is one of the

[157] Theoretically, most *NPs* in the TSSSs identified in this study can be realized as *您/你* or 我, just like the *you* in the lexicalized sentence stem *'NP be-TENSE sorry to keep-TENSE you waiting'* (Pawley and Syder 1983:210). However, for simplicity sake, *NPs* are used instead because the notions of *您/你* and 我 can be expressed with more varieties in Chinese. Respectful *老师, 老师您, 张博士您, 博士您*, etc were used instead of *您/你*. On the other hand, though not found in the mini corpus for this study, self-depreciatory expression like *学生* can be used instead of *我*.

five[158].

One point worth noting is that while TSSS forms ①, ②, ③ and ⑤can be viewed as rather fixed Interactional FSs, ④ can also be viewed as a semi-fixed one with a VP slot to be used in other similar situations:

Exp. IV.01

看 NP1 能 不 能 原谅 NP2 再次 VP.

See NP1 can not can forgive NP2 again VP. (Lit. Eng.)

NP2 would like to see if NP1 could kindly forgive NP2 for VP again. (Edited Eng.)

Another interesting fact is NS utterances are significantly longer than NSSs' as shown in Table IV.18 (on the average 11 versus 7 characters), due to a frequent use of extensions (Pawley and Syder 1983:210). Among the 9 NS utterances, 4 employed extensions like *对您带来的不便* (Sentence Crown frame meaning 'for the inconvenience caused to you'), *再 VP 一次* (Sentence Core frame meaning 'VP one more time'), *在这一点上* (Sentence Crown frame meaning 'concerning this point'), and *只能说 VP 了* (Sentence Core frame meaning '(though with hesitation) I have to say'), as illustrated below:

Exp. IV.02

所以 只 能 说 请 您 原谅 了

So only can say beg you to forgive PRT (Lit. Eng.)

So I dare not say anything but to beg for your forgiveness (Edited Eng.)

Exp. IV.03

我 希望 您 能够 再 原谅 我 一 次

[158] Sum of NSS and NS frequencies equal the total number of utterances in this case because all utterances are of satisfactory quality, but it might not be so if some utterances are Nonnative-like utterances (NNL) or Native-like but Untypical utterances (NL-UT).

I hope you can again to forgive me one time (Lit. Eng.)

I hope you can forgive me one more time (Edited Eng.)

Concerning the great disparity in number of users of this type of utterances, the researcher checked the textbooks used by the NNSs and found that the verbs *原谅* and *体谅* were not covered in the two-year course they took. The researcher also found that the only NNS utterance using *原谅* was produced by an English-speaking ethnic Chinese, seemingly acquired elsewhere. As stated above, the language task at issue is quite complicated and embarrassing, and many might felt the need to employ a speech act of 'asking for forgiveness' on top of offering apologies etc. However, as the verb *原谅* and *体谅* are indispensable in realizing the function of 'asking for forgiveness', the disparity can be well explained.

IV.3.1.1.2 Analysis of 'Overt Apologies' Utterances

This group has the highest occurrence under broad function type 'Apologizing'. As shown in the 5th line of Table IV.19, most of the 35 NNS and 49 NS utterances in this group are classified as Native-like and Typical (NL-T) utterances, and there are 4 Nonnative-like (NNL) utterances in NNS group. All the NNLs contain the same mistaken TSSS: '*NP Adverb 对不起*'[159]. It seems that there are two possible sources of such error: L1 inferences and textbook or teacher-caused factors[160].

[159] The behavior of *对不起* is quite different from 'sorry', as it is used either with or without extensions and without any subject (e.g. *对不起！很对不起！太对不起了！*) or with the conveyer of apology as its subject and receiver of the apology as its object. In the latter case, whenever there is a subject, there must be an object. In CCL corpus, 276 out of 278 tokens of *我对不起* are followed by objects (e.g. *我对不起 她/学校/大家/父母/自己/国家*). The only two exceptions are structurally different.

[160] Firstly, the 4 errors were produced by 3 English speakers and structurally they are very similar

Table IV.21 shows the distribution of the 5 FSs meaning 'sorry' or 'apology' functioning as the key word strings in 84 'overt apology' utterances (see below). It is as clear as crystal that NNSs rely on a smaller number of choices. Almost 70% of NNSs chose *对不起* (including the above-mentioned 4 mistaken ones) but only 40% NSs did so. In NNS group, occurrence of *对不起* is more than 3 times of *不好意思* while in NS group, occurrences of them are almost identical. NSs employed all 5 FSs while NNSs only employed 3.

Table IV.21 FSs functioning as keywords in 'overt apology' function type

	NNS	percentage	NS	percentage
对 不 起 (verb-complement collocation literally meaning 'cannot face')	24	68.6%	19	38.8%
不好意思 (adjectival polyword literally meaning 'feel embarrassed')	6	17.1%	18	36.7%
抱歉 (verb-object collocation literally meaning 'embrace apology')	5	14.3%	9	18.4%
表达歉意 (verb-object collocation literally meaning 'express apology' (formal))	0	0.0%	1	2.0%
道歉 (verb-object collocation literally meaning 'express apology')	0	0.0%	2	4.1%

NNS utterances in this section are again shorter, though not as significantly as in 'asking for forgiveness/understanding', than NSs' as shown in Table IV.18 (on the average 6.9 versus 9.5 characters) because they used fewer adverbs or adverbial FSs (see Table IV.22 below).

Table IV.22 Task-specific sentence stems (TSSS) in 'overt apology'

	NNS freq.	TSSSs (with sample extensions)	NS freq.
		Core FS 1: 抱歉	
①	2	*(真的) (真的(是)) 很抱歉 (很抱歉).* (really) (really) very sorry (very sorry). (Lit. Eng.) (so so) so sorry! (Edited Eng.)	5
②	3	*(真的(是)) (感到) (非常) 非常 (的) 抱歉.* (really) (feel) (very) very sorry. (Lit. Eng.) (really) feel so sorry about this! (Edited Eng.)	3
③	0	*NP 要对 NP 说 一句 万二分 的 抱歉.* NP want to NP say one sentence twelve thousand unit PRT sorry.	1

to 'I am sorry' in English (no object is needed in the end). Secondly, the researcher found that the disparity in the grammatical behavior of '*对不起*' and 'sorry' was not mentioned in the textbooks the NNS participants used and was not even noticed by most of their CSL teachers, and consequently not conveyed to students (personal communications).

		(Lit. Eng.) NP would like to extend millions of apologies to NP. (Edited Eng.)	
		Core FS 2: 不好意思	
①	6	*(真 (的)) (是) (很) 不好意思 (啊).* (really) (very) sorry (PRT) . (Lit. Eng.) (really) so sorry! (Edited Eng.)	12
②	0	*NP 觉得 (真的是) 非常 (非常) 不好意思.* NP feel (really) very (very) sorry. (Lit. Eng.) NP be (really) so sorry! (Edited Eng.)	3
③	0	*实在 不好意思.* Indeed sorry. (Lit. Eng.) Really sorry. (Edited Eng.)	2
④	0	*太 不好意思 了.* Too sorry PRT. (Lit. Eng.) So sorry ! (Edited Eng.)	1
		Core FS 3: 对不起	
①	18	*(真 (的)) (很) 对不起 (啊).* (really (PRT)) (very) sorry (PRT). (Lit. Eng.) So sorry ! (Edited Eng.)	9
②	1	*实在 (是) 非常 (非常的) 对不起.* Indeed (be) very (very PRT) sorry. (Lit. Eng.) (so so) so sorry ! (Edited Eng.)	8
③	0	*实在是 太 对不起 了.* Indeed too sorry PRT. (Lit. Eng.) (so so) so sorry ! (Edited Eng.)	1
④	1	*NP1 要 向 NP2 说 对不起.* NP1 want towards NP2 say sorry. (Lit. Eng.) NP1 needs to apologize to NP2. (Edited Eng.)	0
⑤	0	*对不起了对不起了.* sorry PRT sorry PRT. (Lit. Eng.) So sorry so sorry ! (Edited Eng.)	1
		Core FS 4: 道歉	
①	0	*NP 得 道歉.* NP need make apology. (Lit. Eng.) NP needs to apologize. (Edited Eng.)	1
②	0	*NP1 向 NP2 道歉.* NP1 towards NP2 apologize. (Lit. Eng.) NP1 apologize to NP2. (Edited Eng.)	1
		Core FS 5: 表达歉意	
①	0	*NP1 向 NP2 表达 歉意.* NP1 towards NP2 express apology. (Lit. Eng.) NP1 would like to express apologies to NP2. (Edited Eng.)	1

As shown above, the 5 FSs function as cores of the TSSSs. Around them are various extensions, most of which were produced by NSs, that can be grouped to make the picture clearer. The fewer extensions used by NNSs can be illustrated by their 6 utterances using *不好意思* (on the average only one character was added):

Exp. IV.04

不好意思;

Exp. IV.05

不好意思　　啊;

Exp. IV.06

真　　不好意思;

Exp. IV.07

真　　不好意思;

Exp. IV.08

很　　不好意思;

Exp. IV.09

真的　　不好意思;

On the other hand, while there was a NS who also produced an utterance with barely *不好意思*, on the average they added 6 characters and many used two adverbs, such as the following:

Exp. IV.10

真的是		*很*	*不好意思*	
Really		very	sorry	(Lit. Eng.)
polyword Adverb +		Adverb +	core FS	(Structure)

The following used a Sentence Core frame *太... 了* around *不好意思*:

Exp. IV.11

太		*不好意思*		*了*	
Too		sorry		PRT	(Lit. Eng.)
head of frame	+	core FS	+	end of frame	(Structure)

And the next one even had a complicated VP after *不好意思*:

Exp. IV.12

觉得	*非常*	*非常*	*不好意思*	*让*	*您*	*等*	*那么*	*长*	*时间*
feel	very	very	sorry	keep	you	wait	such	long	time (Lit. Eng.)
Verb + repeated Adverbs			+ core FS		+ VP				(Structure)

The NSs also demonstrated their superiority in terms of formality, such as:

Exp. IV.13

NP1	*诚心诚意地*	*向*	*NP2*	*表达*	*歉意.*	
NP1	most sincerely	towards	NP2	express	apology	(Lit. Eng.)
Sub. + formal Adverb		+ Preposition Obj. +		formal core FS		(Structure)

and repeated usage of complete utterances[161]:

Exp. IV.14

对不起	*了*	*对不起*	*了*	
Sorry	PRT	sorry	PRT	(Lit. Eng.)
complete utterance	+	complete utterance		(Structure)

[161] Repeating short utterances of this kind to show eagerness, sincerity, hospitality, etc. is quite common in Chinese, such as *早, 早, 早* (Good morning, good morning, good morning), *你好, 你好, 你好* (Hello, hello, hello), 请坐, *请坐* (please take a seat, please take a seat).

The table also shows that the (groups of) extensions NNSs chose to go with *抱歉*, *不好意思* and *对不起* are quite similar to NSs'. Under *抱歉*, NNSs produced 2 and 3 utterances in the 1st and 2nd groups, in which NSs also produced the most utterances. Under *不好意思*, all 6 NNS output fall into 1st group which NSs favor the most, and under *对不起*, NNSs' favorable choice coincide with NSs again. In summary, NNSs were able to choose the right core FSs and the right (groups of) extensions, though not as sophisticated as NSs'. This seems to be a good indication of NNSs' satisfactory mastery of 'overt apology' speech acts, though there is a disparity in terms of extensions.

Statistically and intuitively, 'overt apology' seems to be a core function for the task at issue.

IV.3.1.1.3 Analysis of 'Self-reproving' Utterances

As depicted in Table IV.18 and IV.19, among broad function type 'Apologizing', both NNSs and NSs devoted the 2nd most of their output to specific type 'Self-reproving', but with the most undesirable quality, esp. in NNS data. Among 12 NNS utterances, 6 are NL-UT and 2 are NNL. And Among the 6 NL-UTs, 4 involved the misuse of *借口*[162] (= excuse/pretext), as illustrated below:

Exp. IV.15

(X)	*我*	*真*	*没办法*	*给*	*您*	*一*	*个*	*借口.*	
	I	really	unable	give	you	one	unit	excuse.	(Lit. Eng.)
(X)	I really can't make up an excuse for you (to …….)								(Edited Eng.)

As an utterance constructed to admit one's own fault, this sounds like a reply (from a person who for some reason was reluctant to tell a lie, for instance) to a request like 'Just make up some excuses so that I can use to help you'. Apparently the NNSs had not mastered the native ways of expressing the

[162] All the examples of *借口* found in The Contemporary Chinese Dictionary (2002) and The Standard Dictionary of Contemporary Chinese (2004), just to name two, are with negative connotations.

English phrase *I have no excuse* in Chinese and just translated it word for word without noticing the cross-linguistic differences (Aijmer 1996:25)[163].

The next utterance has been labeled NNL because it is very difficult to imagine under what circumstances a NS would produce such a word string, though the meaning can be conveyed clearly: 'I missed the makeup test out of my stupidity', with a seemingly grammatical structure (i.e. NP Verb-be Adv. Adj. PRT NP).[164]

Exp. IV.16

(X)	*我*	*是*	*真*	*笨*	*的*	*学生.*	
	I	am	really	stupid	PRT	student.	(Lit. Eng.)
(X)	I am a really stupid student.						(Edited Eng.)

TSSSs identified in 'Self-reproving' are tabulated in Table IV.23 below. Among the 4 NL-T TSSSs produced by NNS, 3 are similar to one of NSs' favorable types, again showing NNSs' reliance on a small number of common TSSSs. And as revealed by previous sections, NNSs produced shorter utterances because they used fewer adverbs, i.e. *却*, *竟然* and *又* etc, and fewer Sentence Crown frames, i.e. *无论如何* and *不管怎么样*, etc. Besides, they lacked very native inventories like the 11th TSSS, which is fixed in form and clear in intention, yet composed of very basic patterns and lexis[165]. Actually, it is not surprising for NNSs to do quite unsatisfactorily in this specific function. A glance at the TSSSs in Table IV.23 shows that most of them are syntactically or even culturally challenging. One utterance

[163] See Table IV.23 & IV.43 below for how NSs self-reprove and how they use *借口*.

[164] Although sentences like *他是很好的学生* abound in Chinese, no *真笨的 NP* can be found in CCL corpus. Moreover, all the 63 tokens of *真笨* are used in the end of a sentence, such as *你（可）真笨！*, except for three which are followed by exclamation particle 啊 or question particle 吗. On the other hand, *NP 真是个笨 NP* seems to be a probable sentence frame and has five token in CCL, but does not seem suitable for the task at issue. It can tentatively be concluded that as an adverb, *真* and *很* go with different sentence frames and behave substantially differently.

[165] In the CSL program the NNSs took, all the components of this TSSS were covered in the first semester. However, the TSSS as a whole has never been taught and is unlikely to be composed with the syntactic rules learned.

containing the 3rd TSSS produced by a NS is shown below as an illustration:

您 特地 为我 安排 了 今天早上 的 补考, 我 竟然 没 来.

You especially for me arrange PRT this morning PRT makeup test, I unreasonably didn't come. (Lit. Eng.)

You made a lot of effort to arrange the makeup test for me, and I did come for it. (Edited Eng.)

However, statistically and intuitively, 'self-reproving' seems to be a core function for the task at issue. The TSSSs identified need to be carefully evaluated in order to decide which, when and how many of them need to be introduced in a course, and even more fundamentally, if language tasks like the one being studied need to be included.

As a expedient measure to instantly improve NNSs' performance, *都是我的错* or *都是我不好* as whole phrases can be introduced with a suitable task.

Table IV.23 Task-specific sentence stems (TSSS) in 'Self-reproving'

	NNS freq.	TSSSs (with sample extensions)	NS freq.
①	3	*(无论如何/) (这) (都) 是 NP 的 错（误）/不对.* (in any event)(this) (all) is NP PRT fault. (Lit. Eng.) It is all NP's fault. (Edited Eng.)	2
②		*NP1 VP1, NP2 (却) 错过 了 一 次 考试/机会.* NP1 VP1, NP2 (unreasonably) miss PRT one unit test/chance. (Lit. Eng.) NP1 VP1, but NP2 (unreasonably) missed the test/chance. (Edited Eng.)	4
③		*NP1 VP1, NP2 竟然/又/还 没 来.* NP1 VP1, NP2 unreasonably/again/yet didn't come. (Lit. Eng.) NP1 VP1, but NP2 didn't show up again. (Edited Eng.)	4
④		*(不管怎么样) NP 错了/犯了错误.* (in any event) NP wronged. (Lit. Eng.) In any event, it is all NP's fault. (Edited Eng.)	2
⑤		*(无论如何) 责任/错误 在（于） NP.* (in any event) responsibility/fault on NP. (Lit. Eng.) In any event, it is all NP's fault. (Edited Eng.)	2
⑥		*NP 犯 了 不可 原谅 的 错误.* NP made PRT cannot forgive PRT mistake. (Lit. Eng.) NP's made an unforgivable mistake. (Edited Eng.)	1
⑦		*NP1 浪费 了 NP2 的 时间.* NP1 waste PRT NP2 PRT time. (Lit. Eng.) NP1 wasted NP2's time. (Edited Eng.)	1

⑧		*NP1 对 NP2 表示 了 不 尊重.* NP1 to NP2 show PRT not respect. (Lit. Eng.) NP1 did not show duly respect to NP2. (Edited Eng.)	1
⑨		*NP1 没有 办法 跟 NP2 解释.* NP1 have not means to NP2 explain. (Lit. Eng.) NP1 doesn't have any excuses. (Edited Eng.)	1
⑩	1	*NP 真 糊涂 了.* NP really muddled PRT. (Lit. Eng.) NP was so stupid. (Edited Eng.)	
⑪		*千 错 万 错 都 是 NP 的 错.* thousand faults ten thousand faults all be NP PR fault. (Lit. Eng.) It is all my fault. (Edited Eng.)	1

IV.3.1.1.4 Interim Summary:

The disparities shown above between NSs and NNSs show clearly that NSs tend to have more extensions, such as in *希望NP 能 (够) (再) 原谅 NP (一次)* and *(真 (的)) (是) (很)不好意思* etc. It can be argued that, if the learners' aim is to attain very high proficiency, the extensions should be systematically included in the curriculum because of their high usage by NSs. However, when doing this, on top of frequency, syntactic complexity and cultural differences also need to be taken into consideration. While *(真 (的)) (是) (很)* seems to be teachable even at elementary levels, *希望NP 原谅 NP, 希望NP 能原谅 NP, 希望NP 能够原谅 NP, and 希望NP 能够再原谅 NP 一次* can be introduced at intermediate and advanced levels, or even not covered in textbooks at all (one such extreme example can be found in IV.3.1.2.3 below on 'gloss-over').

However, for TSSSs that are statistically and intuitively useful but syntactically and culturally challenging, such as 'self-reproving', it seems that a more lexical and task-based approach should be adopted: introduce phrases like *都是我的错* or *都是我不好* as FSs and touch on the grammar points involved minimally.

IV.3.1.2 Analysis of Utterances in Broad Function 'Excusing'

Table IV.18 and IV.24 show that participants use slightly more than half of Chinese characters and utterances in this broad function type to describe the details of how they missed the makeup test. Most of them are purely descriptive utterances that can collectively form a complete and sensible story to explain why

they missed the makeup test but, when standing alone, might not denote typical speech acts[166]. As confined by the scope and focus of this study, we will only analysis the specific types on how they talked about what kept them from coming for the test (excusing-exonerating), how they tell the professor frankly or skillfully that they forgot (excusing-forgot and excusing-gross over), and how they testify that they missed the test with no bad intentions and how well-prepared they were.

A point worth noting here is that the quality of utterances under this broad function type is significantly lower than that of 'Apologizing', with merely over half of NNSs' and slightly less than 90% of NSs' classified as NL-T, as shown at the bottom of the table, indicating that 'Excusing' might be a more complicated and demanding language function than 'Apologizing', and/or this language function was not properly taught in the course the NNSs took.

Table IV.24 Specific Function Types in Broad Type (Excusing)

		NNS				NS			
Specific function types	Examples (might not be the whole utterance)	No. of NL-UT utter-ances	No. of NNL utter-ances	No. of NL-T utter-ances	No. of users	No. of NL-UT utter-ances	No. of NNL utter-ances	No. of NL-T utter-ances	No. of users
excusing-detail	昨晚我复习功课复习晚了. I reviewed until very late last night (Edited Eng.)	12	23	36	24	8		77	20
excusing-exonerating	这个真的是没办法. This is really out of my control (Edited Eng.)		1	6	6	4		11	9
excusing-forgot	所以把这个补考的事情给忘记了. Therefore I forgot about the makeup test (Edited Eng.)	2	4	18	17	1		19	14
excusing-gloss over	就是不知道为什么突然就忘记了. Don't why but suddenly forgot (Edited Eng.)		1		1	3		11	9
excusing-testifying	我不是故意缺考的. I did not miss the exam deliberately (Edited Eng.)	1	1	4	2	1		23	13
Total		15/ 13.8%	30/ 27.6%	64/ 58.7%		17/ 10.8%	0/ 0%	141/ 89.2%	

[166] Most of the sentences in the story-telling parts in the data collected for this study are of predicating nature, such as *我同学昨天晚上生病了* (= my classmate was sick last night) and *这段时间比较忙* (= recently I have been quite busy).

IV.3.1.2.1 Analysis of 'Exonerating' Utterances

One fourth of the participants (6 NNSs and 9 NSs) employed 22 utterances (7 by NNSs and 14 by NSs) to show that though they indeed missed the makeup test, it was not entirely their faults or there were something beyond their control. As shown in the table below, 7 groups of TSSSs have been used, 3 by NNSs and 6 by NSs. The most popular TSSS is ① *没(有) 办法 VP* (meaning 'unable to VP' or 'there's no way I could VP'), which was chosen by 2 NNSs and 3 NSs. A slight difference between NNSs and NSs in using *没(有) 办法* can be demonstrated by the following two examples:

Exp. IV.17

我	*没办法*	*来*	*这里.*	
I	unable	come	here.	(Lit. Eng.)
I could not come here.				(Edited Eng.)

Exp. IV.18

真的是	*没办法.*	
Really	no way out.	(Lit. Eng.)
I really could not make it.		(Edited Eng.)

While both NNSs used the TSSS in the way as the first example, only one NS did, and the other two did it in the way as the second example. The second way is less specific (because the VP is dropped), but with stronger impact. It seems that NSSs have narrower varieties of TSSSs.

The most popular core FSs among the NNSs is ③ *不能* (= unable) and all the 3 utterances by three different participants are structurally identical as illustrated below:

Exp. IV.19

所以	*不*	*能*	*VP.*

Therefore	not	able	VP.	(Lit. Eng.)
conjunction	TSSS		VP.	(Structure)

The above TSSS was not chosen by any NSs seemingly due to its syntactic, semantic and rhetoric simplicity. *不能* is the commonest way to express the notion 'cannot' and is introduced in very early stages of any CSL curriculum. The utterances are accepted by judgers as NL-T but not employed by NSs.

Table IV.25 Task-specific sentence stems (TSSS) in 'Exonerating'

	NNS freq.	TSSSs (with sample extensions)	NS freq.
①	2	*(NP) (真的 (是)) 没(有) 办法 (VP).* (NP) (really) haven't means (VP). (Lit. Eng.) (NP) (really) could not find a way (to VP). (Edited Eng.)	3
②		*这次 (真的(是)) 有特殊的原因/情况比较特殊/有点儿原因/是有原因的/事出有因* This time (really (is)) have special PRT reason/situation comparatively special/have some reason/PRT have reason PRT/things happen have cause (Lit. Eng.) There was some special situation/the situation was special/there were some reasons/it was not without cause/apparently with a cause this time (Edited Eng.)	5
③	3	*所以 不 能 VP.* Therefore not can VP. (Lit. Eng.) Therefore I couldn't VP. (Edited Eng.)	
④	1	*其实 NP1 记得 要 VP / NP1 TIME 还 记得清清楚楚.* Actually NP1 remember need VP/ NP1 TIME still remember clearly. (Lit. Eng.) Actually NP1 did not forget about VP/ NP1 still remembered this so clearly TIME. (Edited Eng.)	1
⑤		*其实 NP 很 早 就 VP 了/其实 也是 有 点儿 原因 的.* Actually NP very early already VP PRT/Actually also have some reason PRT. (Lit. Eng.) Actually NP VP long time ago/Actually there were something out of NP's control. (Edited Eng.)	2
⑥		*TIME 还 特意 VP 来 VP.* TIME actually intentionally VP to VP. (Lit. Eng.) Actually NP VP to VP TIME. (Edited Eng.)	1
⑦		*NP 也 没有 想到 会 突然 发生 这 种 事情.* NP also didn't think will suddenly take place this kind thing. (Lit. Eng.) NP did not expect something like this to happen so suddenly. (Edited Eng.)	1

In ②, NSs adopted 5 utterances with *（原）因* (= cause) or *特殊* (= special) or both to show that they did not miss the test without justified causes. The most elegant one among them is presented below:

Exp. IV.20

这　一　次　　真　的　是　　事出有因.

This one time really PRT be not without a cause. (Lit. Eng.)

This time there was a good reason. (Edited. Eng.)

Polyword Noun polyword Adverb core FS. (Structure)

IN ④ and ⑤, three participants (1 NNS and 2 NSs) employed adverb *其实* (= actually) plus a VP to show that they did not forget the test and did try their best to come, but in vain because of something uncontrollable, such as the following:

Exp. IV.21

其实　　我　很　　早　　就　起来　　了.

Actually I very early already get up PRT. (Lit. Eng.)

Actually I got up very early this morning. (Structure)

The next utterance (⑥) is used by NS to show the effort he had made before regrettably forgot about the test:

Exp. IV.22

早上　　还　　特意　　醒　了　一下 来　想一想
有　没有　　什么　　事情.

Morning actually intentionally wake PRT a bit to think a bit
have haven't what thing. (Lit. Eng.)

Actually I intentionally woke up for a while this morning to see if there was anything I needed to do today. (Edited Eng.)

IV.3.1.2.2 Analysis of 'Forgot' Utterances

The majority of the participants in both groups employed one or more utterances to mention that they forgot about the test. Among the 57 utterances they used, 44 are very straightforward (just admit that they forgot) but 13 were expressed in a more skillful way that sounds more subtle. Therefore they are categorized into two specific function types. 'Forgot' (the former) will be analyzed in this section and 'Gloss-over' (the latter) will be handled in the next section.

As shown in Table IV.18 and IV.24, NNSs produced 24 'Forgot' utterances (4 more than NSs) but only 1 'Gloss-over' (far fewer than 12 by NSs) most likely because of the higher language proficiency required in producing the latter. The table below shows the TSSSs adopted in 'Forgot' utterances.

Table IV.26 Task-specific sentence stems (TSSS) in 'forgot'

	NNS freq.	TSSSs (with sample extensions)	NS freq.
①	9	*NP (完全/ 就/ 确实是/ 突然间) 忘（记） 了 (VP (的事情)).* NP (completely/then/indeed/suddenly) forget PRT (VP (PRT thing)). (Lit. Eng.) NP (actually/completely/suddenly) forgot about VP. (Edited Eng.)	9
②		*(NP) (就/确实是/突然间/一事) (把) (VP (的事情)) (给) 忘(记)了.* (NP) (then/indeed/suddenly/accidentally) (PREP.) (VP (PRT thing)) (AUX.) forget PRT. (Lit. Eng.) NP (actually/completely/suddenly) forgot about VP. (Edited Eng.)	11
③	6	*NP1 (完全) 忘（记）了 (NP1) 有 NP2/需要 VP.* NP1 (completely) forget PRT (NP1) have NP2/need VP. (Lit. Eng.) I (completely) forgot about VP. (Edited. Eng.)	
④	1	*NP (就是) 记不起来.* NP (simply) couldn't remember. (Lit. Eng.) NP just forgot. (Edited. Eng.)	

Only four types of TSSSs have been found, 3 by NNSs and 2 by NSs. NNSs produced 13 utterances in the first type, mostly classified as NL-T, but, again,

with fewer number and varieties of adverbs than NSs (altogether only 1 *完全* and 1 *就* were used). For example:

Exp. IV.23

我　　完全　　忘记　　了.

I　　completely　　forget　　PRT.　　(Lit. Eng.)

I totally forgot.　　(Edited Eng.)

NSs tend to have more extensions as illustrated below

Exp. IV.24

忙　起来　就　　突然间　　忘记　　了　　补考　　的　　事情

Busy begin　then　　suddenly　　forget　　PRT　　makeup　　PRT　　matter(Lit. Eng.)

While I was busy with something I suddenly forgot about the makeup exam.　(Edited Eng.)

While 9 NSs also employed first type of TSSS, the second type was favored by 11 NSs but none of the NNSs. On top of more varieties of adverbs, these 11 NSs used a frame with preposition *把* (normally found before an objects which is placed before the verb) and auxiliary verb *给* (often used before the verb to add a negative implication), which is one of the most difficult patterns in CSL but very common among native speakers of Chinese, especially in spoken language. An example is given below:

Exp. IV.25

我　真的　　把　　这　件　事　　给　　忘　了.

I　　really　　PREP.　　this　　unit　　matter　　AUX.　　forget PRT.　(Lit. Eng.)

I couldn't believe I forgot all about this.　　(Edited Eng.)

NNSs produced 8 utterances, again, mostly classified as NL-T, in the third type, which is very similar to the first type, except mostly with an additional verb before *考试/补考* in the end. For example:

Exp. IV.26

我	*忘*	*了*	*我*	*有*	*考试.*	
I	forget	PRT	I	had	test.	(Lit. Eng.)
I forgot that I had a test to take.						(Edited Eng.)

The only utterance in the forth type was produced by an ethnic Chinese NNS:

Exp. IV.27

我	*就是*	*记不起来.*	
I	just	couldn't remember.	(Lit. Eng.)
I just forgot.			(Edited Eng.)

As seen above and below, NNSs employed shorter extensions and simpler syntactic structures when producing TSSSs of this type.

Statistically, 'forgot' should be a core function for the task at issue (especially when the 'gloss-over' below is combined with this one). Syntactically, it can be as short and simple as *我突然忘了*[167]. All the elements are normally introduced in elementary or intermediate course books. If tasks like the one being studied need to be introduced in early stage of a certain course, the phrase can be taught as a FS with minimal explanation on its syntactic

[167] Actually the shortest version should be *我忘了*, but it does not sound very sincere when standing alone.

structure.

IV.3.1.2.3 Analysis of 'Gloss-over' Utterances

This type of utterances was employed to create a subtle atmosphere to cover up the embarrassment. As mentioned in last section, NNSs produced far fewer 'Gloss-over', because of the higher language proficiency involved. The table below shows the TSSSs adopted in 'Gloss-over' utterances.

Table IV.27 Task-specific sentence stems (TSSS) in 'Gloss-over'

	NNS freq.	TSSSs (with sample extensions)	NS freq.
①		*(NP) 不 知道 怎么样/为什么/忙 些 什么（就）（忽然） VP 了.* (NP) don't know how/ why/ busy some what (then) (suddenly) VP PRT. (Lit. Eng.) Don't know how/why but just (suddenly) VP. (Edited Eng.)	6
②		*(NP) 也 不 知道 是 怎么 了.* (NP) also not know be how PRT. (Lit. Eng.) (NP) really don't know why but …… (Edited Eng.)	2
③		*NP 有 点儿 头脑 发昏, VP 了.* NP have some mind muddled, VP PRT. (Lit. Eng.) NP was not clear-minded and just VP. (Edited Eng.)	1

As shown in Table IV.24 and IV.27, among 11 utterances by NSs, 3 are pragmatically problematic, showing that even some educated native speakers are not good with this type of speech acts. The only utterance by NNS is formally and semantically similar to the first type of TSSSs in IV.27, but seemingly a word-for-word translation from Japanese and consequently not counted as a TSSS:

Exp. IV.28

我 不 知道 为 什么 可是 我 忘 了.

I not know for what but I forget PRT. (Lit. Eng.)

I don't know why but I forgot. (Lit. Eng.)

Two utterances by NSs are presented below as illustrations of first and second

type TSSSs respectively:

Exp. IV.29

不　知道 为 什么　把　考试　的　事情　给　忘　了.

Not know　for what　PREP. take test　PRT　matter　AUX.　forget PRT. (Lit. Eng.)

Don't know why but somehow I forgot about the test.　(Edited. Eng.)

Exp. IV.30

我　也　不　知道　我　是　怎么　了.

I　also　not　know　I　be　how　PRT.　(Lit. Eng.)

I really don't know what's wrong with me.　(Edited. Eng.)

One additional point is worthy of mentioning here: Besides Sentence Core frame *把 NP 给 VP　了*, the noun frame *VP　的事情* (= the matter about VP), was also used many times in 'Gloss-over' and 'Forgot' by NSs only. On the other hand, *完全* was used 4 times but exclusively by NNSs.

'Gloss-over', are culturally and probably psycholinguistically too challenging and not recommended even in intermediate or advanced levels. Fortunately, its function is the same as 'forgot' analyzed above and can be easily replaced.

IV.3.1.2.4　Analysis of 'Testifying' Utterances

As shown in Table IV.18 and IV.24, NNSs produced only 6 utterances of this specific function type (one forth of that of NSs) but they are not significantly shorter in length (15.6 versus 18.8 characters).

The researcher's first impression after his first glance at the 30 utterances was that they were very diverse in form. Therefore, the analysis started with participants' testifying strategies. Strategically, NNSs and NSs testify their integrity with quite similar arguments, as depicted in the table below:

Table IV.28 Arguments in 'Testifying'

NNS freq.	Arguments	NS freq.
3	Conceived no bad intention	8
2	Did not forget/Was occupied by other urgent matters	1
	Gave priority to the course	5
1	Have always been a good student	4
	Had sufficient preparation	6

The table below shows how the arguments were translated into the TSSSs of 'Testifying'.

Table IV.29 Task-specific sentence stems (TSSS) in 'Testifying'

	NNS freq.	TSSSs (with sample extensions)	NS freq.
①	2	*(NP) (真的/确实) 不 是 故意/有心/特意 (VP) 的.* (NP) (really/actually) not be deliberately VP PRT. (Lit. Eng.) NP really did not miss the test deliberately. (Edited. Eng.)	8
②	2	*NP 真 的 没有 VP/ NP 真 的 VP 了 / NP 最近 有 些 事情 VP.* NP really PRT didn't VP/ NP really PRT VP PRT/ NP recently have something VP. (Lit. Eng.) NP really didn't VP/ NP really did VP/ NP had something to VP recently. (Edited. Eng.)	1
③		*NP1 真 的 很 重视 NP2 / NP1 对 NP2 真 的 是 很 重视 /NP1 很 认真 地 对待 NP2.* NP1 really PRT very think highly of NP2 / NP1 toward NP2 really PRT be very think highly of/ NP1 very serious PRT treat NP2. (Lit. Eng.) NP1 is really serious about NP2 /NP1 really take NP2 very seriously/ NP1 take NP2 very seriously. (Edited Eng.)	5
④		*NP1 知道 NP2 不是 那 种 NP3/ NP1 知道 NP2 在 NP3 上 非常 用心/ 不 是 态度 不 对 这 样 的 问题/ NP1 的 NP2 也 都 非常的 好.* NP1 know NP2 not be that kind NP3/ NP1 know NP2 at NP3 up very serious/ Not be attitude not right this kind PRT problem/ NP1 PRT NP2 also all very PRT good. (Lit. Eng.) NP1 know NP2 is not that kind of NP3 /NP1 knows NP2 has been working very hard in NP3/ It's not because NP's attitude was wrong/ NP1's NP2 has always been very good. (Edited. Eng.)	4
⑤		*NP1 为 NP2 VP 了 很长时间/ 非常久/ NP VP 得 很 充分/ 每 次 NP1 前 NP2 都会 VP/ NP 熬 了 几个 通宵 来 VP.* NP1 for NP2 VP PRT very long time/very long/ NP VP PRT very sufficiently/ Every time NP1 before NP2 all will VP/ NP burn PRT several midnight oil to VP. (Lit. Eng.)	5

		NP1 spent a lot of time to VP NP2/ NP VP very sufficiently/ Every time beforeNP1, NP2 VP/ NP did not sleep for a few days to VP. (Edited. Eng.)	

Utterances employed to show that the speakers did not have bad intentions are the most unanimous in form, as shown below:

Exp. IV.31

我 真 的 不 是 故意 的.

I indeed PRT not be deliberately PRT. (Lit. Eng.)

I didn't (do this) deliberately/out of bad intention. (Edited. Eng.)

Exp. IV.32

我 不 是 故意 缺 考 的.

I not be deliberately miss test PRT. (Lit. Eng.)

I didn't miss the test deliberately/out of bad intention. (Edited. Eng.)

A NNS used this utterance to convince the professor that she happened to be occupied by an urgent matter:

Exp. IV.33

我真 的 把 我的 外婆 送 到 医院 去 了.

I indeed PRT (PREP.) my grandma take to hospital go PRT. (Lit. Eng.)

I really took my grandma to the hospital. (Edited. Eng.)

A NS adopted the following utterance to emphasize that the test is of top priority. Note that many adverbial FSs were used probably as a means to intensify the argument as well as fluency devices or hesitant fillers to cover

up the embarrassment:

Exp. IV.34

我 真 的 是对 这 个 考试 其实 是 真 的 是 非常 的 重视.

I indeed PRT be to this unit test actually be really PRT be very PRT regard highly. (Lit. Eng.)

I really take the test really very seriously. (Edited. Eng.)

Participants argued in many ways to prove that they had been doing very well in the course and missing the test was not their normal state of behavior, such as the one below:

Exp. IV.35

我 以前 的 出勤 情况 也 都 非常 的 好.

I before PRT attendance situation also all very PRT good. (Lit. Eng.)

My attendance rate has always been very high. (Edited. Eng.)

The last type of argument is employed to prove that the speakers had already prepared sufficiently for the test and they did not miss the makeup that morning in order to buy some more time for reviewing.

Exp. IV.36

其实 我 都 准备 了, 准备 得 很 充分.

Actually I all prepared PRT, prepare PRT very sufficiently. (Lit. Eng.)

Actually I have prepared everything so sufficiently. (Edited. Eng.)

Statistically, ‘testifying’ stands out to be a core function for the task at issue.

Syntactically, it can be as short and simple as *我真的不是故意的.* All the elements are very basic and are normally introduced in elementary course books, except for *故意*. *我真的不是故意的* can be taught as a FS earlier in a course if necessary.

IV.3.1.2.5 Interim summary:

Once again, it has been shown that NSs tend to have more extensions, and consequently more complex structures, which are worth learning if the learners aim at high proficiency levels. However, some type of TSSSs, e.g. 'Gloss-over', are culturally and probably psycholinguistically too challenging and not very common in daily life and thus might not be suitable to be included in textbooks, or may be included only for training receptive skills.

On the other hand, in order to improve NNSs' performance, simple phrases like *我真的是没有办法*, *我突然忘记了*, *我真的不是故意的* can be introduced with tasks and taught as FSs.

IV.3.1.3 Analysis of Utterances in Broad Function 'Requesting'

Table IV.30 demonstrates that participants in both groups employed similar strategies to obtain another chance. They begged, overtly asked for another chance, proposed a new time for next appointment, demonstrated their eagerness, tried to win the professor's sympathy, or even made vows. NSs also used 3 utterances to show their readiness but they will not be analyzed in this research due to their limited occurrence.

The quality of NNS utterances under this broad function type is even lower than that of 'Excusing', with merely over 20% classified as NL-T, as shown at the bottom of the table, indicating that 'Requesting' might be an even more

complicated and demanding language function than 'Excusing'. Quality of NS production is also lower but not significantly.

Table IV.30 Specific Function Types in Broad Type (Requesting)

		NNS				NS			
Specific function types	Examples (might not be the whole utterance)	No. of NL-UT utter-ances	No. of NNL utter-ances	No. of NL-T utter-ances	No. of users	No. of NL-UT utter-ances	No. of NNL utter-ances	No. of NL-T utter-ances	No. of users
requesting-another chance	请再给我一次机会. Please give me one more chance. (Edited Eng.)	4	10	8	19	3		43	27
requesting-general	这个要麻烦老师了. I know it is a huge trouble but please help me. (Edited Eng.)	3	2	1	5	1		10	6
requesting-new appointment	可不可以改天再考? Can I take the test again on another day? (Edited Eng.)	11	10	5	18	1		9	9
requesting-showing eagerness/ readiness	我非常希望参加补考. I really want to take the make-up exam. (Edited Eng.)	3	3	1	5	3		5	7
requesting-vowing	我想您保证下一次的补考我一定不会忘记的. I promise that I will by no means miss the next make-up exam. (Edited Eng.)	3	3		6	2		10	11
requesting-winning sympathy	如果没有这个考试我是毕不了业的. Without the exam I have no chance to graduate. (Edited Eng.)	5	4	1	9	5		16	14
Total		30/ 38.5%	32/ 41%	16/ 20.5%		15/ 13.9%	0/ 0%	93/ 86.1%	

IV.3.1.3.1 Analysis of 'Another Chance' Utterances

As shown in Table IV.18 and IV.30, 19 NNSs and 27 NSs produced 22 and 46 utterances of this type (on the average 1.2 and 1.7 utterances each). About two third of NNS production is problematic.

Almost all utterances (17/22 in NNS and 45/46 in NS) used a similar FS '*NP 给 NP 机会*' (meaning 'NP gives NP a chance') which is a verb-object frame, indicating that most of the TSSSs used in this specific function type

might be more unanimous, as shown in the table below (see TSSSs ① to ⑥).

Table IV.31 Task-specific sentence stems (TSSS) in 'Another chance'

	NNS freq.	TSSSs (with sample extensions)	NS freq.
①	5	***(NP1)*** *(恳)请 NP2 (可不可以)* *(再)* ***给 NP1*** *(一次/个) (VP1 的) 机会 (让 NP1 VP1).* (NP1) beg NP2 (can or cannot) (again) give NP1 (one time) (VP1 PRT) chance (let NP1 VP1). (Lit. Eng.) NP1 would like to ask if NP2 can give NP1 one more chance to VP1. (Edited Eng.)	22
②	1	*希望* ***NP1*** *(能) (再)* ***给 NP2*** *(一次/个) (VP1 的)* ***机会*** *(让 NP2 VP1).* Hope NP1 can (again) give NP2 (one time) (VP1 PRT) chance (let NP2 VP1) (Lit. Eng.) NP2 hope NP1 can give NP2 another chance to VP1. (Edited Eng.)	19
③	4	*请* ***给 NP*** *最后 一个/ 多 一 个* ***机会.*** Please give NP last one/ more one unit chance (Lit. Eng.) Please give NP one last chance. (Edited Eng.)	
④		*(可不可以) (恳) 请 NP1 (再)* ***给 NP2*** *第二次 (VP1 的)* ***机会*** *(让 NP2 VP1)* (can or cannot) beg NP1 (again) give NP2 2nd (VP PRT) chance (let NP2 VP1) (Lit. Eng.) Is it possible that NP1 can give NP2 the 2nd chance to VP1. (Edited Eng.)	2
⑤		*不 知道* ***NP1*** *是 不 是 还 能够* ***给 NP2 机会?*** Don't know NP1 be not be still can give NP2 chance. (Lit. Eng.) NP2 wonders if NP1 can give NP2 one more chance. (Edited Eng.)	1
⑥		*请* ***NP1*** *一定要* ***给 NP2*** *这次 VP 的****机会.*** Beg NP1 definitely should give NP2 this time VP PRT chance.(Lit. Eng.) Oh please please give NP2 one more chance to VP. (Edited Eng.)	1
⑦		*希望 NP1 能够 给*[168] *NP2 VP* Hope NP1 can let NP2 VP. (Lit. Eng.) Hope NP1 can let NP2 VP. (Edited Eng.)	1

Most utterances employed the first type of TSSS. One example below is an imperative while the other is a question:

Exp. IV.37

请 您 再 给 我 一 个 机会 让 我 参加 考试.

[168] As specified in CSL textbooks, e.g. Jamieson & Lin (2002:317), *让* should be used instead of *给*(e.g.*希望您能够让我再考一次*). This is also confirmed by an informal survey of 10 colleagues of the author on 2012/5/10. However, the author has heard quite a few NSs of Chinese other than this participant using *给*. This might be due to inference of dialects.

Beg you again give me one unit chance let me participate test. (Lit. Eng.)

Please give me one more chance to take the test. (Edited. Eng.)

Exp. IV.38

能 不 能 请 您 再 给 我 一 次 机会 呢?

Can not can beg you again give me one unit chance PRT? (Lit. Eng.)

Can I ask you to give me one more chance? (Edited. Eng.)

The second type was also favored by many NSs but is structurally alike. So, the following two examples are given to illustrate the third type of TSSS which was adopted only by NNSs but still valid and sound, and the forth type adopted only by NSs respectively:

Exp. IV.39

请 您 给 我 最后 一 个 机会.

Beg you give me last one unit chance? (Lit. Eng.)

Please give me one last chance. (Edited. Eng.)

Exp. IV.40

可 不 可以 恳请 您 再 给 我 第二次 补考 的 机会?

Can not can beg you again give me second makeup PRT chance?(Lit. Eng.)

May I ask you to give me a second chance to take the makeup? (Edited. Eng.)

Statistically, ‘another chance’ stands out to be a core function for the task at issue. In addition, intuitively it can be used widely in many other speech

events in which speakers need to plead for another chance. Syntactically, it can be as short and simple as *请再给我一个机会*, with all the elements being very basic and are normally introduced in elementary course books.

IV.3.1.3.2 Analysis of 'Requesting-general' Utterances

As shown in Table IV.18 and IV.30, 5 NNSs and 6 NSs produced 6 and 11 utterances of this type (on the average 1.2 and 1.8 utterances each). The vast majority of NNS production is problematic, and the TSSSs used by NSs are relatively simple and short (compared with those analyzed above).

As shown in the table below, almost all acceptable utterances (1/6 in NNS and 10/11 in NS) are typical 'institutionalized expressions' in Nattinger and DeCarrico (1992) and Lewis (1993), or 'interactional routines' in Wray (2002). They are quite fixed in form and specific in manipulative and socio-interactional function (Wray, 2002:250). They are also semantically fused when translated (the meaning of their components are not as clear as when used in other utterances), especially in the first few types.

Table IV.32 Task-specific sentence stems (TSSS) in 'requesting-general'

	NNS freq.	TSSSs (with sample extensions)	NS freq.
①		*麻烦 NP.* Trouble NP. (Lit. Eng.) Please! (Edited Eng.)	3
②		*(那就)/(真的) (要) 麻烦 NP 了.* (Then)/(really) (need) trouble NP PRT. (Lit. Eng.) Please help me! (Edited Eng.)	3
③		*拜托 了.* Request a favor PRT. (Lit. Eng.) Please! (Edited Eng.)	1
④		*请 NP1 一定 要 让 NP2 VP.* Beg NP1 must must let NP2 VP. (Lit. Eng.) Oh please please let NP2 VP. (Edited Eng.)	1
⑤	1	*请 NP1 再/特别 帮 NP2 一 个 忙/一 次 吧.* Beg NP1 again/especially help NP2 one unit favor/one time PRT.(Lit. Eng.) Please do NP1 one more favor/help NP2 one more time. (Edited Eng.)	1

Despite of their structural simplicity and short length on average, these utterances seem unfamiliar to NNSs. This is probably because the NNS participants either did not come across them or did not pay enough attention

when learning them, resulting in lots of NL-UT or NNL. Below is an example which is composed of similar lexical items and syntactically correct, but just sound non-nativelike:

Exp. IV.41

(x) *请 帮 我 忙.*

Beg help me favor. (Lit. Eng.)

(x) Please give me a hand. (Edited. Eng. with pragmatic mistake)

IV.3.1.3.3 Analysis of 'New Appointment' Utterances

Compared with 'Requesting-Another chance' (see below), utterances in 'Requesting-New appointment' type are more concrete requests literally mentioning lexical items such as *安排补考* (= arrange a makeup) or *安排时间* (= fix a time) or proposing a time such as *现在* (= now), as shown in Table IV.33.

Table IV.33 Task-specific sentence stems (TSSS) in 'new appointment'

	NNS freq.	TSSSs (with sample extensions)	NS freq.
①	1	*能 不 能/请 NP1 (再) (给 NP2) (另外)安排 一个/次 机会 (VP).* Can not can/beg NP1 (again) (give NP2) (again) arrange one time chance (VP). (Lit. Eng.) Could NP1/please arrange another chance for NP2 to VP. (Edited Eng.)	7
②		*那 这 次 时间 NP 安排.* Then this time time NP arrange. (Lit. Eng.) NP may appoint any time. (Edited Eng.)	1
③		*(现在) 能 不 能 再 VP 一 次?* (now) can no can again VP one time? (Lit. Eng.) Can NP VP one more time now? (Edited Eng.)	1
④	3	*可 不 可以 改天 再 VP?* Can not can another day again VP? (Lit. Eng.) Can NP VP again on some other day? (Edited Eng.)	

If we are to select one type of core function utterances in the language task being studied, by common sense, getting a new appointment must be the first choice. However, NSs were not keen on using this kind of expressions. As

shown in Table IV.18 and IV.30, NNSs produced far more number of utterances only in this specific function type (compared to 10 utterances by 9 NSs, 18 NNSs produced 25 utterances, but only 4 are NL-T). Their average length is also amazingly close to NSs' (16.3 versus 16.7). Their approach is far more aggressive than NSs (see below). This abnormality deserves a closer investigation.

There are three reasonable inferences that can be drawn to help explain this abnormality:

a) The difference might be due to cultural differences. Under such an embarrassing situation after missing the test and the makeup test especially arranged for them, NSs, as cultural beings, preferred more low-key utterances such as those in 'Another chance', 'Vowing' and 'Winning sympathy' to show that they looked upon the professor's mercy or would comply with the professor's instructions than to initiatively ask or propose a new appointment which might lead to bad feeling or even confrontation. This can be supported by the scarcity of time words in NS utterances. There are only 2 specific time words for new appointments (one is 'now' and the other 'this afternoon') and 1 vague 'another time', while in NNS data, 5 specific time words and 4 general time words ('any time') are identified, plus 14 occurrences of 'another time' or 'another day'. A typical NS utterance is more indirect, with a Sentence Crown FS to show concessions, and in the form of a question, such as:

Exp. IV.42

如果 可以 的话，能 不 能 再 另外 安排 一 个 时间 让 我 参加 考试？

If can if, can not can again specially arrange one unit time let me participate test? (Lit. Eng.)

If possible, would you mind arranging another time for me to take the test? (Edited Eng.)

On the other hand, typical NNS utterances are more direct:

Exp. IV.43

(x)如果　*下　个　　星期四　　下午　的话, 我 有　 时间　考.*

If　　next unit　Thursday afternoon　if,　I　have　time　 take test.　(Lit. Eng.)

(x) If it's Thursday afternoon next week, I will have time to take the test.　　(Edited Eng.)

Exp. IV.44

(x) 所以 请　 你　 决定　　什么　　时候　　比较　　好.

So　　beg　you　decide　　what　　time　　comparative good. (Lit. Eng.)

(x) So please decide when is better.　　(Edited Eng.)

b) The difference might also be attributed to the perceived prestigious status by NNS students in China as foreigners from developed countries. This is not impossible because ordinary people in China (including Hong Kong) tend to treat foreigners from developed countries more nicely, and CSL teachers in China also tend to 'spoil' their foreign students like quests or customers. The researcher has personally heard many complaints from CSL teachers teaching overseas that when their students were sent to China, many of them studied less diligently, skipped more classes and did not make progress as expected, despite of far better language environment. The main problem seemed to be their teachers who were too polite, not strict enough and dared not teach their foreign student in the traditional way, even if they knew that learning of language and culture should not be detached. This might hold true in Hong Kong as well, as all CSL teachers involved in the program are from either from mainland China or Taiwan.

c) The difference can also be attributed to NNSs' insufficiency or deficiency in their linguistic repertoire while they sensed the need to say something along this vein because, by common sense, without getting a new appointment, the whole language task would not be successful. Although

they had learned a lot of patterns and lexis (as a matter of fact, all the patterns and lexis NSs used had been covered and practiced intensively in the first half of the 2-year program NNSs took), they did not have enough ready-to-use interactional routines in hand and had to create with the patterns and lexis they possessed, or to employ something semantically viable but pragmatically not very appropriate. In the face of a demanding communicative situation in which the demand exceeds the resources (Segalowitz 1997:105), probably some NNSs chose to say something they would not say if they had other alternatives. Actually, among the 26 NNS utterances, 10 were given by 8 English speakers and 16 by 10 Japanese and Korean. To the researcher's dismay, the quality of their output does not differ, even though the Japanese and Korean group is culturally very similar to Chinese, or even more traditional after Cultural Revolution in China. Among their 16 utterances, 7 are NL-T and 6 are NNL, with similar density of specific time words, general time words ('any time') and vague time words ('another time' or 'another day')[169].

IV.3.1.3.4 Analysis of 'Showing Eagerness/Readiness' Utterances

Number of utterances of this type is quite small, and NSs produced only 5, fewer than the 7 by NNSs. Quality of NNSs utterances is very low (6 out of 7 are problematic), like other specific types in the same broad type 'Requesting'. Table IV.34 captures all the TSSSs employed.

Table IV.34 Task-specific sentence stems (TSSS) in 'Showing eagerness/readiness'

	NNS freq.	TSSSs (with sample extensions)	NS freq.
①		*NP 能 不 能 再 VP?* NP can not can additionally VP? (Lit. Eng.) Can NP VP (in order to have another chance)? (Edited Eng.)	1
②	1	*NP (真的 (是)) 很/ 非常 想/ 希望 (再) VP.* NP (indeed (be)) very/very want/ hope (again) VP. (Lit. Eng.) NP really want to VP again. (Edited Eng.)	2
③		*NP 希望 能 有 机会 来 VP.* NP hope can have chance come VP. (Lit. Eng.) NP really hope NP can have another chance to VP. (Edited Eng.)	1
④		*NP 已经 准备 好 了.*	1

[169] Exp. VI.3.3 b and c were produced by a Japanese and a Korean participant.

		NP already prepare well PRT. (Lit. Eng.) NP is prepared well for it. (Edited Eng.)	

Despite of their lower occurrence in the data collected for this study, the following utterances produced by NSs are very typical Interactional FSs to express ones eagerness to retake a test or do something else. The first one is syntactically and lexically the simplest (all the components were included in the NNSs' elementary course material), while the third is the most difficult (most of the components were included in their elementary courses except for the verb-object collocation *补偿错误* in the end of utterances).

Exp. IV.45

我 真的 很 希望 很 希望 再 考 一 次.

I really very hope very hope again take test one time.(Lit. Eng.)

I really really hope that I can take the test one more time. (Edited. Eng.)

Exp. IV.46

我 真的 很 想 再 得到 这 次 机会.

I really very want again obtain this time chance.(Lit. Eng.)

I really hope that I can be given this chance. (Edited. Eng.)

Exp. IV.47

我 希望 能 有 其他 机会 来 补偿 这 次 错误.

I hope can have other chance come make up this time fault. (Lit. Eng.)

I hope that I can have another chance to make up for my fault. (Edited. Eng.)

Though functionally slightly different from the first three, the fourth form in Table IV.34 seems to be the most direct and highly learnable because of its simplicity in form and in the lexical items involved.

IV.3.1.3.5 Analysis of 'Vowing' Utterances

Utterances of this type are usually found towards the ending of discourses, typically after requesting for another chance or proposing a new schedule for retaking the makeup, to convince the professor that granting another chance

to the students would not be fruitless.

Number of utterances of this type is quite small but with normal ratio (6 by NNSs with 16 characters on the average, and 12 by NSs with 20.4 characters). Although the messages can be conveyed, quality of NNSs utterances is very low (all utterances are problematic[170]), like other specific types in the same broad type 'Requesting'.

Strategically, NNSs and NSs vowed with quite similar approaches, as displayed in Table IV.35. The TSSSs are presented in Table IV.36:

Table IV.35 Approaches in 'Vowing'

NNS freq.	Approaches	NS freq.
4	Will definitely not forget again	8
1	If forget again, will not ask for another chance	2
0	Will cultivate new habits to avoid the same fault in the future	1
1	Will work hard in the test to repay	1

Table IV.36 Task-specific sentence stems (TSSS) in 'Vowing'

	NNS freq.	TSSSs (with sample extensions)	NS freq.
①	0	*(NP) 一定不会 (再) VP 的 / 下次 (一定/肯定) 不会 (再) VP 了/ 下次一定会 VP 的.* (NP) definitely not will (again) VP PRT/next time (definitely/surely) not will (again) VP PRT/next time definitely will VP PRT (Lit. Eng.) NP will never VP again/next time NP will definitely not VP again/next time NP will definitely VP. (Edited Eng.)	8
②	0	*如果 NP1 这 次 再 忘掉 的话, 那 NP1 真的 是 就 该死了.* If NP1 this time again forget if, then NP1 really be then should die PRT. (Lit. Eng.) If NP1 forget again this time, NP1 really should kill himself. (Edited Eng.)	1
③	0	*NP 应该 养成 良好 的 习惯, 把 重要 的 事情 记录 下来.* NP should cultivate good PRT habit, PREP important PRT matter jot down. (Lit. Eng.) NP will jot down everything important from now on. (Edited Eng.)	1
④	0	*如果 NP1 能 再 给 NP2 机会, NP2 一定 会 好好儿 把握 的.* If NP1 can again give NP2 chance, NP2 definitely will well grasp PRT. (Lit. Eng.) If NP1 let NP2 retake the test, NP2 promises to do his very best. (Edited	1

[170] NNSs' utterances include 我一定记得住吧；如果我下次又忘了那就不用了；那个时候我可以真的我别忘；你可以一个机会给我我一定考考试得很好。

		Eng.)	

The VPs in TSSS ① can be *忘记* (= forget) and *有这种情况发生* (= have something like this happen), etc, when expressed negatively, or *想办法准时到达* (= try my best to arrive in time) and *赶得及来考试* (= arrive in time to take the test) when expressed positively, such as the following:

Exp. IV.48

下 一 次 的 补考 我 一定 不 会 忘记 的.

Next one time PRT makeup I definitely not will forget PRT.(Lit. Eng.)

I will definitely not forget about the makeup again next time. (Edited. Eng.)

The NNSs performed very poorly with their 6 utterances in this category seemingly because they did not master any whole phrases and relied on their grammatical competence. The simplest way of expressing 'Vowing' is 下一次我一定不会忘记的. Though it is of considerable length, its constituents are very basic items and are normally covered in elementary textbooks.

IV.3.1.3.6 Analysis of 'Winning Sympathy' Utterances

Table IV.18 shows that number of utterances of this type is not too small and with normal ratio (10 by NNSs with 17.8 characters on the average, and 21 by NSs with 21.2 characters). Quality of NNSs utterances is again very low (9 out of 10 utterances are problematic), like other specific types in the same broad type 'Requesting'.

Strategically, most NNSs and NSs tried to win the professor's sympathy with quite similar approaches, as displayed in Table IV.37, and the TSSSs are presented in Table IV.38:

Table IV.37 Approaches in 'Winning sympathy'

NNS freq.	Approaches	NS freq.
7	The test is too important not to retake.	20
3	The student is currently facing physical, mental or general problems.	0
0	The student can definitely pass the test, once given the chance.	1

In order to support their argument that the test was too important not to retake, 4 NNSs mentioned the importance of the test to their study, graduation, future job-hunting and even parents and while 4 NSs mentioned the same concerns except for parents.

Table IV.38 Task-specific sentence stems (TSSS) in 'Winning sympathy'

	NNS freq.	TSSSs (with sample extensions)	NS freq.
①	0	*这 个 NP1 (对 (于) NP2 (来说)) 非常/很 重要/意义重大/关键.* This unit NP1 (to NP2) very/very important/meaningful/critical.(Lit. Eng.) This NP1 is very important/meaningful/critical (to NP2). (Edited Eng.)	13
②		*这 个 NP1 (对 (于) NP2 (来说)) 太 重要 了.* This unit NP1 (to NP2) too important PRT. (Lit. Eng.) This NP1 is so important (to NP2). (Edited Eng.)	2
③		*没有 这 个 NP1, NP2 是 毕不了业 的.* Without this unit NP1, NP2 be cannot graduate PRT. (Lit. Eng.) Without this NP1 NP2 can't possibly graduate. (Edited Eng.)	1
④		*会 影响 NP 将来 的 学习 和 工作.* Will affect NP future PRT study and career. (Lit. Eng.) Will affect NP's study and work in the future. (Edited Eng.)	1
⑤		*NP 一定 会 考 过去 的.* NP definitely will take test pass PRT. (Lit. Eng.) NP definitely will pass the test (if given the chance). (Edited Eng.)	1
⑥	1	*哎呀, NP 快 要 烦 死 了.* Oh, NP soon will troubled die PRT. (Lit. Eng.) Oh, NP has a big problem and is deeply troubled. (Edited Eng.)	

Among the 6 TSSSs in Table.IV.38, the first two are very similar and are the most common ones that can be adapted to fit almost anyone's need, while the third and forth are more impactful because they mention about the consequences of not taking the test. The third can be reformulated as below, in which the VP represents the consequences:

Exp. IV.49

没有 NP 我 是 VP 的.

Without NP I be VP PRT. (Lit. Eng.)

The fifth is acceptable but the argument might not be applicable to everyone. It can be formulated to be of the same structure as one of the TSSSs in 'Vowing' (Table.IV.36 ①) showing determination. In this case, VP represents a highly likely result:

Exp. IV.50

我　一定　　会　VP　的.

I　definitely　will　VP　PRT.　　(Lit. Eng.)

The sixth is an interactional routine that can be used verbatim when complaining, grumbling or crying out for help when feel deeply troubled (in this case it is preceded by 'my parents are going to divorce'). It is the only error-free utterance by NNSs (an ethnic Chinese) in this section.

Statistically, 'winning sympathy' is also a core function for the task at issue. Syntactically, the shortest and simplest form is *这个考试对我来说非常重要*. All the elements are normally introduced, at the latest, in intermediate course books. The NNSs did not do it well seemingly because they did not master enough whole phrases (the only satisfactory utterance was produced by an ethnic Chinese). If tasks like the one being studied need to be introduced in early stage of a certain course,*这个考试对我来说非常重要* can be taught as a FS with minimal explicit grammar teaching (actually *NP 对NP 来说非常重要* is a very useful frame in many circumstances and is worthy of learning).

IV.3.1.3.7 Interim summary:

As in the above two broad function types, more extensions and more complex structures are observed in NS data. The performance of NNSs was quite unsatisfactory but most of the basic utterances needed (e.g. *请再给我一个机会, 麻烦老师了, 我已经准备好了, 下一次我一定不会忘记的, 这个考试对我来说非常重要*) are syntactically not too difficult. It can be argued that a

more lexical and task-based approach can help improve this.

IV.3.1.4 Analysis of Utterances in Broad Function 'Winning Favorable Impression'

Table IV.39 demonstrates that NNSs employed all kind of specific function type of utterances as NSs to win favorable impressions except for 'Showing modesty', which was only used once by NSs. They complimented the professor and showed appreciation of his help in the past or in the future, demonstrated their honesty, modesty and flexibility, and tried to be considerate when asking for another chance. Due to their limited occurrence, 'Complimenting', 'Showing modesty' and 'Showing submission' will not be analyzed here.

The quality of NNS utterances under this broad function type is the lowest of the 4 being scrutinized in this chapter, with 40% classified as NL-UT, and 40% as NNL, as shown at the bottom of the table. Quality of NS production is also the lowest of all.

Table IV.39 Specific Function Types in Broad Type (Winning favorable impression)

		NNS				NS			
Specific function types	Examples (might not be the whole utterance)	No. of NL-UT utter-ances	No. of NNL utter-ances	No. of NL-T utter-ances	No. of users	No. of NL-UT utter-ances	No. of NNL utter-ances	No. of NL-T utter-ances	No. of users
Winning favorable impression-complimenting	我真的非常非常的喜欢您教的这一门课. I really really love the course you teach. (Edited Eng.)		1		1			1	1
Winning favorable impression-expressing gratitude	我会感激您的. I will be very thankful to you. (Edited Eng.)	1	1		2			8	8
Winning favorable impression-showing consideration	我知道这给您添了很大的麻烦. I know I have caused a huge trouble to you. (Edited Eng.)	3	3	3	6	2		6	5
Winning favorable impression-showing honesty	我不想编任何的谎言. I don't want to tell any lies. (Edited Eng.)	2	1		3	1		6	5

Winning favorable impression-showing modesty	我想再多的解释都不是很有意义. I think I should not give more excuses. (Edited Eng.)					1		1	2
Winning favorable impression-showing submission	甚至是说一定要下一个学期再选修您的课，我都会绝对服从. Even if you ask me to retake your course next semester, I will obey without reservation. (Edited Eng.)		1		1			1	1
Total		6/ 40%	6/ 40%	3/ 20%		4/ 15.4%	0/ 0%	22/ 84.6%	

IV.3.1.4.1 Analysis of 'Expressing Gratitude' Utterances

Table IV.18 shows that NNSs only produced 2 utterances (average number of characters: 8.5) with low quality (Table IV.39) while NSs had 8 NL-T (average number of characters: 19.6). When looking at the time dimension, 7 out of 8 NS utterances expressed gratitude to what the professor had done before (see TSSS ① and ② in Table IV.40). One NS utterance and both NNS ones expressed appreciation in advance to what the professor might do to help in the future (TSSS ③).

Table IV.40 Task-specific sentence stems (TSSS) in 'Expressing gratitude'

	NNS freq.	TSSSs (with sample extensions)	NS freq.
①		*NP1 特地/特意 为 NP2 安排 了 补考.* NP1 specially for NP2 arrange PRT makeup. (Lit. Eng.) NP1 arranged a makeup specially for NP2. (Edited Eng.)	5
②		*NP1 (非常 好), 给 了 NP2 这/一 次 机会.* NP1 (very good), give PRT NP2 this/one time chance. (Lit. Eng.) (NP1 were so kind to) gave NP2 the second chance. (Edited Eng.)	2
③		*这个, NP1 会 很 感激 NP2 的.* This, NP1 will very appreciate NP2 PRT. (Lit. Eng.) NP1 will appreciate it very much if NP2 can do this. (Edited Eng.)	1

As in many other cases, NS TSSSs are escorted by many different peripheral phrases, such as the first 3 characters in the utterance below:

Exp. IV.51

还　　麻烦　　您　特意　　为　我　安排　了　补考.

Unreasonably　trouble　you　specially　for　me　arrange PRT makeup.(Lit. Eng.)

It was so kind of you to have scheduled a makeup test solely for me. (Edited Eng.)

Among TSSSs in 'expressing gratitude', the first two with highest frequencies seem to be syntactically and culturally challenging. The third, though low in frequency, is much easier. Intuitively *我会感激您的* can be used in many other speech events. As *感激* is normally not introduced in elementary stages, *我会感激您的*（or a even easier and commoner version *我会感谢您的*）can be introduced as a FS[171].

IV.3.1.4.2 Analysis of 'Showing Consideration' Utterances

Table IV.18 shows that number of NNS utterances is slightly more than NS (9 versus 8) but with shorter length (average number of characters: 12.9 versus 21.7). NNS quality is again very low while NS quality is also the lowest, with 2 out of 8 utterances being NL-UT (Table IV.39). Strategically, 5 out of 6 NNSs used 7 of their utterances to show their consideration by acknowledging the inconvenience caused to or would be caused to the professor. NSs adopted more diversified approaches, as shown in Table IV.41.

Table IV.41　　Approaches in 'Showing consideration'

NNS freq.	Approaches	NS freq.
7	Acknowledge inconvenience caused or would be caused later	3
	Acknowledge the bad feeling the professor might have	1
	Acknowledge the difficulties the professor might be facing	2
	Offer solutions to help the professor solve his problem	2
2	Emphasize what has been done to help ease professor's problem	

The TSSSs are presented in Table IV.42 (note that ① and ② correspond to the first and second approaches above. Both ③ and ④ correspond to the

[171] In CCL corpus, as of 22 June 2012, there are there are 14 tokens of *NP 会很感激 NP 的*, but 21 tokens of *NP 会感谢 NP 的*, indicating the wider usage of the latter.

third approach and no acceptable TSSSs are found for the fourth and fifth approaches):

Table IV.42 Task-specific sentence stems (TSSS) in 'Showing consideration'

	NNS freq.	TSSSs (with sample extensions)	NS freq.
①	2	*(NP1 知道 这) 给 NP2 添/造成/带来 (了) (很 大 的) 麻烦 (了).* (NP1 know this) to NP2 add/create/bring (PRT) (very big PRT) trouble (PRT). (Lit. Eng.) (NP1 knows) this has caused/will cause a lot of troubles to NP2. (Edited Eng.)	3
②		*可能 NP1 会 觉得 NP2 对 NP1 不 尊重/重视.* Maybe NP1 will feel NP2 to NP1 not respectful/regard highly. (Lit. Eng.) What NP2 has done might cause NP1 to think that NP2 doesn't respect NP1/doesn't care about NP1. (Edited Eng.)	1
③		*NP1 知道 可能 这个 比较 难 一点儿, 因为 NP1 已经 是 第二次 VP 了.* NP1 know maybe this comparatively difficult a bit, because NP1 already be second time VP PRT. (Lit. Eng.) NP1 understands that this might be quite difficult because it is already the second time NP1 VP. (Edited Eng.)	1
④		*不 知道 NP1 可 不 可以 接受. 但是 NP2 希望 NP1 可以接受啦.* Not know NP1 can not can accept. But NP2 hope NP1 can accept PRT. (Lit. Eng.) NP2 doesn't know if NP1 can accept this, but really hope they can. (Edited Eng.)	1

A point worth noting here is that all of the syntactic pattern and lexis in TSSS ①had been covered and intensively practiced in the elementary courses taken by NNSs, except for *添* and *造成* which were also introduced at intermediate and advanced level. In other words, theoretically the following utterance by a Japanese participant might have been produced by all NNSs with considerable ease, had they learnt more lexically in a task-based curriculum:

Exp. IV.52

真的 给 老师 添 麻烦 了.

Really to teacher add trouble PRT. (Lit. Eng.)

I know this is going to cause a lot of troubles to you. (Edited Eng.)

However, while 5 NNSs produced 6 utterances with the right core noun *麻烦*

(= trouble) , it collocates with the right verb(s) only in 2 utterances.

IV.3.1.4.3 Analysis of 'Showing Honesty' Utterances

Table IV.18 shows that NNSs produced only 3 utterances of this type while NSs had 7. NNS productions are much shorter (average number of characters: 11.7 versus 21). All NNS utterances are problematic (Table IV.39). Strategically, only two related approaches have been found, i.e. 'tell the truth' and 'don't want to lie'.

The TSSSs are presented below:

Table IV.43 Task-specific sentence stems (TSSS) in 'Showing honesty'

	NNS freq.	TSSSs (with sample extensions)	NS freq.
①		*NP1 (本来) 也 想 (过) 找/扯 一 个 理由/借口 跟 NP2 解释/ 让 NP2 再 给 NP1 一 个 机会.* NP1 (originally) also want (PRT) find one unit excuse with NP2 explain/ ask NP2 again give NP1 one unit chance. (Lit. Eng.) NP1 did thought about making up excuses (and ask NP2 for another chance). (Edited Eng.)	2
②		*NP 不 想 撒谎/编任何理由.* NP not want lie/make up any excuses. (Lit. Eng.) NP don't want to lie/make up any excuses. (Edited Eng.)	2
③		*NP1 想 NP1 也 应该 非常 真诚地 对 NP2.* NP1 think NP1 also should very sincerely treat NP2. (Lit. Eng.) NP1 think NP1 should be sincere to NP2. (Edited Eng.)	1
④		*NP1 还 是 想 跟 NP2 实话实说.* NP1 still be want with NP2 tell the truth. (Lit. Eng.) In the end NP1 decided to tell NP2 the truth. (Edited Eng.)	1

Lexically the NL-T utterances in this specific function type are comparatively difficult. The core verb *解释* and noun *理由* in TSSS ① were not taught until intermediate level, and TSSS ② to ④ contains core lexical items that are not covered until advanced level (such as *撒谎* and *坦诚*) or not covered

in the whole program (*实话实说*). Therefore, it is not surprising to see utterances by NNSs like the following which are grammatically and semantically correct but not acceptable by judges:

Exp. IV.53

(x)	我	应该	说	真	话.	
	I	should	speak	true	speech.	(Lit. Eng.)
(x)	Let me tell you the truth.					(Edited Eng. with pragmatic mistake)

However, phrases as simple as *我不想找借口* can suffice. As analyzed in IV.3.1.1.3 above, NNSs tend to use *借口* in the way they use 'excuse' in 'I have no excuses'. Hopefully this kind of mistakes can be reduced if *我不想找借口* is introduced as a whole with a task.

IV.3.1.4.4 Interim summary:

As in the previous three broad function types, more extensions and more complex structures are observed in NS data. The performance of NNSs was quite unsatisfactory but most of the basic utterances needed (e.g.*我会感激您的, 给老师添麻烦了, 我不想找借口*) are syntactically not too difficult. It seems sensible to argue that a more lexical and task-based approach can help improve this.

IV.3.2. Quantitative Summary of TSSSs

In the above sections, fewer TSSSs can be found in NNSs data in each Specific Function type. Besides, while in some types the distribution of different TSSSs is quite even (e.g. 'Winning favorable impression-Showing honesty'), in most other types certain TSSSs stand out far more frequent than others (e.g. 'Requesting-Another chance'), indicating that there are some core TSSSs that should be given priorities in CSL.

As summarized in Table IV.44, 16 types of utterances denoting different language functions are analyzed and for each type 3 to 15 TSSSs are found, mostly used by NSs (the last line shows that only 29 or 32% of the 90 TSSSs are used by NNSs, compared to 83 or 92% by NSs).

Table IV.44 TSSSs identified in each Broad and Specific Function Types of utterances

Broad function types	Specific Function types	NNS No. of utterance	NS No. of utterance	NNS No. of TSSSs	NS No. of TSSSs	NNS TSSSs Type	NS TSSSs Type	Total TSSSs Type
Apologizing	apologizing-asking for forgiveness	1	8	1	9	1	5	5
	apologizing-overt apologies	35	49	31	49	6	14	15
	apologizing-self-reproving	12	21	4	19	2	9	10
Excusing	excusing-exonerating	7	15	6	12	3	7	7
	excusing-forgot	24	20	16	19	3	2	4
	excusing-gloss over	1	12	0	10	0	3	3
	excusing-testifying	6	24	4	23	2	5	5
Requesting	requesting-another chance	22	46	10	42	3	6	7
	requesting-general	6	11	1	10	1	5	5
	requesting-new appointment	26	10	4	9	2	3	4
	requesting-showing eagerness	7	5	1	4	1	4	4
	requesting-vowing	6	12	0	10	0	4	4
	requesting-winning sympathy	10	21	1	17	1	5	6
Winning favorable impression	winning favorable impression-expressing gratitude	2	8	0	8	0	3	3
	winning favorable impression-showing consideration	9	8	2	6	1	4	4
	winning favorable impression-showing honesty	3	7	0	6	0	4	4
Total		177	277	81	253	26	83	90

Among the 177 NNS utterances analyzed, only 81 or 45.8% of them contain TSSSs (compared with 252 or 91% in NS utterances). This implies that a significantly higher percentage of NNS utterances did not used proper sentence stems. This helps to explain from another perspective why NNS data contain a high percentage of NL-UT and NNL.

When comparing NNSs' number of utterances and number of TSSSs, we can see that their performance are the best with 'Overt apologies', 'Exonerating', 'Forgot'

and ‘Testifying’. Quantitatively they produced 6 to 35 utterances and qualitatively 66.6% (4/6) to 88.6% (31/35) of the utterances are with proper sentence stems. On the other hand, they are the weakest with ‘Gloss-over’, ‘Expressing gratitude’ and ‘Showing honesty’, with small number of utterances and no decent sentence stems.

Altogether NNSs only contributed to 26 types of TSSSs, far less than NSs’ 83 types. As a whole 90 types of TSSSs are found, implying that 7 types that are only used by NNSs and 64 types only used be NSs.

As mentioned earlier, besides ‘Excusing-details’, 7 Specific Function types are not included in the above analysis due to their limited number of utterances. If all the TSSSs are identified, altogether there should be around 100 types of TSSSs. As analyzed above, among these TSSSs, some are essential and teachable but some are not. TSSSs identified need to be graded and introduced at different stages.

IV.3.2.1 Quantitative Results Pertaining to Hypothesis 4

Altogether NNSs and NSs participants produced 252 and 377 utterances respectively in this study, and they were categorized into four Broad function types as shown in Table IV.45 below (a summary of Table IV.44). Only 177 NNS and 277 NS utterances were subjected to analysis of TSSS (see Section VI.3 for details), and 81 TSSSs (of 26 distinctive forms) and 253 TSSSs (of 83 distinctive forms) were extracted from them respectively. The following points can be observed:

1. Compared with 253 TSSSs extracted from 277 NS utterances, 177 NNS utterances only contributed 81 TSSSs because many NNS utterances are native-like but untypical (NL-UT) for this particular task, as shown in Section IV.1 above.

2. While there are 83 distinctive forms of TSSSs in NS data, there are only 26 in NNS data. Among the 26 distinctive forms of TSSSs by NSSs, 19 overlap with NSs’, yielding to a total of 90 distinctive forms.

3. As NSs employed a much greater number and variety of TSSSs, Hypothesis 4 (Given the language task, NSs employed significantly more typical TSSSs than NNSs) can preliminarily be proved.

Table IV.45 TSSSs identified in each Broad Function Types of utterances

Broad function types	NNS No. of utterances	NS No. of utterances	NNS No. of TSSSs	NS No. of TSSSs	NNS TSSS Types	NS TSSS Types	Total TSSS Types
Apologizing	**48**	**78**	**36**	**77**	**9**	**28**	**30**
Excusing	**38**	**71**	**26**	**64**	**8**	**17**	**19**
Requesting	**77**	**105**	**17**	**92**	**8**	**27**	**30**
Winning favorable impression	**14**	**23**	**2**	**20**	**1**	**11**	**11**
Total	**177**	**277**	**81**	**253**	**26**	**83**	**90**

In Table IV.46 below, numbers of TSSSs extracted from each participant in NNS and NS groups are compared. One the average, NNSs provided 2.70 TSSSs, which is significantly less than NSs' 8.67 ($p < 0.001$).

Table IV.46: t-test result comparing average no. of TSSSs extracted from NNS and NS data

Gp 1	Gp 2	Avg1	Avg2	Std 1	Std 2	n1	n2	CL (%)	Z-value	T-Stat	Implication
no. of TSSSs Extracted from NNS	no. of TSSSs extracted from NS	2.70	8.67	2.20	3.58	30	30	99.9%	3.2368	7.6515	NS > NNS

However, as NNSs produced far less utterances, and as not all utterances are analyzed in this study, it makes more sense to compare average number of TSSS per utterance analyzed. As shown below, for each NNS and NS utterance analyzed, there are 0.45 and 0.90 TSSS. Density of TSSS in NS data is two times as much as in NNS data ($p < 0.001$). Hence it is rather safe to say that Hypothesis 4 (NSs employed significantly more typical TSSSs than NNSs) can be proved.

Table IV.47: t-test result comparing average no. of TSSSs extracted from NNS and NS data divided by number of utterances analyzed

Gp 1	Gp 2	Avg1	Avg2	Std 1	Std 2	n1	n2	CL (%)	Z-value	T-Stat	Implication
NNS no. of TSSSs divided by utterances analyzed	NS no. of TSSSs divided by utterances analyzed	0.45	0.90	0.28	0.14	30	30	99.9%	3.2368	7.7137	NS > NNS

IV.4 Summary of the Chapter

The analysis in this chapter yields the following findings:

1. Both NNS and NS data are highly formulaic. It seems that Chinese language is not exceptional in its heavy reliance on formulaic word strings.

2. Quantitatively density of FSs is found to correlate with oral proficiency.

3. Given the language task, within the same time limit, NNSs produced significantly fewer number of syllables/characters, employed significantly fewer non-TSSS FSs (i.e. Collocations, Frames and Polywords) and TSSSs (Task-Specific Sentence Stems) than NNSs[172].

4. Moreover, while density of non-TSSS FSs in NNS data is only slightly (though statistically significant) lower than NS (60.4% vs. 71.9%), number of TSSS in NNS data is far lower than in NS data (45% vs. 90%). It might logically be speculated that the language training the NNS participants received was not sufficiently task-oriented, because they were able to use single-word lexis and many small FSs but were much weaker in using FSs at utterance (or speech act) level.

[172] This echoes the findings in Taguchi (2007) and Forsberg and Fant (2010).

5. A broader variety of non-TSSS FSs identified: we managed to identify many new types of word strings that are highly likely to be formulaic but yet neglected by studies on Chinese FSs, e.g. *很早就* (Adverb + Adverb Collocation), *如何才能* (Adverb + Verb Collocation), *比较 ADJ 一点儿* (Adverbial Frame), *VP 的事情* (Noun Frame), *一大半* (Polyword Noun) etc.. These new FSs might help to facilitate the awareness-raising process and form-focused exercises in lexical approach or task-based learning (Willis and Willis 1996b).

6. A broader variety of TSSSs identified: Most NNSs completed the language task at issue minimally (i.e. able to communicate the message minimally but with lots of mistakes and non-native like expressions). Among the TSSSs found, some can enable learners to complete the task sufficiently, and some can even enable them to do it graceful or impressively. Yet many of these TSSSs are not syntactically complex. They were just not provided in the course the NNSs took, or provided but not in task-based mode.

All the above findings point to more lexical and task-based approaches in L2 teaching and learning.

Chapter V Discussions, Implications and Conclusions

In this chapter we will discuss the findings presented in the previous chapters based on the research questions and hypothesis and draw some implications based on the discussions.

V.1 Discussions:

V.1.1 Wide Usage of FSs:

Although the criteria used in this study need to be further tested and tighter criteria such as the ones proposed by Wray (2008) to identify morpheme equivalent units (MEUs; see Section II.6) might lower the percentages, it is confirmed in this study that both NSs and NNSs of Chinese rely on heavy 'doses' of formulaic word strings.

V.1.1a In their oral production lasting up to 55 seconds, an average of 25.07[173] and 66.1[174] FSs of different sizes and types are employed by NNSs and NSs respectively. This implies on the average almost one FS per two seconds by NNSs and more than one FS per second by NSs!

V.1.1b As much as 69.4% and 56.8% of NS and NNS data (in terms of number of characters) are inside FSs, implying that most of both NS and NNS oral productions are formulaic! These figures are far higher than those, for example, reported by Foster (2001) and Moon (1998) but significantly lower than those by Altenberg (1998) etc. The disparities might mainly be attributed to difference in identifying criteria.

Many of the FSs, esp. TSSSs, identified were not included in the teaching materials used by the NNS participants. As Chinese language also heavily

[173] Sum of 22.37 (non-TSSS FSs) and 2.7 (TSSSs).

[174] Sum of 57.43 (non-TSSS FSs) and 8.67 (TSSSs).

relies on FSs, this can be rectified when developing new teaching materials in the future, especially in TBL (Task-based Learning) in which TSSSs can typically be highlighted and utilized.

V.1.2 Correlation between degree of Formulaicity and Proficiency:

In the spoken data studied, degree of formulaicity (in terms of characters inside FSs as percentage of total characters) was found to correlate with quality of oral production (i.e. whether grammatically correct and pragmatically appropriate). NS data is more formulaic than NNS data, advanced NNS data is more formulaic than less advanced NNS data, and, in NNS data, native-like utterances are more formulaic than non-native-like utterances. These findings can provide preliminary evidence to show that the density of FSs might serve as an indicator of quality of language use or even language proficiency. This further attests to the significance of FSs in teaching and learning of CSL.

V.1.3 Broader Varieties of Non-TSSS FSs Identified:

Besides those FSs that have been mentioned by other researchers, as inspired by literatures on 'duanyu' and patterns in CSL textbook, we managed to identify many new types of word strings that are highly likely to be formulaic but yet neglected by studies on Chinese FSs, e.g. *改天再* (Noun + Adverb Collocation), *忙起来就* (Verb + Adverb Collocation), *整个 NP 都* (Adverbial Frame), *NP 把 NP 给错过了* (Preposition + Verb Frame), *听见 NP 响* (Verb + Verb Frame), *真的是* (Polyword Adverb) etc.. These new FSs, once confirmed as MEUs in future studies, can significantly broaden the scope of FSs in Chinese and open up new directions on related studies. They are a good source of input for the awareness-raising and form-focused exercises, especially in lexical approach or TBL.

V.1.4 All-round Varieties of TSSSs Targeting on a Task

Identified:

Research papers and books on Chinese Interactional FSs typically provide small number of examples, and those examples are normally not organized for teaching or learning purposes. In this study, over 600 utterances are found and categorized into 27 function types. 16 representative function types containing 358 utterances are analyzed in depth to generate 90 TSSSs (together with those not analyzed results in almost a hundred TSSSs). These TSSSs, though of varying degree of difficulty and might need to be taught at different stages, are organized around a task, i.e. can provide an overall picture on how a language task is tackled and can constitute remarkable resources for teaching and researching the language task being studied. These TSSSs constitute a good source of input for TBL.

V.1.5 Disparities between NNSs and NSs:

Disparities between NNSs and NSs are bound to exist. The interesting thing is that the disparity is the greatest with TSSSs. NNSs are capable of expressing most of what they want to express, except with low quality, and the main problem with quality seems not to be due to insufficient words or grammar rules, but insufficient FSs, esp. TSSSs. NNSs only possess less than one third of necessary TSSSs. This fact leads us to consider the urgent necessity of explicitly teaching TSSSs. Most necessary ingredients were covered in early stages of CSL studies, and actually most of the TSSSs are syntactically and lexically not complicated. However, the results show that knowing the parts does not equal knowing the wholes, and the assembling of the wholes is not completely rule-based. Although creativity should be encouraged, it is worth teaching TSSSs explicitly for specific language tasks because we cannot assume that they can be produced once the components are learnt.

On a related issue, unlike NNSs, low quality utterances constituting 12.6% of NS production are those classified as untypical or pragmatically inappropriate. On the one hand, this is quite understandable, because it is impossible for all

NSs to produce only pragmatically appropriate TSSSs under pressure, and plausibly because the NSs did the task not for a real test.

V.1.6 Number of TSSSs or Sentence Stems in a Language:

Sentence stems, i.e. sentence-level formulaic word strings, in English are estimated to be of hundreds of thousands in number (Pawley and Syder 1983). It seems reasonable to speculate that there might be similar number of TSSSs in Chinese, as number of TSSSs identified in current study alone (with only 30 NS and 30 NNS speech samples and one language task!) is already around a hundred. Although many of these TSSSs might be used in performing other language tasks as well, there can be hundreds, if not thousands, of such tasks in real life, ranging from similar ones (e.g. asking for a second chance from your boss after screwing up an important business in a company; asking for a second chance from your girlfriend after forgetting her birthday; asking for a second chance from a company after being late for a job interview, etc.) to those very different ones. Even with the same task, when the demographics of the speaker and listener (student and professor in this case) are different or when the power structure between them are different, for instance when it is between a 55-year-old postgraduate student and a 25-year-old tutor, or when the professor and the student are long-time friends since childhood, many more varieties of TSSSs will definitely emerge.

While theorists might be interested in speculating the total number of TSSSs in a language, the language teachers and teaching material compilers should be more interested in identifying the TSSSs learners need at different stages, or identifying the TSSSs needed for completing the language tasks in a task-based syllabus. The way spoken data was collected and TSSSs were extracted in this research should be able to shed some light in this regard.

V.1.7 Core and Peripheral TSSSs and How They can be Used:

For a given task, it is unlikely that all the TSSSs, or all types of TSSSs, identified are of equal importance in teaching and learning. In-depth analysis

in Section IV.3 shows that statistically, among the broad function types, ‘apologizing’ and ‘requesting’ are more crucial than ‘excusing’ and ‘winning favorable impression’. If we look at the specific function types, ‘overt apologies’, ‘asking for another chance’ and ‘testifying’ are the most salient. When we go deeper into each specific function type, some TSSSs stand out as commoner or easier for NNSs to learn, and are of higher value for pedagogical purposes. For example, in the ‘asking for forgiveness’ type, it seems that the following TSSSs are more recommendable than others, based on the findings of this research, and the first one is syntactically and semantically easier and can be taught in the first time when such speech act is in need:

请 (NP1) 原谅 (NP2).

Ask (NP*1*) forgive (NP*2*). (Lit. Eng.)

Please forgive NP*2*. (Edited Eng.)

NP1 希望 NP2 能(够) 原谅 NP1.

NP*1* hope NP*2* can forgive NP*1*. (Lit. Eng.)

NP*1* hope NP*2* can forgive NP*1*. (Edited Eng.)

On the other hand, some culturally and psycholinguistically challenging TSSSs such as ‘gloss –over’ might need to be avoided at elementary or even at advanced stages. In other words, while tasks need to be graded according to some criteria, TSSSs also need to be graded, and core ones and peripheral ones should be distinguished and categorized. Difficult ones can be introduced in later stages or only be used to train receptive skills.

Below is a list of the most common TSSSs of each specific function type identified in this study[175].

1. *请 (NP1)原谅 (NP2).* (Please forgive NP.)

[175] Arranged not in their order in the discourse, but in alphabetical order of their function type (from Apologizing-asking forgiveness to Winning favorable impression-showing honesty).

2. *(真 (的)) (很)对不起 (啊).* (So sorry !)
3. *(无论如何)是 NP 的错(误)/不对.* (It is all NP's fault)
4. *NP (真的 (是))没(有) 办法 VP.* (NP (really) could not find a way to VP)
5. *NP (完全/就/确实是/突然间)忘 (记)了(VP (的事情)).* (NP (actually/completely/suddenly) forgot about VP.)
6. *(NP) 不知道怎么/为什么/忙些什么 (就) (忽然)VP 了.* (Don't know how/why/what NP was busy with but just (entirely) forgot about this)
7. *(NP) (真的/确实)不是故意/有心/特意 VP 的.* (NP really did not miss the test deliberately)
8. *(NP1) (恳)请 NP2 (可不可以) (再)给 NP1 (一次/个) (VP1 的)机会 (让 NP1 VP1).*(NP1 would like to ask if NP2 can give NP1 one more chance to VP1)
9. *(那就)/(真的) (要)麻烦 NP 了.* (Please help me)
10. *能不能/请 NP1 (再) (给 NP2) (另外) 安排一个/次机会 (VP).* (Could NP1 /please arrange another chance for NP2 to VP)
11. *NP (真的 (是)) 很/非常想/希望 (再) VP.* (NP really want to VP again)
12. *(NP) 一定不会 (再) VP1 的/下次 (一定/肯定) 不会 (再) VP1 了/下次一定会 VP2 的.*(NP will never VP1 again/next time NP will definitely not VP1 again/next time NP will definitely VP2)
13. *这个 NP1 (对(于) NP2 (来说)) 非常/很 重要/关键/意义重大.* (This NP1 is very important/meaningful/critical (to NP2))
14. *NP1 特地/特意 为 NP2 安排 了 补考.* (NP1 arranged a makeup specially for NP2)
15. *(NP1 知道这)给 NP2 添/造成/带来 (了) (很大的) 麻烦.* ((NP1 knows) this has caused/will cause a lot of troubles to NP2)
16. *NP 不想 撒谎/编任何理由.* (NP don't want to lie/make up any excuses)

Obviously some TSSSs cannot be used together in one discourse (such as no. 4 and no. 5) because they are semantically mutually excluding. However, by choosing and rearranging/remixing a few of the above, and with necessary amendments, it should be very easy to construct a new discourse to complete the language task being studied. The bigger such repertoire of TSSSs or groups of TSSSs a learner possesses, theoretically the easier for him/her to tackle a new task.

If a language task is too challenging but necessary to be included in a syllabus, TSSSs identified can be graded and regrouped to make texts of different

difficulty to be used at different stages. In other words, similar tasks can be introduced quite a few times in a syllabus, with more complexity each time.

For example, the easiest monologue text based on the task at issue can be as short as containing three simple TSSSs as follows, to be introduced as early as at elementary stage:

1. *老师，很对不起*. (I am very sorry teacher!)
2. *是我的错*. (It is my fault.)
3. *请再给我一个机会*. (Please give me one more chance.)

The next one might fit intermediate level:

1. *老师，真的很对不起*. (I am so sorry teacher!)
2. *是我不对*. (It is my fault)
3. *我忘了考试*. (I forgot about the test)
4. *请您再给我一个机会吧*. (Please give me one more chance.)
5. *这次考试对我非常重要*. (The test is so important to me)
6. *麻烦老师了*. (Please help me, teacher)
7. *这次我一定不会忘记的*. (I will not forget again)
8. *麻烦老师了*. (Please help me, teacher)

The third one is an edited version of an authentic NS production[176], likely to be used at advanced level or above, at least for listening exercises:

1. *教授~不好意思*! (I am sorry, Professor!)
2. *今天早上因为家里出了点儿急事*. (It is because something urgent happened in my home this morning)
3. *我的外公突然生病了*, (My grandfather suddenly got very ill)
4. *然后我要送他到医院去*, (so I had to take him to hospital)
5. *所以忙起来就突然间忘记了补考的事情*. (and I suddenly forgot my makeup test while I was busy taking care of all these)

[176] This is a clean version of the speech sample quoted in III.4.1.1. Seven hesitant fillers and two false starts were removed as a result of streamlining, and a collocation error was corrected.

6. *请求您再给我一次机会补考*. (Please give me one more chance to do the makeup test)
7. *对你带来的不便希望您能原谅*. (I hope you can forgive me for any inconvenience caused)
8. *真是不好意思*! (I am really very sorry!)
9. *我向您保证下一次的补考我一定不会忘记的*. (I swear that I will not forget again in the next makeup test)
10. *请求您再给我一次机会吧*. (Please give me one more chance)

There are also some tasks that might require very different TSSSs at different stages, e.g. casual self-introduction (e.g. between classmates) at elementary level, semi-formal self-introduction (e.g. job interview) at intermediate level, and formal self-introduction (e.g. inauguration ceremony) at advanced level (Wu, 2012). In such cases, we might need different spoken data to extract their respective TSSSs.

V.1.8 The Use of NNS data:

In TBL, NS data is of primary significance as input, and NNS data can also be of considerable value. They can serve as input in designing help boxes containing common errors to warn learners in dictionaries like those described in Gillard and Gadsby (1998:164-170), in designing L1-sensitive teaching materials (Kaszubski 1998:184) and in compiling error-recognition exercises (Milton 1998:192). Like NS data used as a way of conducting form-focused instruction in data-driven learning (DDL) (Johns 1991a & 1991b), NNS data can be used to enhance form-focused instruction and data-driven learning to remedy overlooked issue of accuracy in communicative approaches (Granger and Tribble 1998:199), especially overuse and underuse (ibid. p205).

Besides the above, there may be one way NNS TSSSs can directly contribute to TBL. The acceptable TSSSs produced by NNSs should be a valuable source of input in compiling teaching materials. As they were correctly used by some NNSs, intuitively they might be easier to learn by other NNSs. Of course this needs further confirmation with more empirical research.

V.2 Implications

Several implications for CSL and related research can be drawn below:

V.2.1 FSs as core elements in CSL

As FSs are tangible and sensible units of languages, their significance in language teaching and learning should not be neglected. They should be at least of the same importance as grammar and vocabulary. Through studies like this, FSs of a certain language task performed by NSs can be identified, grouped and prioritized to be included in teaching materials, and NNS errors can be highlighted and analyzed to facilitate teacher training and learners' acquisition, and provide more meaningful and alternative input for inter-language studies.

V.2.2 A Lexical and Task-based Approach in CSL

Michael Lewis, the proponent of 'The Lexical Approach' asserts that his approach 'has less to say about innovative methods than might be expected. This is because it is explicitly an approach, not a syllabus or method. It advocates a total re-evaluation of the language which is offered to students, and how that language is analyzed' (1996:13). Richard and Rodgers (2001) also points out the following:

> The status of lexis in language teaching has been considerably enhanced by developments in lexical and linguistic theory, by work in corpus analysis, and by recognition of the role of multiword units in language teaching and communication. However, lexis still refers to only one component of communicative competence. Lewis and others have coined the term lexical approach to characterize their proposals for a lexis-based approach to language teaching. However, such proposals lack the full characterization of an approach or method ..... It remains to be convincingly demonstrated how a lexically based theory of language and language learning can be applied at the levels of design and procedure in language teaching, suggesting that it is still an idea in search of an approach and a methodology (p138).

As in the way spoken data was collected and FSs were identified in the study, there is a possibility to incorporate the Lexical Approach and TBL and develop a lexical and task-based CSL syllabus and approach. Before talking about designing a new syllabus, let us take a look at traditional CSL textbook compilation.

A typical lesson in a traditional textbook contains the following items: one or more texts in the form of dialogues, paragraphs and/or connected discourses, vocabulary list(s) with phrases or sentences as examples, grammar points with phrases or sentences as examples, and grammar-focused exercises. Most, if not all, of the texts and examples were created by intuition, and most of them were created to illustrate the grammar points, while topics, functions and tasks were secondary or even not taken into considerations. As a result, the language items covered in a lesson (words, phrases, sentences, paragraphs etc) were often diverged and difficult to be put together to complete a task.

With tasks and TSSSs, i.e. speech acts to complete the tasks, in mind, a syllabus can be designed to be lexical and task-based, with the following steps, for example (cf. pragmatic syllabus proposed by Wu (2008; 2012)):

1. Identify tasks learners need;
2. Collect NS data[177] (and some NNS data if possible) and extract the TSSSs (and other FSs, if needed);
3. Evaluate the difficulties of the language items, i.e. TSSSs, other FSs, vocabulary and syntactic structure[178] etc, involved in each task;
4. Sequence the task according to their level of difficulties and urgencies in use.
5. Group certain number of tasks that can share similar language items to

[177] Although it might be unrealistic and discouraging to use a NS as a model in learning a L2 (Byram 1997:11-12), speech samples produced by NSs on tasks that learners need to complete should be well justified to be used as major inputs.

[178] A lexical and task-based approach does not deny the importance of grammar teaching (Lewis 1997a:14; Willis and Willis 1996a; Skehan 1996, Hayashi 1995). Wray (2008) does notice the limitations of a strictly formulaic way of learning in adulthood: narrow topic and difficulties in extrapolating to new situations (p229). It seems that a blend of holistic and analytic learning is more desirable, especially in a long language program lasting many semesters like the one NNS participants in this study took.

form units of study in which the language items can be reused for several times in different but related tasks[179].

6. TSSSs can be regrouped to make dialogues, paragraphs and/or connected discourses to serve as texts in teaching materials. TSSSs can also be used as examples in vocabulary lists and grammar points.

With the above steps, textbooks making use of TSSSs extracted from task-based authentic data can be created to facilitate a lexical and task-based syllabus (also see V.1.7 for more details).

V.2.3 Language strategies in Lexical and Task-based Approach

Language strategies need to be investigated and taught explicitly. For instance, for 'Apology' and 'Pleading', it seems that repeated use of some varieties of short TSSSs, e.g. *对不起了对不起了* is highly recommendable. However, in 'Asking for forgiveness', one go of TSSS seems enough. When it comes to culture sensitive speech acts like 'making new appointment' in the task at issue, it seems that NNS learners needed to be reminded that only very tactful TSSSs can be used, and cultural differences should be borne in mind, and they might be advised not to use any in real life situations, as vast majority of NS, unless they are very proficient.

V.2.4 Memorization in Lexical and Task-based Approach

Languages are, to a large extend, formulaic in nature. While it is of utmost importance that we encourage learners to create with the language, it is of similar importance that we remind them of the formulaic nature of language. Learners need to acquire generative rules, but at the same time, also need to note that a huge number of language units of various sizes (FSs in this study) are to be acquired and used as wholes rather than composed with rules. Instead of creating with words and rules of target languages, they may be encouraged to create with FSs as described by Wong-Fillmore (1976:603) and,

[179] Gatbonton and Segalowitz (1988; 2005) recommend task-based activities in which useful utterances can be repeatly and meaningfully used.

recourse to words and rules only when necessary, like NSs do (Skehan 1996:21-22). But how? Do learners have to memorize the FSs?

While depth of processing and meaningful communication etc are definitely core elements in TBL[180], the importance of memorization cannot be neglected, as Lewis (1996:11) puts it:

> 'All these factors suggest a vastly greater role for memory in language learning, and a greatly reduced role for (implicit or explicit) understanding of 'grammar rules', however that term is understood'

Peters (1983) also points out that 'There is a pedagogical bias against the idea of rote memorization of long chunks of speech', especially in cultures in which imitation is looked down upon (p109), and 'though many of the objections to memorization and pattern practice are valid ……memorization and pronunciation practice of long chunks do at least allow the learner to concentrate on fluent phonological production of relatively lone pieces in a situation where other aspect of the processing lead have been minimized' (p110). Wray and Fitzpatrick (2008)'s empirical study report that memorized sentences in anticipated conversations gave the learners opportunity to sound nativelike and 'promoted their fluency, reduced the panic of on-line production in stressful encounters, gave them a sense of confidence about being understood, and provided materials that could be used in other contexts too' (p143). Ellis and Sinclair also view memorization of FSs as central to successful learning (Ellis and Sinclair 1996:246-7; Ellis 1996:91)[181].

Yorio (1980) argues that in order to have communicative competence, learners must be able to use the language grammatically, appropriately and effectively

---

[180] As Richard and Rodgers (2001:228) points out, TBL 'shares the general assumptions about the nature of language learning underlying Communicative Language Teaching.'

[181] In a culture where rote memorization is not looked down upon, Dai and Ding (2010:83-4) found that Chinese learners of English, esp. the low achieving ones, make faster and greater progress in English proficiency and writing ability, and make significantly greater progress in the accuracy and variation of FSs, through practicing text memorization.

(p433). For L2 learners, especially adults who do not have native speaking acquaintance to practice with, memorizing (not necessarily in mechanical ways[182]) FSs and refraining from excessive analysis might be an effective way to enhance effectiveness in communication. After all, if FSs are just longer words, intuitively there should be nothing wrong with memorizing them, as we memorize single-lexi words.

V.3 Limitations

Several areas might potentially limit the generalizability of the findings of this research:

1. Even if most, if not all, of the FSs in the data are identified, as there is only one spoken language task, the results only represent a very special sector of language in use rather than a representative sample. Many FSs found are obviously very task-specific and colloquial. When looking at the FSs classified under each categories and subcategories, it is also as clear as crystal that they are far from being all-encompassing. The results are only indicative.

2. In spite of the effort to identify exhaustively, as some FSs are discontinuous, and as FSs can be multiply stored, it cannot be guaranteed that all FSs are singled out, though NS and NNS data are scrutinized in exactly the same way.

3. As participants were asked to speak as much as possible within the time allowed, NSs might be quantitatively or even qualitatively advantaged because of their obvious higher proficiency in making up something to fill up the time, after the message was sufficiently conveyed. On the other hand, as there are 4 ethnic Chinese among the advanced NNSs, the disparities between NNSs and NSs might not be sufficiently manifested in this research.

[182] After reviewing eight studies on memorization for successful learning, Wray and Fitzpatrick (2008) observe that 'contrary to popular perceptions in the west, memorization does not need to be a superficial and therefore rather pointless activity' (p125). Cooper (2004) also concludes that 'while surface approaches to learning can be associated with mechanical rote learning…… memorization through repetition can be used to deepen understanding and achieve high levels of academic performance' (p289).

4. As there are only 12 NNSs participants in advanced group and 18 in intermediate group, the findings concerning their difference should be interpreted with care. On a related issue, as there are too few NNSs participants in each group, i.e. native English, Japanese and Korean speakers, NNS output was analyzed and described as a whole and some language-specific features might be blurred. This might further impair the generalizability of some of the findings in this study.

5. It is debatable whether all FSs identified in this research are truly formulaic items, i.e. MEUs in Wray (2008), because further scrutiny is not performed, due to the limitations of time and resources. This research can only serve as a preliminary trial in the empirical study of Chinese FSs.

6. Data in this study were collected with computers. After the recording, two Japanese participants did express their dissatisfaction with their performance as they were not used to talking to a computer. Whether this had any impact on the findings and generalizability of this research is unclear.

V.4 Conclusions

90 distinct forms of TSSSs and 984 distinct forms of non-TSSS FSs were identified from a transcript of NNS and NS task-based oral production containing 9268 Chinese characters. Both NNS and NS data were found to be highly formulaic, and density of FSs was found to be positively correlated with level of proficiency in the data. An in-depth analysis of non-TSSS FSs found many formulaic word strings which have been largely neglected in past research and textbook compilation, as a result of excessive reliance on intuition and written data. Another in-depth analysis of TSSSs revealed the severe deficiency in NNSs' mastery of Chinese speech acts in carrying out the language task at issue. The analysis also found that many of the necessary speech acts NNSs did not perform well are syntactically and lexically undemanding, with most of the constituents having been learnt. NNSs' poor performance was partly attributed to the fact that many of the speech acts were simply not included in the teaching materials and unlikely to be

supplemented in classroom activities which were centred around grammar points rather than tasks. In view of the significance of FSs, in particular TSSSs, in proficiency-based CSL, it is suggested that L2 teaching and learning, esp. syllabus design and textbook compilation, should be more lexical and task-based, and when speaking and listening are concerned, based on corpuses of spoken language.

Bibliography

Anonymous. (2006). *Collins COBUILD advanced learner's English dictionary* Glasgow: HarperCollins.

Anonymous. (2004). *The Standard Dictionary of Contemporary Chinese* Beijing, PRC: Foreign Language Teaching and Research Press.

Anonymous. (2002). *The contemporary Chinese Dictionary* Beijing, PRC: Foreign Language Teaching and Research Press.

Anonymous. (2000). *Yingyong Hanyu Cidian* Beijing, PRC: Commercial Press.

Adolphs, S. (2008). *Corpus and context: investigating pragmatic functions in spoken discourse* Amsterdam; Philadelphia: John Benjamins Pub. Co.

Adolphs, S. & Durow, V. (2004). Social-cultural integration and the development of formulaic sequences. In N. Schmitt (Ed.) *Formulaic Sequences: Acquisition, processing and use* (pp. 107-126). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Aijmer, K. (1996). *Conversational routines in English:convention and creativity* London; New York: Longman.

Aijmer, K. & Altenberg, B. (1991). Introduction. In K. Aijmer, B. Altenberg & J. Svartvik (Eds.) *English corpus linguistics :studies in honour of Jan Svartvik* (pp. 1-7). London; New York: Longman.

Altenberg, B. & Eeg-Olofsson, M. (1990). Phraseology in spoken English: Presentation of a project. In J. Aarts & W. Meijs (Eds.) *Theory and Practice in Corpus Linguistics* (pp. 1-26). Amsterdam and Atlanta: Rodopi.

Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combinations. In A. Cowie (Ed.) *Phraseology: theory, analysis and applications* (pp. 101-122). Oxford: Clarendon Press.

American Council on the Teaching of Foreign Languages (2012). *The ACTFL Proficiency Guidelines 2012—Speaking*. http://actflproficiencyguidelines2012.org/speaking

Arnaud, P. J. L. & Savignon, S. J. (1997). Rare word, complex lexical units and the advanced learner. In J. Coady & T. Huckin (Eds.) *Second Language Vocabulary Acquisition* (pp. 157-173). Cambridge, UK: Cambridge University Press.

Austin, J. L. & Urmson, J. O. (1962). *How to do things with words: the William James lectures delivered at Harvard University in 1955* Oxford: Clarendon Press.

Bardovi-Harlig, K. & Hartford, B. S. (1992). Refining the DCT: Comparing Open Questionnaires and Dialogue Completion Tasks. Pragmatics and Language Learning (Volume 4) Selected papers presented at the Annual Meeting of the International Conference on Pragmatics and Language Learning (April 2-4, 1992).

Benton, A. L. & Joynt, R. J. (1960). Early descriptions of aphasia. *Archives of Neurology, 3*, 109-126-205-222.

Biber, D. (1999). *Longman grammar of spoken and written English* Harlow, Essex: Longman.

Bloomfield, L. (1933). *Language* New York: Holt, Rinehart and Winston.

Boers, F. & Lindstromberg, S. (2008). How cognitive linguistics can foster effective vocabulary teaching. In F. Boers & S. Lindstromberg (Eds.) *Cognitive Linguistic Approaches to Teaching Vocabulary and Phraseology* (pp. 1-61). Berlin; New York: Mouton de Gruyter.

Bolander, M. (1989). Prefabs, patterns and rules in interaction? formulaic speech in adult learners' L2 Swedish. In K. Hyltenstam & L. K. Obler (Eds.) *Bilingualism across the lifespan : aspects of acquisition, maturity, and loss* (pp. 73-86). Cambridge ; New York: Cambridge University Press.

Bolinger, D. (1976). Meaning and memory. *Forum Linguisticum, 1*, 1-14.

Bradley, J. (2003). *Formulaic language in learner discourse: How study abroad affects oral production*. Unpublished PhD thesis, The University of Tennessee.

Brazil, D. (1995). *A grammar of speech* Oxford; New York: Oxford University Press.

Brown, R. W. (1973). *A first language; the early stages* Cambridge, Mass.: Harvard University Press.

Brumfit, C. & Johnson, K. (1979). *The communicative approach to language teaching* Oxford: Oxford University Press.

Byram, M. (1997). *Teaching and assessing intercultural communicative competence* Clevedon, England; Philadelphia: Multilingual Matters.

Candlin, C. N. (1987). Towards task-based language learning. In C. N. Candlin & D. Murphy (Eds.) *Language learning tasks* Englewood Cliffs, N.J.: Prentice Hall International.

Carter, R. & McCarthy, M. (1997). *Exploring spoken English* Cambridge: Cambridge University Press.

Chang, Y. (1993). *Kouyu Xiyongyu Gongneng Cidian (口语习用语功能词典)* Beijing, PRC: Beijing Yuyan Xueyuan Chubanshe.

Chanier, T., Colmerauer, C., Fouqueré, C., Abeillé, A., Picard, F. & Zock, M. (1993). Modélisation informatique de l'acquisition des expressions lexicales en langues secondes. *Communication présentée au 9ème colloque international sur l'acquisition des langues*. St.-Etienne. 13-15 mai 1993.

Channell, J. (1994). *Vague language* Oxford: Oxford University Press.

Chen, L. (2010). An investigation of lexical bundles in ESP textbooks and electrical engineering introductory textbooks. In D. Wood (Ed.) *Perspectives on formulaic language: acquisition and communication* (pp. 107-125). London ; New York: Continuum.

Clark, R. (1974). Performing Without Competence. *Journal of Child Language, 1(1)*, 1-10.

Coady, J. (1997). L2 vocabulary acquisition: A synthesis of the research. In J. Coady & T. Huckin (Eds.) *Second Language Vocabulary Acquisition* (pp. 273-290). USA: Cambridge University Press.

Cong, S. (2010). *Chinese Chunk and Their Implications in Learners Second Language Acquisition (汉语语块及其在留学生第二语言习得中的应用).* Unpublished MA dissertation, Northeast Normal University. Changchun, PRC.

Conklin, K. & Schmitt, N. (2008). Formulaic Sequences: Are They Processed More Quickly than Nonformulaic Language by Native and Nonnative Speakers? *Applied Linguistics, 29(1)*, 72-89.

Cooper, B. J. (2004). The enigma of the Chinese learner. *Accounting Education, 13*, 289-310.

Coulmas, F. (1981). Introduction: Conversational routine. In F. Coulmas (Ed.) *Conversational routine: Explorations in Standardized Communication Situations and Prepatterned Speech* (pp. 1-17). The Hague; New York: Mouton.

Cowie, A. P. & Howarth, P. (1996). Phraseological competence and written proficiency. In G. M. Blue & R. Mitchell (Eds.) *Language and education : papers from the annual meeting of the British Association for Applied Linguistics held at the University of Southampton, September 1995* (pp. 80-93). England; Philadelphia: Clevedon: British Association for Applied Linguistics in association with Multilingual Matters.

Cowie, A. P. (1998). Phraseological dictionaries: Some east-west comparisons. In A. P. Cowie (Ed.) *Phraseology: Theory, Analysis, and Applications* (pp. 209-227). Oxford: Clarendon Press.

Cowie, A. P. (1988). Stable and creative aspects of vocabulary use. In R. Carter & M. McCarthy (Eds.) *Vocabulary and Language Teaching* (pp. 126-139). London and New York: Longman.

Crick, F. H. C. (1979). Thinking about the brain. *Scientific American, 9*, 219-232.

Cruttenden, A. (1981). Item-learning and system-learning. *Journal of Psycholinguistic Research, 10*, 79-88.

Dai, Z. & Ding, Y. (2010). Effectiveness of text memorization in EFL learning of chinese students. In D. Wood (Ed.) *Perspectives on formulaic language: acquisition and communication* (pp. 71-87). London ; New York: Continuum.

De Cock, S., Granger, S., Leech, G. & McEnery, T. (1998). An automated approach to the phrasicon of EFL learners. In S. Granger(Ed.) *Learner English on computer* (pp. 67-79). London; New York: Longman.

de Saussure, F., Bally, C., Riedlinger, A. & Sechehaye, A. (1966; 1959). *Course in general linguistics* New York: McGraw-Hill.

Ding, J. (2006). *Exploration of SL Acquisition of Oral Chinese Idioms by Foreign Students (留学生汉语口语习用语块习得研究).* Unpublished MA dissertation, Jinan University, Guangzhou, PRC.

Ding, S. (1979). *Xiandai Hanyu Yufa Jianghua (现代汉语语法讲话)* Beijing, PRC: Commercial Press.

Doi, T. (1981). *The anatomy of dependence* Tokyo; New York: Kodansha International.

Dörnyei, Z., Durow, V. & Zahran, K. (2004). Individual differences and their effects on formulaic sequences acquisition. In N. Schmitt (Ed.) *Formulaic Sequences: Acquisition, processing and use* (pp. 87-106). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Ellis, N. C. (2008). Phraseology - the periphery and the heart of language. In S. Granger & F. Meunier (Eds.) *Phraseology in foreign language learning and teaching* (pp. 1-13). Amsterdam; Philadelphia: John Benjamins Pub. Co.

Ellis, N. C. (1996). Sequencing in SLA: phonological memory, chunking and points of order. *Studies in Second Language Acquisition, 18(91)*, 126.

Ellis, N. C. & Sinclair, S. G. (1996). Working memory in the acquisition of vocabulary and syntax: putting language in good order. *Quarterly Journal of Experimental Psychology, 49A(a)*, 234-250.

Ellis, R. (1994). *The study of second language acquisition* Oxford: Oxford University Press.

Erman, B. & Warren, B. (2000). The idiom principle and the open choice principle. *Text, 20*, 29-62.

Fan, X. (1996). *San ge Pingmian de Yufaguan (三个平面的语法观)* Beijing, PRC: Beijing Yuyan Xueyuan Chubanshe.

Ferguson, C. (1981). The structure and use of politeness formulas. In F. Coulmas (Ed.) *Conversational routine: Explorations in Standardized Communication Situations and Prepatterned Speech* (pp. 21-35). The Hague; New York: Mouton.

Firth, J. R. (2004; 1972). J. S. La Fontaine(Ed.) *The interpretation of ritual: essays in honour of A.I. Richards* (pp. 1-38). London: Routledge.

Firth, J. R. (1964). *The tongues of men; and Speech* London: Oxford University.

Fitzpatrick, T. & Wray, A. (2006). Breaking up is not so Hard to do: Individual Differences in L2 Memorization. *The Canadian Modern Language Review/La Revue Canadienne Des Langues Vivantes, 63(1)*, 35-57.

Forsberg, F. & Fant, L. (2010). Idiomatically speaking: Effects of task variation on formulaic language in highly proficient users of L2 French and Spanish. In D. Wood (Ed.) *Perspectives on formulaic language: acquisition and communication* (pp. 47-70). London ; New York: Continuum.

Foster, P. (2001). Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehen & M. Swain (Eds.) *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing* (pp. 75-93). Harlow: Longman.

Francis, W. N. (1992). Language copora B.C. In J. Svartvik(Ed.) *Directions in corpus linguistics: proceedings of Nobel Symposium 82* (pp. 17-32). Berlin: Mouton de Gruyter.

Gatbonton, E. & Segalowitz, N. (2005). Rethinking Communicative Language Teaching: A Focus on Access to Fluency. *Canadian Modern Language Review, 61(3)*, 325-353.

Gatbonton, E. & Segalowitz, N. (1988). Creative automatization: Pinciples for promoting fluency within a communicative framework. *TESOL Quarterly, 88*, 473-492.

Gerbig, A. & Shek, A. (2007). The phraseology of tourism: A central lexical field and its cultural construction. In P. Skandera (Ed.) *Phraseology and culture in English* (pp. 303-322). Berlin ; New York: Mouton de Gruyter.

Gibbs, R. W. J. (2007). Idioms and formulaic language. In D. Geeraerts & H. Cuyckens (Eds.) *The Oxford handbook of cognitive linguistics* (pp. 697-725). Oxford; New York: Oxford University Press.

Gibbs, R. W. (1994). *The poetics of mind: figurative thought, language, and understanding* Cambridge England; New York, NY, USA: Cambridge University Press.

Gillard, P. & Gadsby, A. (1998). Using a learners' corpus in compiling ELT dictionaries. In S. Granger (Ed.) *Learner English on computer* (pp. 159-171). London; New York: Longman.

Gläser, R. (1988). The grading of idiomaticity as a presupposition for a taxonomy of idioms. In W. Hullen & R. Schulze (Eds.) *Understanding the lexicon : meaning, sense, and world knowledge in lexical semantics* (pp. 264-279). Tubingen: M. Niemeyer.

Gleason, J. B. & Weintraub, S. (1976). The acquisition of routines in child language. *Language in Society 5*, 129-136.

Goffman, E. (1971). *Relations in public; microstudies of the public order* New York: Basic Books.
Granger, S. & Tribble, C. (1998). Learner corpus data in the foreign language classroom: Form-focused instruction and data-driven learning. In S. Granger(Ed.) *Learner English on computer* (pp. 199-209). London; New York: Longman.
Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. Cowie (Ed.) *Phraseology: theory, analysis and applications* (pp. 145-160). Oxford: Clarendon Press.
Gries, S. T. (2008). Phraseology and linguistic theory. In S. Granger & F. Meunier (Eds.) *Phraseology: An interdisciplinary perspective* (pp. 3-25). Amsterdam/Philadelphia: John Benjamins Publishing Company.
Hakuta, K. (1976). A case study of a Japanese child learning English as a second language. *Language Learning, 26(2)*, 321-351.
Halliday, M. A. K. (1991). Corpus studies and probabilistic grammar. In K. Aijmer, B. Altenberg & J. Svartvik (Eds.) *English corpus linguistics :studies in honour of Jan Svartvik* (pp. 30-43). London; New York: Longman.
Hayashi, K. (1995). Form-focused instruction and second language proficiency. *RELC Journal, 26(1)*, 95-115.
Hickey, T. (1993). Identifying formulas in first language acquisition. *Journal of Child Language, 20*, 27-41.
Howarth, P. (1998). The phraseology of learners' academic writing. In A. P. Cowie (Ed.) *Phraseology: Theory, Analysis, and Applications* (pp. 161-186). Oxford: Clarendon Press.
Hu, Y. (1992). *Xiandai Hanyu (现代汉语)*. Hong Kong: Joint Publishing.
Hudson, J. (1998). *Perspectives on fixedness: applied and theoretical*. Lund: Lund University Press.
Hunston, S. (2003). Lexis, wordforms and complementation pattern. *Functions of Language, 10(1)*, 31-59.
Hymes, D. H. (1979). On communicative competence. In Brumfit, C. & Johnson, K. (Eds.) *The Communicative Approach to Language Teaching* (pp. 5-26). Oxford: Oxford University Press.
Hymes, D. H. (1962). The ethnography of speaking. In Anonymous *Anthropology and human behavior* (pp. 13-53). Washington D. C.: The Anthropological Society of Washington.
Irujo, S. (1993). Steering clear: avoidance in the production of idioms. *International Review of Applied Linguistics in Language Teaching, 31(3)*, 205-219.
Jackendoff, R. (1995). The boundaries of the lexicon. In M. Everaert, E. van der Linden, A. Schenk & R. Schreuder (Eds.) *Idioms: Structural and Psychological Perspectives* (pp. 133-166). Hillsdale NJ: Erlbaum.
Jamieson, J. C. & Lin, T. (2002). *Kungfu (I): An Elementary Chinese Text* Hong Kong: The Chinese University Press.
Jespersen, O. (1924). *The philosophy of grammar, by Otto Jespersen* London: G. Allen & Unwin, ltd.
Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning. In T. Johns & P. King (Eds.) *Classroom concordancing* (pp. 1-16). Birmingham, England: Centre for English Language Studies, University of Birmingham.
Johns, T. (1991). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In T. Johns & P. King (Eds.) *Classroom concordancing* (pp. 27-45). Birmingham, England: Centre for English Language Studies, University of Birmingham.
Kang, Y. & Lai, S. (2005). *Conversational Chinese 301* Beijing, PRC: Beijing Language and Culture University Press.
Kaszubski, P. (1998). Enhancing a writing textbook: A national perspective. In S. Granger (Ed.) *Learner English on computer* (pp. 172-185). London; New York: Longman.
Kecskés, I. (2003). *Situation-bound utterances in L1 and L2* Berlin; New York: Mouton de Gruyter.
Keller, E. (1981). Gambits: Conversational strategy signals. In F. Coulmas (Ed.) *Conversational Routine: Explorations in Standardized Communication Situations and Prepatterned Speech* (pp. 93-113). The Hague; New York: Mouton Publishers.
Kjellmer, G. (1991). A mint of phrases. In K. Aijmer, B. Altenberg & J. Svartvik (Eds.) *English corpus linguistics :studies in honour of Jan Svartvik* (pp. 111-127). London; New York: Longman.

Koprowski, M. (2005). Investigating the usefulness of lexical phrases in contemporary coursebooks. *ELT Journal, 59(4)*, 322-331.
Krashen, S. D. & Scarcella, R. (1978). On routines and patterns in language acquisition and performance. *Language Learning, 28(2)*, 283-300.
Kuiper, K. (2004). Formulaic sequences in conventionalised varieties of speech. In N. Schmitt (Ed.) *Formulaic Sequences: Acquisition, processing and use* (pp. 37-54). Amsterdam/Philadelphia: John Benjamins Publishing Company.
Kuiper, K. (1996). *Smooth talkers: the linguistic performance of auctioneers and sportscasters* Mahwah, N.J.: L. Erlbaum Associates.
Kuiper, K. & Austin, P. (1990). They are off and racing now: The speech of the new zealand race caller. In A. Bell & J. Holmes(Eds.) *New Zealand ways of speaking English* (pp. 195-220). Clevedon [England] ; Philadelphia : Multilingual Matters.
Kuusi, A. L. (1974). An approach to the categorisation of phrases. *Proverbium, 23*, 895-904.
Leech, G. (2000). Grammars of Spoken English: New Outcomes of Corpus-Oriented Research. *Language Learning 50(4)*, 675-724.
Leech, G. (1998). Preface. In S. Granger (Ed.) *Learner English on computer* (pp. xiv-xx). London; New York: Longman.
Leech, G. (1991). The state of the art in corpus linguistics. In K. Aijmer, B. Altenberg & J. Svartvik (Eds.) *English corpus linguistics :studies in honour of Jan Svartvik* (pp. 8-29). London; New York: Longman.
Levinson, S. C. (2000). *Presumptive meanings :the theory of generalized conversational implicature* Cambridge, Mass.: MIT Press.
Lewis, M. (2000). Learning in the lexical approach. In M. Lewis (Ed.) *Teaching Collocation: Further Developments in the Lexical Approach* (pp. 155-185). England: Language Teaching Publications.
Lewis, M. (1997). *Implementing the lexical approach: putting theory into practice* Hove, England: Language Teaching Publications.
Lewis, M. (1997). Pedagogical implications of the lexical approach. In J. Coady & T. Huckin (Eds.) *Second Language Vocabulary Acquisition* (pp. 255-270). Cambridge, UK: Cambridge University Press.
Lewis, M. (1996). Implications of a lexical view of language. In D. Willis & J. Willis (Eds.) *Challenge and change in language teaching* (pp. 16). Oxford: Heinemann.
Lewis, M. (1993). *The lexical approach: the state of ELT and a way forward* Hove, England: Language Teaching Publications.
Li, M. (1995). *Intermediate Chinese Listening and Speaking* Beijing PRC: Beijing Language and Culture University Press.
Little, D. (1994). Words and their properties: Arguments for a lexical approach to pedagogical grammar. In T. Odlin (Ed.) *Perspectives on pedagogical grammar* (pp. 99-122). Cambridge ; New York, NY, USA: Cambridge University Press.
Liu, D., Liu, X., Tao, Y. (2005). *Exemplification of common sentence patterns in spoken Chinese (汉语口语常用句式例解)* Beijing, PRC: Beijing Daxue Chubanshe.
Liu, L. (2012). The selection and training of instructors of teaching CSL: A case study of the TCSOL program at Columbia University. In W. Wu (Ed.) *The First Academic Seminar on CSL (Putonghua) Education Under the New Curriculum at Universities in Hong Kong (Conference proceedings)* (pp. 8-8). Hong Kong: Yale-China Chinese Language Center, CUHK.
Lü, S. (1982). *Zhongguo Wenfa Yaolüe (中国文法要略)* Beijing, PRC: Commercial Press.
Ma, X. (2010). *The Theory About Chunks of Language and TCFL (语块理论与对外汉语教学)*. Unpublished MA dissertation, Liaoning Normal University. Liaoning Province, PRC.
Mackey, W. F. (1965). *Language teaching analysis* London: Longman.
Maclagan, M., Davis, B. & Lunsford, R. (2008). Fixed expressions, extenders and metonomy in the speech of people with alzheimer's disease. In S. Granger & F. Meunier (Eds.) *Phraseology: An interdisciplinary perspective* (pp. 175-187). Amsterdam/Philadelphia: John Benjamins Publishing Company.
McCarthy, M. (1991). *Discourse analysis for language teachers* Cambridge ; New York: Cambridge University Press.
McCarthy, M. (1990). *Vocabulary* Oxford: Oxford University Press.

Mel'cuk, I. (1998). Collocations and lexical functions. In A. Cowie (Ed.) *Phraseology: Theory, Analysis, and Applications* (pp. 23-53). Oxford: Clarendon Press.

Mel'cuk, I. (1995). Phrasemes in language and phraseology in linguistics. In M. Everaert, E. van der Linden, A. Schenk & R. Schreuder (Eds.) *Idioms: Structural and Psychological Perspectives* (pp. 167-232). Hillsdale NJ: Erlbaum.

Milton, J. (1998). Exploiting L1 and interlanguage corpora in the design of an electronic language learning and production environment. In S. Granger (Ed.) *Learner English on computer* (pp. 186-198). London; New York: Longman.

Moon, R. (1998). *Fixed expressions and idioms in English: a corpus-based approach* Oxford England: Clarendon Press.

Namba, K. (2010). Formulaicity in code-switching: Criteria for identifying formulaic sequences. In D. Wood (Ed.) *Perspectives on Formulaic Language: Acquisition and Communication* (pp. 129-150).

Nation, I. S. P. (1990). *Teaching and learning vocabulary* Boston: Heinle & Heinle Publishers.

Nattinger, J. R. (1988). Some current trends in vocabulary teaching. In R. Carter & M. McCarthy (Eds.) *Vocabulary and Language Teaching* (pp. 407-82). London and New York: Longman.

Nattinger, J. R. & DeCarrico, J. S. (1992). *Lexical phrases and language teaching* Oxford: Oxford University Press.

Pawley, A. (2009). Grammarians' languages versus humanists' languages and the place of speech act formulas in models of linguistic competence. In R. Corrigan, E. A. Moravcsik, H. Ouali & K. M. Wheatley (Eds.) *Formulaic Language volume 1: Distribution and Historical Change* (pp. 3-26). Philadelphia: John Benjamins.

Pawley, A. (2007). Developments in the study of formulaic language since 1970: A personal view. In P. Skandera (Ed.) *Phraseology and Culture in English* (pp. 3-45). New York: Mouton de Gruyter.

Pawley, A. & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmitt (Eds.) *Applied linguistics and language study.* (pp. 191-225). London: Longman.

Peters, A. M. (1983). *The units of language acquisition* Cambridge: Cambridge University Press.

Peters, A. M. (1977). Language learning strategies: does the whole equal the sum of the parts? *Language, 53(3)*, 560-573.

Peterson, S. L. (1998). *The research writer's phrase book: a guide to proposal writing and research phraseology* San Francisco: International Scholars Publications.

Qi, H. (2000). *Xiandai Hanyu Duanyu (现代汉语短语)* Shanghai, PRC: East China Normal University Press.

Qian, X. (2008). A Preliminary Study on Chinese Chunk. *Journal of Peking University (Philosophy & Social Sciences)* *（北京大学学报(哲学社会科学版)）, 45(5)*, 139-146.

Read, J. & Nation, P. (2004). Measurement of formulaic sequences. In N. Schmitt(Ed.) *Formulaic Sequences: Acquisition, processing and use* (pp. 23-36). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly, 10(1)*, 77-89.

Richards, J. C. & Rodgers, T. S. (2001). *Approaches and methods in language teaching* Cambridge; New York: Cambridge University Press.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics 11(2)*, 129-158.

Schmidt, R. W. (1983). Interaction, acculturation, and the acquisition of communicative competence: A case study of an adult. In N. Wolfson & E. Judd (Eds.) *Sociolinguistics and language acquisition* (pp. 137-174). Rowly, MA: Newbury House.

Schmitt, N. (2000). *Vocabulary in language teaching* Cambridge; New York: Cambridge University Press.

Schmitt, N. & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.) *Formulaic Sequences: Acquisition, processing and use* (pp. 1-22). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Schmitt, N., Dörnyei, Z., Adolphs, S. & Durow, V. (2004). Knowledge and acquisition of formulaic sequences: A longitudinal study. In N. Schmitt (Ed.) *Formulaic Sequences:*

Acquisition, processing and use (pp. 55-86). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Schmitt, N., Grandage, S. & Adolphs, S. (2004). Are corpus-derived recurrent clusters psycholinguistically valid? In N. Schmitt (Ed.) *Formulaic Sequences: Acquisition, processing and use* (pp. 127-152). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Schmitt, N. & Underwood, G. (2004). Exploring the processing of formulaic sequences through a self-paced reading task. In N. Schmitt (Ed.) *Formulaic Sequences: Acquisition, processing and use* (pp. 173-190). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Searle, J. R. (1979). *Expression and meaning: studies in the theory of speech acts* Cambridge, Eng.; New York: Cambridge University Press.

Searle, J. R. (1976). A Classification of Illucutionary Acts. *Language in Society, 5(1)*, 1-23.

Searle, J. R. (1969). *Speech acts: an essay in the philosophy of language* London: Cambridge U.P.

Shapira, R. G. (1978). The non-learning of English: Case study of an adult. In E. M. Hatch (Ed.) *Second Language Acquisition: a book of readings* (pp. 246-255). Rowley, MA: Newbury House.

Shi, Z., Hu, L. & Wang, X. (2010). *Zhuoyue Hanyu: gongsi shizhan pian (卓越汉语: 公司实战篇)* Beijing, PRC: Foreign Language Teaching and Research Press.

Shortall, T. (1996). What learners know and what they need to learn. In D. Willis & J. Willis (Eds.) *Challenge and change in language teaching* (pp. 31-41). Oxford: Heinemann.

Sinclair, J. (2008). The phrase, the whole phrase, and nothing but the phrase. In S. Granger & F. Meunier (Eds.) *Phraseology: An interdisciplinary perspective* (pp. 407-410). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Sinclair, J. (2008). Preface. In S. Granger & F. Meunier (Eds.) *Phraseology: An interdisciplinary perspective* (pp. xv-xviii). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Sinclair, J. & Renouf, A. (1988). A lexical syllabus for language learning. In R. Carter & M. McCarthy (Eds.) *Vocabulary and Language Teaching* (pp. 407-160). London and New York: Longman.

Sinclair, J. (1991). *Corpus, concordance, collocation* Oxford: Oxford University Press.

Skehan, P. (1998). *A cognitive approach to language learning* Oxford: Oxford University Press.

Skehan, P. (1996). Second language acquisition research and task-based instruction. In D. Willis & J. Willis (Eds.) *Challenge and change in language teaching* (pp. 17-30). Oxford: Heinemann.

Song, W. (2009). *语块与句子加工 - 以框架语为例.* Unpublished MA dissertation, East China Normal University. Shanghai, PRC.

Su, F. (2010). *A Study on Chinese Chunks and the Learning Strategies Thereof (汉语语块及其学习策略研究)*. Unpublished MA dissertation, Shanghai Jiaotong University. Shanghai, PRC.

Swales, J. (1990). *Genre analysis :English in academic and research settings* Cambridge; New York: Cambridge University Press.

Taguchi, N. (2007). Task difficulty in oral speech act production. *Applied Linguistics, 28(1)*, 113-135.

Teliya, V., Bragina, N., Oparina, E. & Sandomirskaya, I. (1998). Phraseology as a language of culture: Its role in the representation of a cultural mentality. In A. Cowie (Ed.) *Phraseology: theory, analysis and applications* (pp. 55-75). Oxford: Clarendon Press.

Tremblay, A. & Baayen, H. (2010). Holistic processing of regular four-word sequences: A behavioural and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.) *Perspectives on formulaic language: acquisition and communication* (pp. 151-173). London; New York: Continuum.

Underwood, G., Schmitt, N. & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (Ed.) *Formulaic Sequences: Acquisition, processing and use* (pp. 153-172). Amsterdam/Philadelphia: John Benjamins Publishing Company.

van Ek, J. A., Alexander, L. G., Fitzpatrick, M. A., Council of Europe & Pergamon Institute of English (1980). *Waystage English: an intermediary objective below Threshold Level in a European unitcredit system of modern language learning by adults* Oxford; New York: Published for and on behalf of the Council of Europe by Pergamon Press.

Wang, H. (2007). *A Study on the Chinese Chunk in Second Language Acquisition (二语习得中的汉语语块研究)*. Unpublished MA dissertation, Jinan University. Guangzhou, PRC.

Weinert, R. (2010). Formulaicity and usage-based language: Linguistic, psycholinguistic and acquisitional manifestations. In D. Wood(Ed.) *Perspectives on formulaic language: acquisition and communication* (pp. 1-20). London ; New York: Continuum.

White, R. V. (1988). *The ELT curriculum :design, innovation, and management* Oxford: Blackwell.

Widdowson, H. G. (1989). Knowledge of language and ability for use. *Applied Linguistics, 10(2)*, 128-137.

Widdowson, H. G. (1979). Directions in the teaching of discourse. In C. Brumfit & K. Johnson (Eds.) *The communicative approach to language teaching* (pp. 49-60). Oxford: Oxford University Press.

Widdowson, H. G. (1978). *Teaching language as communication* Oxford: Oxford University Press.

Wilkins, D. A. (1976). *Notional syllabuses : a taxonomy and its relevance to foreign language curriculum development* London: Oxford University Press.

Williams, M. (1988). Language taught for meetings and language used in meetings: is there anything in common? *Applied Linguistics, 9(1)*, 44-58.

Willis, D. (1990). *The lexical syllabus: a new approach to language teaching* London: Collins ELT.

Willis, D. & Willis, J. (1996). *Challenge and change in language teaching* Oxford: Heinemann.

Willis, D. & Willis, J. (1996). Consciousness-raising activities in the language classroom. In D. Willis & J. Willis (Eds.) *Challenge and change in language teaching* (pp. 63-76). Oxford: Heinemann.

Willis, J. (1996). A flexible framework for task-based learning. In D. Willis & J. Willis (Eds.) *Challenge and change in language teaching* (pp. 52-62). Oxford: Heinemann.

Wong Fillmore, L. (1976). *The second time around: cognitive and social strategies in second language acquisition.* Unpublished PhD thesis, Stanford University.

Wong, H. (2011). *A New Taxonomy of Chinese Formulaic Sequences and Its Implications on Teaching Chinese as L2*. Paper presented at 9th BCLTS International Symposium on Innovation and Standardisation in Teaching and Learning Chinese. 29 Jun – 1 July 2011. Edinburg.

Wood, D. (2010). *Formulaic language and second language speech fluency: background, evidence and classroom applications* London: Continuum.

Wood, D. (2010). Lexical clusters in an EAP textbook corpus. In D. Wood (Ed.) *Perspectives on formulaic language: acquisition and communication* (pp. 88-106). London; New York: Continuum.

Wood, D. (2002). Formulaic language in acquisition and production: Implications for teaching. *TESL Canada Journal, 20*, 1-15.

Wray, A. (2009). Identifying formulaic language: Persistent challenges and new opportunities. In R. Corrigan, E. A. Moravcsik, H. Ouali & K. M. Wheatley (Eds.) *Formulaic Language volume 1: Distribution and Historical Change* (pp. 27-51). Philadelphia: John Benjamins.

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics, 21(4)*, 463-489.

Wray, A. & Fitzpatrick, T. (2008). Why can't we just leave it alone? deviations from memorized language as a gauge of nativelike competence. In F. Meunier & S. Granger (Eds.) *Phraseology in foreign language learning and teaching* (pp. 259). Amsterdam ; Philadelphia: John Benjamins Pub. Co.

Wray, A. & Namba, K. (2003). Use of formulaic language by a Japanese-English bilingual child: A practical approach to data analysis. *Japan Journal of Multilingualism and Multiculturalism, 9(1)*, 24-51.

Wray, A. & Perkins, M. (2000). The function of formulaic language: An integrated model. *Language and communication 20(1)*, 1-28.

Wray, A. (2008). *Formulaic Language: Pushing the Boundaries* Oxford University Press.

Wray, A. (2004). 'Here's one I prepared earlier': Formulaic language learning on television. In N. Schmitt (Ed.) *Formulaic Sequences: Acquisition, processing and use* (pp. 249-268). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Wray, A. (2002). *Formulaic language and the lexicon* Cambridge ; New York: Cambridge University Press.

Wu, W. (2012). *Speaking proficiency assessment: theories and practices, ideals and reality*. Paper presented at The First Academic Seminar on CSL (Putonghua) Education Under the New Curriculum at Universities in Hong Kong. 30 Mar 2012. Hong Kong.

Wu, W. (2008). Pragmatic framework and its role in language learning: With special reference to Chinese. In W. E. A. Chan (Ed.) *Processes and Process-Orientation in Foreign Language Teaching and Learning* Singapore: National University of Singapore.

Wu, Y., He, S. & Wu, X. (2009). *The classification, the Lexicalization Degree and Teaching Thoughts of Chinese Chunks*. Paper presented at The First National Symposium on Formulaic Language Teaching and Research. 16-7 May 2009. Beijing, PRC.

Yang, C. (2010). *Lexical Chunks in Teaching Chinese as a Second Language (论对外汉语教学中的语块)*. Unpublished MA dissertation, Shaanxi Normal University. Shaanxi, PRC.

Yorio, C. A. (1989). Idiomaticity as an indicator of second language proficiency. In K. Hyltenstam & L. K. Obler (Eds.) *Bilingualism across the lifespan* (pp. 55-72). Cambridge: Cambridge University Press.

Yorio, C. A. (1980). Conventionalized Language Forms and the Development of Communicative Competence. *TESOL Quarterly, XIV(4)*, 433-442.

Zhang, B. (1998). *Hanyu yufaxue (汉语语法学)* Shanghai, PRC: Shanghai Jiaoyu Chubanshe.

Zhang, B. (1989). *Xian dai Han yu jing jie (现代汉语精解)* Shanghai, PRC: Shanghai Wenyi Chubanshe.

Zhang, F. (2005). Two questions in the study of Chinese idioms. *Yuyan Wenzi Yingyong (语言文字应用), 2*, 49-55.

Zhou, J. (2009). *Study of Lexical Chunk in TCFL - Taking 《an Outline Ranking Chinese Vocabularies》 as an Example (对外汉语语块研究-以《汉语水平词汇等级大纲》为例)*. Unpublished MA dissertation, East China Normal University.Shanghai, PRC.

Zhou, J. (2007). Reinforce the Language Chunk Teaching to Foster the Intuition of Chinese. *Journal of Jinan University (Philosophy and Social Sciences), 126(1)*, 99-104.

Zhou, J. (2004). *Hanyu cihui jiegoulun (汉语词汇结构论)* Shanghai, China: Shanghai cishu chubanshe.

Zimmerman, C. B. (1997). Historical trends in second language vocabulary instruction. In J. Coady & T. Huckin (Eds.) *Second Language Vocabulary Acquisition* (pp. 5-19). USA: Cambridge University Press.

Appendices

Appendix III.1 Sample NNS and NS transcripts

Non-native speaker sample

张老师~真对不起~~我~~我忘了~~哦~真的~最近我太忙~每天都有考试~~我 ~~我注~不太注意~跟你约~约好~~~~做考试~真对不起~~老师~可不可以给我~~啊可不可以~再一次~约好这个考试的~时间~行不行~~~我~真的~嫌~麻烦你~但是~最近~我~~我~好象我的身体也~比较弱~~~~~~~可能我~太忙

Native speaker sample

啊~不好意思啊教授~嗯那个~我同学昨天晚上生病了~然后~我带她去医院结果~~就在医院里面就陪她一直到~到早上结果睡过头了~然后结果就没~就没有赶回学校来参加考试~真是不好意思我们可不可以再~找一个时间或今天下午~我就可以来考试我已经准备好了~请再给我一次机会~~~~嗯真的是很不好意思~~~~~~我也没有想到会突然发生这种事情但突然就~~就说头很疼然后就就陪她去医院~~~~~~~~真的是不好意思麻烦你~拜托了~~~~下次不会再有这种情况发生了

Appendix III.2 Instructions and Examples to Judges (Identification of FSs)

请您标出所有您认为是语块的字串。

1) 语块必须由两个或两个以上经常同时出现的词组成。典型的语块包括成语、俗语、习用语（如'吃饭了没有？'、'你说呢？'、'你知道什么'和'哪儿的话'）、句头（如'总的来说'）、句框（如'能不能麻烦您 VP'、'请您 VP 一下好吗'）、动宾结构（如'吃'和'饭'）、状中结构（如'热烈'和'欢迎'）、动补结构（如'吓得'和'浑身哆嗦'）等。
2) 语块也包括对外汉语教科书中的句型、句子框架、连接词，如'又…又'、'越…越'、'非…不可'、'不但…而且'、'因为…所以'、'NP 跟 NP 结婚'、'向 NP 转达 NP 的问候'等。
3) 如果您认为某个句子框架是语块，请检查一下该框架是否包含多于一个词。例如，同样是表达请求功能的句子框架，'请您 VP '、'请您 VP 一下'、'请您 VP 一下，

好吗？’都算得上是语块，’请 VP’则不算。

4) 如果一个字串是您上课时经常作为一个整体来领说的，那么它就有可能是一个语块。是否标示出来请您自行判断。
5) 标示语块时，不用考虑字串所在的句子是否有语法、语义或语用错误。不论句子有什么问题，请照样把其中的所有语块标出来。
6) 如果一个字串属于语块，但其中部分有误，只要无误的部分包含多于一个词，便请将无误部分都标出，例如’顶天立土’中的’顶天立’。
7) 如果一个字串属于改动过的语块，至少将未改动部分都标出，例如’一举四得’中的’一举’和’得’。若您认为’一举四得’也是语块，请整个标出。
8) 标示语块时，请尽量采用划一的标准。
9) 尽可能一次完成。
10) 标示完成后，请尽量在另外一天重新检查一遍。

Appendix III.3 Instructions and Examples to Judges (Error Correction)

改正的单位可以是字、词、词组或整个句子，改正的原因可以是用错词、搭配不当、语法错误或语用错误，如下面的例子（’~’表示大约 0.5 秒的停顿）。

请注意以下事项：

1) 跟平常判作业一样，有错从严改，不论是语法、语义还是语用的错误。
2) 把错处划掉，在旁边写上正确的。
3) 若整句话不妥/不地道/不合宜，那就整句换掉，用中国人常用的话取代之。
4) 需要改正的是口语语料，其中会有开了头但没说完的句子（如英语的‘I would like to er I am very er may I ask you a question’ 中的前半句）或重复的句子成分（如‘I I I don’t er don’t know’中的‘I’ 和 ‘don’t’），还有 ‘ah’, ‘uh’ 和 ‘er’ 等。若属正常范围，不必改。
5) 请尽量采用划一的标准。
6) 尽可能一次完成。
7) 完成后，请尽量在另外一天重新检查一遍。

例子：

‘~~怎么可以说~~~我~~没有特别的书喜欢看~~~有的时候~~看~~谈恋爱书~~看~~情书~~~

怎么说呢　　没有什么特别喜欢看的书　　爱情小说/恋爱小说

不过~我~但是~~我也看~~有的旅游关系的杂志~~~~嗯~~啊~~虽然我不常去旅行~~~

一些旅游杂志/一些跟旅游有关系的杂志

不常~~~有机会~~~但是一边看杂志~~一边想一想也~很快乐~~~~没有其他~~~~没有~~

一边想象一下也挺开心的　　没有别的了　　没有了

~~啊~我~~看过不太少不同的~~漫画~~从小孩子的时候到现在看了~~’

不少各种各样的　　从小时候到现在一直都看

Appendix IV.1 Overall distribution of 3 big formal types of FSs

	NNS FSs				NS FSs			
	NNS Total	**CR (correct)**	**LK (likely)**	**ULK (unlikely)**	**NS Total**	**CR (correct)**	**LK (likely)**	**ULK (unlikely)**
COLLOCATIONS	**326**	**169**	**79**	**78**	**805**	**739**	**29**	**37**
% of sub-total	**100%**	**51.84%**	**24.23%**	**23.93%**	**100%**	**91.80%**	**3.60%**	**4.60%**
% of total	**48.58%**	**25.19%**	**11.77%**	**11.62%**	**46.72%**	**42.89%**	**1.68%**	**2.15%**
FRAMES	**239**	**139**	**40**	**60**	**657**	**595**	**20**	**42**
% of sub-total	**100%**	**58.16%**	**16.74%**	**25.10%**	**100%**	**90.56%**	**3.04%**	**6.39%**
% of total	**35.62%**	**20.72%**	**5.96%**	**8.94%**	**38.13%**	**34.53%**	**1.16%**	**2.44%**
POLYWORDS	**106**	**84**	**8**	**14**	**261**	**247**	**7**	**7**
% of sub-total	**100%**	**79.25%**	**7.55%**	**13.21%**	**100%**	**94.64%**	**2.68%**	**2.68%**
% of total	**15.80%**	**12.52%**	**1.19%**	**2.09%**	**15.15%**	**14.34%**	**0.41%**	**0.41%**
Total non-TSSS FSs	**671**	**392**	**127**	**152**	**1723**	**1581**	**56**	**86**
%	**100%**	**58.42%**	**18.93%**	**22.65%**	**100%**	**91.76%**	**3.25%**	**4.99%**

Appendix IV.2.1 Collocation: Adverb + (adjective/verb) + adverb

(CR: correct; LK: likely; ULK: unlikely)

		NNS	NS

FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
不是不[183]	not be not					1	1		
不是很	not be very					3	3		
不太	not very	4	1	2	1	3	3		
不用再	no need again					1	1		
还可以	unexpectedly especially					2	2		
还一直	unexpectedly all the way					1	1		
很晚才	very late then					2	2		
很早就	very early already					1	1		
忽然就	suddenly then					1	1		
就一直	then + all the way[184]	1			1	1	1		
就马上	then immediately	1	1						
就只好	then have to					1	1		
天天总	everyday unexceptionally					1	1		
突然就	suddenly already					2	2		
也都	also unexceptionally					1	1		
也就	also then					1	1		
一直都	all the way + unexceptionally					1	1		

[183] In order to facilitate the explanations of the FSs, when writing Pinyin, we do not always follow the basic rules of Hanyu Pinyin Orthography as specified by the government of People's Republic of China (last version published 1 July 1996. http://www.china-language.gov.cn/gfbz/shanghi/025.htm)

[184] '+' is added only when difficult to tell the relationship between Pinyin words and their meanings.

Totals	6	2	2	2	23	23	0	0

Analysis:

1. FS ratio: **0.27**, with 6 by NNSs and 22 by NSs (All collocations: 0.40; overall: 0.39)
2. Error ratios: **66.7%** by NNSs (All collocations: 48.2%; overall: 41.6%) and **0%** by NSs (All collocations: 8.21%; overall: 8.25%).
3. Other findings:
   - 3.1 Overlapping rate (NNS FSs that are also used by NSs[185]): 83.3% (5 out of 6).
   - 3.2 NNSs heavily rely on *不太* (4 out of 6), resulting in a smaller variety of FSs.
   - 3.3 NSs' FSs are of 16 different types and distributed more evenly, each with 1 to 3 tokens.
   - 3.4 There are 5 FSs in NS data ended with *就* and 3 others ended with *都* or 总, but there are none in NS data, indicating that FSs of this kind might be difficult for NNS to acquire. However, FSs beginning with *就* seems to be easier as NNSs produced 2, as NSs did.
4. Conclusions:
   - 4.1.1 Proportionally NNSs used fewer FSs of this type.
   - 4.1.2 NNSs made more mistakes than their average level.
   - 4.1.3 Most NNSs FSs are also used by NSs.
   - 4.1.4 NNSs rely on smaller varieties of FSs.

Appendix IV.2.2 Collocation: Adverb + (adverb) + verb/adjective

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
并不是	actually not be					1	1		
不太注意	not + very + pay attention	1			1				
诚恳地说	sincere PRT say					1	1		
从来没有	all along + not have					1			

[185] Whether used correctly is not considered here.

									1
到底为什么	exactly for what					1	1		
都不好	all not good	1			1				
都不是	all not be					1	1		
都不是很好	all not be + very good					1	1		
都好	all good					1	1		
都会	all will					2	2		
都可以	all can	3	3						
都没有	all + not have	2	1		1				
都是	all be	1	1			1	1		
都是没有	all + be + not have					1			1
都要	all need	1	1						
都有	all have	1	1			3	2		1
都准备了	all prepare PRT					1	1		
还记得	still remember					1	1		
还没有	still + not have					2	1		1
还能够	still can					1	1		
还要	still need	1			1				
还是想	still want					1	1		
还是希望	still hope					2	2		
还是要	still need					1	1		
好好儿把握	well grasp					1	1		

忽然忘了	suddenly forget PRT					1	1		
竟然错过了	unexpectedly miss PRT					1	1		
竟然忘记了	unexpectedly forget PRT					1	1		
就会	then will	1		1		2	1		1
就可以	then can					5	2		3
绝非	absolutely not					1	1		
绝对服从	absolutely obey					1	1		
肯定不会	definitely not will					1	1		
可能会	probably will					1	1		
努力学习	diligently study	1		1					
认真对待	seriously treat					1	1		
认真上课	seriously + attend class					1	1		
如何才能	how exactly can					1	1		
深切地认识到	deep PRT realize					1			1
坦诚地说	frank PRT speak					1	1		
特别安排	specially arrange					1	1		
特别珍惜	specially treasure					1	1		
特地安排	specially arrange					2	2		
特意安排	specially arrange					2	2		
突然发生	suddenly happen					1	1		
突然患病	suddenly + get sick	1	1						

突然生病	suddenly + get sick					1	1		
突然撞	suddenly + bump into	1		1					
完全不行	completely not okay					1	1		
完全没有	completely + not have	1			1				
完全忘了	completely forget PRT	2	2						
完全忘记了	completely forget PRT	2	2						
一定不会	definitely not will	1	1			4	4		
一定会	definitely will					5	5		
一定要	definitely need					3	3		
应该会	likely will					1	1		
应该能	likely can					1	1		
一时忘了	accidentally forget PRT					1	1		
一直到	all the way + to					2	2		
有点儿发烧	a bit + have a fever					1	1		
有点儿感冒	a bit + catch a cold					1	1		
有点儿过分	a bit + excessive					1	1		
早醒	early wake					1	1		
怎么办	how handle					1	1		
真诚地道歉	sincere PRT + express apologies					1	1		
只可以	only can	1			1				
只能	only can					3	3		

只能够	only can					1	1		
只有	only have					1	1		
准时到	punctually arrive					2	2		
Total		22	13	3	6	82	73	0	9

Analysis:

1 FS ratio: **0.27**, with 22 by NNSs and 82 by NSs (All collocations: 0.40; overall: 0.39)

2 Error ratios: **40.9%** by NNSs (All collocations: 48.2%; overall: 41.6%) and **12.3%** by NSs (All collocations: 8.21%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate: 13.6% (3 out of 22).

3.2 NSSs' FSs are of 17 different types, each with 1 to 3 tokens.

3.3 NSs' FSs are of 59 different types, each with 1 to 5 tokens.

3.4 NNSs seem to rely on FSs with *都* (9 tokens versus 11 in NS), *完全* (5 versus 1 in NS), *可以* (4 versus 5) and *突然* (2 versus 2).

4 Conclusions:

4.1.1 Proportionally NNSs used fewer FSs of this type.

4.1.2 NNSs made more mistakes than their average level.

4.1.3 NNSs and NSs have very different choices of FSs.

4.1.4 NNSs rely on smaller varieties of FSs.

Appendix IV.2.3 Collocation: Causative verb + pronoun:

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
拜托您	request you	1			1				
给我	let me	1	1			4	2		2
恳请您	beg you					2	2		

麻烦您	trouble you					2	2		
请您	beg you	1	1			2	2		
请求您	beg you					4	4		
求您	beg you					1			1
让您	ask you					3	3		
让我	let me	9	9			7	7		
Total		12	11	0	1	25	22	0	3

Analysis:

1 FS ratio: **0.48**, with 12 by NNSs and 25 by NSs (All collocations: 0.40; overall: 0.39)

2 Error ratios: **8.3%** by NNSs (All collocations: 48.2%; overall: 41.6%) and **12%** by NSs (All collocations: 8.21%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate: 91.7% (11 out of 12).

3.2 NSSs' FSs are of 4 different types, 3 with 1 token and 1 with 9 tokens.

3.3 NSs' FSs are of 8 different types, each with 1 to 7 tokens but distributed more evenly (mostly with 2-4 tokens).

3.4 The FS with 9 tokens from NNSs coincides with the one with 7 tokens from NSs.

3.5 NSs have a far greater repertore of synomyns to express 'beg': *恳请, 请, 请求* and *求*.

4 Conclusions:

4.1.1 NNSs and NSs have very similar choices of FSs.

4.1.2 NNSs made fewer mistakes than their average level.

4.1.3 NNSs seem to rely on one FS: 'rang wo' (75% versus 28% in NS data).

Appendix IV.2.4 Collocation: Conjunction + adverb + (verb)

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
那就	in that case + then	1	1	0	0	3	3		
然后就	afterwards then					6	6		
然后就是	afterwards then be					1	1		
所以就	therefore then					4	4		
Total		1	1	0	0	14	14	0	0

Analysis:

1 FS ratio: **0.07**, with 1 by NNSs and 14 by NSs (All collocations: 0.40; overall: 0.39)

2 Error ratios: **0%** by NNSs (All collocations: 48.2%; overall: 41.6%) and **0%** by NSs (All collocations: 8.21%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate: 100% (1 out of 1).

3.2 NSs' FSs are of 4 different types with uneven distribution (with 1 to 6 tokens).

3.3 This kind of FSs all end with *就* and seems not well mastered by NNSs, like those discussed in IV.2.1.

4 Conclusions:

4.1.1 NNSs seem not good at using FSs of this kind.

4.1.2 This type of FSs seems to be of limited variations but quite high frequencies in NS data.

Appendix IV.2.5 Collocation: Fillers

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
还是	still be					1	1		

就是	then be					16	16		
就是说	then be say					4	4		
就说	then say					2	2		
也是	also be					1	1		
有点儿	have + a bit					1			1
在这里	at here					1	1		
真的	really PRT	1	1			1	1		
Total		1	1			27	26		1

Analysis:

1 FS ratio: **0.04**, with 1 by NNSs and 27 by NSs (All collocations: 0.40; overall: 0.39)

2 Error ratios: **0%** by NNSs (All collocations: 48.2%; overall: 41.6%) and **3.7%** by NSs (All collocations: 8.21%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate: 100% (1 out of 1).

3.2 NSs' FSs are of 8 different types, distributed very unevenly (5 types with 1 token each, and the 3 types with 2 to 16 tokens are *就是, 就是说* and *就说*, which are structurally and semantically similar).

4 Conclusion:

4.1.1 Formulaic fillers were quite popular among NSs but was barely used by NNSs.

4.1.2 NSs seem to favor a small variety of formulaic fillers.

Appendix IV.2.6 Collocation: Juxtaposed nouns

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK

白天和晚上	day and night	1			1				
妈妈爸爸[186]	mom and dad	1	1						
学习和工作	study and work					1	1		
Total		2	1	0	1	1	1	0	0

Analysis:

1 FS ratio: **2**, 2 by NNSs and only 1 by NSs (All collocations: 0.40; overall: 0.39)

2 Error ratios: 5**0%** by NNSs (All collocations: 48.2%; overall: 41.6%) and **0%** by NSs (All collocations: 8.21%; overall: 8.25%).

3 Overlapping rate: 100% (0 out of 2).

4 Due to the limited frequencies, it is difficult to draw any valid conclusions.

Appendix IV.2.7 Collocation: Measure word + noun

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English * M: measure word	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
次补考	M* + makeup test					6	5	1	
次错误	M mistake					1	1		
次机会	M chance	4	1	3		54	51	2	1
次考试	M test	1	1			9	9		
次事情	M matter					1	1		

[186] Although *爸爸妈妈* is more common in Chinese, *妈妈爸爸* is also acceptable (CCL corpus has 744 occurances of the former and 23 of the latter on July 3, 2010).

段时间	M time					2	2		
份声明	M declaration					1	1		
个车	M vehicle	1		1					
个错误	M mistake					1	1		
个东西	M thing	1	1			1	1		
个建议	M suggestion					2	1		1
个借口	M excuse	2			2	1	1		
个机会	M chance	13	6	7		8	7	1	
个考试	M test	9	4	2	3	11	10	1	
个理由	M reason					2	2		
个忙	M favor	1	1						
个情况	M situation	1			1	1	1		
个人	M person					1	1		
个事	M matter					2	2		
个时间	M time	4	2	1	1	2	1		1
个事情	M matter	2	2			7	7		
个事儿	M matter					1	1		
个失误	M mistake					1	1		
个通宵	M sleepless whole					1	1		

	night								
个同学	M classmate					1	1		
个问题	M problem					1	1		
个小时	M hour	1		1					
个学期	M semester					3	2	1	
个要求	M requirement					1	1		
个因素	M factor					1	1		
个邮件	M email					1	1		
个原因	M reason					1	1		
个约会	M date	1			1				
个钟头	M hour	1	1						
个自行车	M bike	1	1						
件事	M matter					3	3		
件事情	M matter					4	3	1	
门考试	M test					4	4		
门课	M course					6	6		
门学科	M course					1	1		
种情况	M situation					1	1		
种事情	M matter					1	1		

种问题	M problem					1	1		
种习惯	M habit					1	1		
种学生	M student	1			1				
Total		44	20	15	9	147	137	7	3

Analysis:

1 FS ratio: **0.30**, with 44 by NNSs and 147 by NSs (All collocations: 0.40; overall: 0.39)

2 Error ratios: **54.5%** by NNSs (All collocations: 48.2%; overall: 41.6%) and **6.8%** by NSs (All collocations: 8.21%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate: 84.1% (37 out of 44).

3.2 NSSs' FSs are of 16 different types, 10 with 1 token, 4 with 2 to 4 tokens, and 2 with 9 or 13.

3.3 NSs' FSs are of 38 different types, 22 with 1 token, 9 with 2 to 4 tokens, 6 with 6 to 11, and 1 with 54.

3.4 As constrained by the nature of the language task, the nouns with highest frequencies are *机会* (17 in NNS, 62 in NS) and *考试* (10 in NNS, 20 in NS). However, their collocations with measure words are different. NNSs used more *个* with both nouns, while NSs also used more *个* with *考试*, but overwhelmingly more *次* with *机会*.

3.5 NNSs used 3 measure words and 14 nouns while NSs used 7 measure words and 29 nouns.

4 Conclusions:

4.1.1 Proportionally NNSs used fewer FSs of this type.

4.1.2 NNSs made more mistakes than their average level.

4.1.3 Most NNSs FSs are also used by NSs.

4.1.4 NNSs covered fewer topics in their discourses, as reflected by their narrower varieties of nouns.

Appendix IV.2.8 Collocation: Modifier + noun as head word

(CR: correct; LK: likely; ULK: unlikely)

		NNS	NS

FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
宝贵的机会	precious PRT chance					1	1		
别的日子	other day	1		1					
别的时间	other time	1			1				
不可宽恕的错误	unable forgive PRT mistake					1	1		
充分的准备	sufficient PRT preparation					1	1		
第二次的机会	number two M PRT chance = second chance					1	1		
第一次考试	number one M test					1	1		
各种各样的状况	various PRT situation					1	1		
好借口	good excuse	1			1				
好机会	good chance					1		1	
很长时间	very long time					2	2		
很大的麻烦	very big PRT trouble	3	3			2	2		
很大的失误	very big PRT fault					1	1		
很多事儿	very many matters					1	1		
合适的理由	suitable PRT excuse					1	1		

家庭的问题	family PRT problem	1	1						
今天这次	today this M					1	1		
考试时间	test time					1	1		
良好的习惯	good PRT habit					1	1		
另外一个机会	other one M chance = another chance	1			1				
某些事情	certain some matters					1	1		
那个事	that matter					1	1		
那么长时间	such long time					1	1		
其他的时间	other PRT time	1			1				
私人的问题	private PRT problem	1	1						
特别原因	special reason					1	1		
特殊的原因	special PRT reason					2	2		
无意识状态	unconscious state of mind					1			1
下一次的补考	next one M PRT + makeup test					1	1		
学习成果	study achievement					1	1		
学校的功课	school PRT work	1	1						
要紧的事	important PRT matter	1	1						

一些事情	some matter					1			1
再一次机会	again one M chance	3			3	1	1		
这次补考	this M + makeup test					6	5	1	
这次错误	this M mistake					1	1		
这次机会	this M chance					5	4		1
这次考试	this M test					7	7		
这段时间	this M time					2	2		
这件事	this M matter					3	3		
这件事情	this M matter					4	3	1	
这门课	this M course					5	5		
这学期	this semester					1	1		
这一次机会	this one M chance					1	1		
这一门课	this one M course					1	1		
这一门学科	this one M course					1	1		
这种事情	this M matter					2	2		
这种问题	this M problem					1	1		
这个问题	this M problem					1	1		
这个东西	this M thing					1	1		
这个机会	this chance					2	1		1

这个考试	this test	8	4	2	2	9	9		
这个情况	this situation	1			1				
这个事情	this matter	1	1			4	4		
这个事儿	this matter					1	1		
这个因素	this factor					1	1		
真话	true words	1			1				
这些部门	these departments					1	1		
这样的问题	this PRT problem					1	1		
重要的事情	important PRT matter					1	1		
最后的机会	last PRT chance	1			1				
最后一次机会	last one M chance					1	1		
最后一个机会	last one M chance	2	2						
Total		29	14	3	12	91	84	3	4

Analysis:

1 FS ratio: **0.32**, with 29 by NNSs and 91 by NSs (All collocations: 0.40; overall: 0.39).

2 Error ratios: **52%** by NNSs (All collocations: 48.2%; overall: 41.6%) and **7.7%** by NSs (All collocations: 8.21%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate: 52% (15 out of 29).

3.2 NSSs' FSs are of 17 different types, mainly with 1-3 tokens, and 1 with 8 tokens (average 1.7 tokens).

3.3 NSs' FSs are of 50 different types, mostly with 1 to 2 tokens and a few with 3-9 (average 1.82).

3.4 *这个考试* is the top combination in both groups.

3.5 NNSs used 14 modifiers and 11 nouns, while NSs used 36 modifiers and 22 nouns.

4 Conclusions:

4.1.1 Proportionally NNSs used fewer FSs of this type.

4.1.2 NNSs made more mistakes than their average level.

4.1.3 NNSs and NSs have moderately similar choices of FSs.

4.1.4 NNSs covered fewer topics in their discourses, as reflected by their narrower varieties of nouns.

Appendix IV.2.9 Collocation: Noun + adverb

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
改天再	another day + then	4	2	2					
结果就	as a result + then					3	3		
以后才	afterwards then	1	1						
一早就	early morning + already					1	1		
最后再	in the end + then					2	2		
Total		5	3	2	0	6	6	0	0

Analysis:

1 FS ratio: **0.83**, with 5 by NNSs and 6 by NSs (All collocations: 0.40; overall: 0.39)

2 Error ratios: **40%** by NNSs (All collocations: 48.2%; overall: 41.6%) and **0%** by NSs (All collocations: 8.21%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate: 0% (0 out of 5).

3.2 NNSs seem to rely on *改天再* (4 tokens used by 3 participants).

3.3 NNSs seem to be weaker in using FSs ending with *就* (consistent with findings in IV.2.1 and IV.2.3).

4 Conclusions: 4.1.1 Proportionally NNSs used more FSs of this type. 4.1.2 NNSs' mistakes are on average level. 4.1.3 NNSs and NSs have completely different choices of FSs. 4.1.4 NNSs rely on smaller varieties of FSs.

Appendix IV.2.10 Collocation: Noun + verb

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
一早起来	early morning + get up					2	2		
Total						2	2	0	0

Analysis:

1 FS ratio: **0**, with 0 by NNSs and 2 by NSs (All collocations: 0.40; overall: 0.39)

2 Error ratios: **0%** by NSs (All collocations: 8.21%; overall: 8.25%).

3 Due to the limited frequencies, it is difficult to draw any valid conclusions.

Appendix IV.2.11 Collocation: Place + direction

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
家里	home in	1	1			1	1		
门外	door out					1	1		
心上	heart up					1	1		

Total	1	1			3	3		

Analysis:

1 FS ratio: **0.33**, with 1 by NNSs and 3 by NSs (All collocations: 0.40; overall: 0.39)

2 Error ratios: **0%** by NNSs (All collocations: 48.2%; overall: 41.6%) and **0%** by NSs (All collocations: 8.21%; overall: 8.25%).

3 Overlapping rate: 100% (1 out of 1).

4 Due to the limited frequencies, it is difficult to draw valid conclusions.

Appendix IV.2.12 Collocation: Preposition + noun:

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
因故	because of + reason					2	1	1	
在一块儿	at + same place					1	1		
Total						3	2	1	

Analysis:

1 FS ratio: **0**, with 0 by NNSs and 3 by NSs (All collocations: 0.40; overall: 0.39).

2 Error ratios: **33.3%** by NSs (All collocations: 8.21%; overall: 8.25%).

3 Due to the limited frequencies, it is difficult to draw any valid conclusions.

Appendix IV.2.13 Collocation: Redundant words

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK

博士您	Dr. you = you (respectful form)					5	5		
都已经	already already					1	1		
事情本身	matter itself					1	1		
统统都	all all	1			1				
我自己	I myself	1			1	2	2		
再继续	continue continue	1	1						
再另外	another another					1	1		
Total		3	1		2	10	10		

Analysis:

1 FS ratio: **0.30**, with 3 by NNSs and 10 by NSs (All collocations: 0.40; overall: 0.39)

2 Error ratios: **66.7%** by NNSs (All collocations: 48.2%; overall: 41.6%) and **0%** by NSs (All collocations: 8.21%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate (NNS FSs that are also used by NSs): 33.3% (1 out of 3).

3.2 FSs of this type can be divided into two types: norminal (*博士您*, *我自己*, *事情本省*) and adverbial[187] (*都已经*, *统统都*, *再继续*, *再另外*). NNSs production covers both types.

3.3 The most important FS consisted of a title and a pronoun, i.e. *博士您*, constitutes half of NS occurrence but is not used by any NNSs[188].

4 Conclusions:

4.1.1 Proportionally NNSs used fewer FSs of this type.

4.1.2 NNSs made more mistakes than their average level.

4.1.3 NNSs seem unable to use the most popular FS of this type used by NSs.

[187] If the two adverbs are exactly the same, the FS is classified as another type: Repeated words (see *非常非常* in V.3.2.14).

[188] Actually there is one such FS candidate in NNS data produced by an ethnic Chinese *我妈妈她* (meaning 'my mother she') but is discarded because of insufficient frequency in CCL online corpus and data collected in this study.

Appendix IV.2.14 Collocation: Repeated words

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
对不起了对不起了	sorry PRT sorry PRT					1	1		
非常非常	very very	2	2			5	5		
好多好多	very many very many					1	1		
清清楚楚	clear clear					1	1		
一点点	a bit + bit					1	1		
真的真的	really PRT really PRT					1	1		
Total		2	2			10	10		

Analysis:

1 FS ratio: **0.20**, with 2 by NNSs and 10 by NSs (All collocations: 0.40; overall: 0.39)

2 Error ratios: **0%** by NNSs (All collocations: 48.2%; overall: 41.6%) and **0%** by NSs (All collocations: 8.21%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate: 100% (2 out of 2).

3.2 There are 6 types of FSs identified, and they are all structurally distnctive, as shown below:

Table IV.2.14.1 Collocation – Repeated words (CR: correct; LK: likely; ULK: unlikely)

	Literal English	Structure

对不起了对不起了	sorry PRT sorry PRT	(FS + PRT) x2 Duibuqi is a FS and the PRT represent an exclamation. So *duibuqi le* can by itself fuction as an utterance. The whole utterance was repeated.	
非常非常	very very	adverb x 2	
好多好多	very many very many	(adverb + adjective) x 2	
清清楚楚	clear clear	The two syllables of a bisyllabic adjective repeated themselves to form an AA-BB structure.	
一点点	a bit + bit	The second part of a special measure word (internal structure: one + measure) repeats itself.	
真的真的	really PRT really PRT	(adverb + PRT) x 2	

3.3 NNSs only have one type *非常非常*, which is also the one outstandingly popular among NSs.

4 Conclusions:

4.1.1 Proportionally NNSs used fewer FSs of this type.

4.1.2 All FSs were correctly used.

4.1.3 All NNSs FSs are also used by NSs.

4.1.4 NNSs rely on only one FS.

Appendix IV.2.15 Collocation: Subject + predicate[189]

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK

[189] In Subject + Predicate Collocations, the subjects are not always immediately followed by the predicates (normally sperated by adverbs or adverbial phrases).

病严重	illness serious	1	1						
肚子痛	stomach ache	1	1						
工作忙	work busy	1	1						
考试重要	test important	1		1		6	6		
脑子糊里糊涂	brain muddled	1			1				
脑子忘记	brain forget					1	1		
情况不好	condition not good	1			1				
情况发生	situation appear					1	1		
情况好	condition good					1	1		
情况特殊	situation special					1	1		
身体不好	health not good	1		1		3	3		
身体不舒服	health not well	1	1			1	1		
身体弱	health weak	1	1						
时间过	time pass	1			1	1	1		
时间晚	time late	1		1					
时间允许	time allow					1	1		
事儿多	matter many					1	1		
态度不对	attitude not right					1	1		
*头疼**	head ache	1	1			1	1		

头晕	head + feel dizzy					2	2		
头脑发昏	mind + feel giddy					1	1		
习惯不好	habit not good					1	1		
心脏有问题	heart has problem	1	1						
学习忙	study busy					1	1		
要求过分	requirement excessive					1	1		
意义重大	significance great					1	1		
状况出现	situation appear					1	1		
Total		13	7	3	3	27	27		

Analysis:

1 FS ratio: **0.48**, with 13 by NNSs and 27 by NSs (All collocations: 0.40; overall: 0.39)

2 Error ratios: **46%** by NNSs (All collocations: 48.2%; overall: 41.6%) and **0%** by NSs (All collocations: 8.21%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate: 38% (5 out of 13).

3.2 NNSs used 10 different subjects and 11 different predicates while NSs used 14 and 18.

3.3 NSSs used 13 types of perfectly distributed FSs (1 token each), while NSs used 19 types but 16 of them have 1 token, and 3 types with 2 - 6 tokens.

3.4 53.8% (7 out of 13) and 14.8% (4 out of 27) of NNS and NS FSs are related to health issues, seemingly indicating that NNSs could cover fewer topics in their discourses (also see IV.2.8 and IV.2.9).

4 Conclusions:

4.1.1 NNSs' performance is on average but far worse than NSs'.

4.1.2 NNSs and NSs have very different choices of FSs.

4.1.3 NNSs rely on fewer areas of topics.

* Pinyin written as one word in Contemporary Chinese Dictionary.

Appendix IV.2.16 Collocation: Verb + adverb

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
不知道怎么就	don't know how already = don't know why but					1	1		
忙起来就	busy begin then = while too busy with					1	1		
Total						2	2		

Analysis:

1 FS ratio: **0**, with 0 by NNSs and 2 by NSs (All collocations: 0.40; overall: 0.39)

2 Error ratios: **0%** by NSs (All collocations: 8.21%; overall: 8.25%).

3 Valid conclusions cannot be drawn but both of the 2 FSs of this type end with *jiu*, which seems difficult for NNSs (see IV.2.1 and IV.2.3).

Appendix IV.2.17 Collocation: Verb + complement of degree

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
记得清清楚楚	remember PRT clearly					1	1		
起得晚	get up + PRT + late					1	1		

疼得厉害	ache PRT severely	1	1						
忘得一干二净	forget PRT completely					1	1		
准备得充分	prepare PRT very sufficiently					1	1		
Total		1	1			4	4		

Analysis:

1 FS ratio: **0.25**, with 1 by NNSs and 4 by NSs (All collocations: 0.40; overall: 0.39)

2 Error ratios: **0%** by NNSs (All collocations: 48.2%; overall: 41.6%) and **0%** by NSs (All collocations: 8.21%; overall: 8.25%).

3 Overlapping rate: 0% (0 out of 1).

4 Valid conclusions cannot be drawn but half of the 4 NS FSs of this type are polyword adjectives (see IV.3.1).

Appendix IV.2.18 Collocation: Verb + complement of direction

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
*出来**	out come = come out					1		1	
*出去**	out go = go out					2	2		
带来	carry come = bring	1	1			1	1		
带去	carry go = take to					1	1		
*过来**	move come = move closer					3	2		1
*过去**	move go = move away					2	2		

*回来**	return come = come back					3	3		
忙起来	busy up = become busy					1	1		
*起来**	get come = get up	3	3			13	12		1
送去	deliver go = send to	2	2						
*下来**	down come = come down					2	2		
醒来	wake come = wake up					1			1
Total		6	6			30	26	1	3

Analysis:

1 FS ratio: **0.2**, with 6 by NNSs and 30 by NSs (All collocations: 0.40; overall: 0.39)

2 Error ratios: **0%** by NNSs (All collocations: 48.2%; overall: 41.6%) and **14%** by NSs (All collocations: 8.21%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate: 67% (4 out of 6).

3.2 Both NNSs and NSs heavily rely on *起来* (50% and 45%).

3.3 While NSs used 10 of the 11 types of FSs, NNSs only used 3, even though they were all learned in elementary levels. This might indicate NNSs' limited repertore in describing actions, on top of their insufficiency in modifiers and nouns (see IV.2.8, IV.2.9 and IV.2.15).

4 Conclusions:

4.1.1 Proportionally NNSs used fewer FSs of this type but with very good quality.

4.1.2 Most NNSs FSs are also used by NSs.

4.1.3 NNSs's reliance on smaller varieties of FSs is well-justified in this case, as NSs also did the same.

* These combinations can be found in Contemporary Chinese Dictionary and do not have spaces between the two syllables.

Appendix IV.2.19 Collocation: Verb + complement of movement

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
安排一下	arrange one action*	3	1	1	1	1	1		
摆弄一下	tinker one action					1			1
帮一次	help one time					1	1		
补考一次	makeup a test one time					1	1		
多一个	more one unit	2			2				
解释一下	explain one action					1	1		
考一次	take (a test) one time	1	1			1	1		
说明一下	explain one action	1		1		1	1		
调一下	adjust one action					1	1		
想一想	think one think					1	1		
原谅一次	forgive one time					1	1		
准备一下	prepare one action					1	1		
做一次	do one time					1	1		
Total		7	2	2	3	12	11	0	1

Analysis:

1 FS ratio: **0.58**, with 7 by NNSs and 12 by NSs (All collocations: 0.40; overall: 0.39)

2 Error ratios: **71%** by NNSs (All collocations: 48.2%; overall: 41.6%) and **8%** by NSs (All collocations: 8.21%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate: 71% (5 out of 7).

3.2 Three types of FSs are identified: verb + 一下; verb + 一次/个 and verb + 一 + verb.

3.3 NNSs heavily rely on 安排一下 (3 out of 7), resulting in a smaller variety of FSs.

3.4 NSs' FSs covered 12 of the 13 types, with perfectly even distribution (1 token each).

4 Conclusions:

4.1.1 Proportionally NNSs used more FSs of this type.

4.1.2 NNSs made more mistakes than their average level.

4.1.3 Most NNSs FSs are also used by NSs.

4.1.4 NNSs rely on smaller varieties of FSs.

* The 'one action' means doing a bit of something or doing for a short period of time

Appendix IV.2.20 Collocation: Verb + complement of potential

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
毕不了业	complete not able work = cannot graduate					1	1		
记不起来	remember not up come = cannot recall from memory	2	1		1				
说不出来	say not out come = don't know how to express					1		1	
Total		2	1		1	2	1	1	

Analysis:

1 FS ratio: **1**, with 2 by NNSs and 2 by NSs (All collocations: 0.40; overall: 0.39)

2 Error ratios: **50%** by NNSs (All collocations: 48.2%; overall: 41.6%) and **50%** by NSs (All collocations: 8.21%; overall: 8.25%).

3 Overlapping rate: 0% (0 out of 2).

4	Due to the limited frequencies, it is unlikely to draw valid conclusions.

Appendix IV.2.21Collocation: Verb + complement of potential (fixed)

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
*对不起**	face PRT rise = sorry	25	21	4		19	19		
*赶得及**	rush PRT reach = still have time					1	1		
*来不及**	come not reach = too late	1	1						
*来得及**	come PRT reach = still have time					1	1		
Total		26	22	4		21	21		

Analysis:

1 FS ratio: **1.24**, with 26 by NNSs and 21 by NSs (All collocations: 0.40; overall: 0.39)

2 Error ratios: **15%** by NNSs (All collocations: 48.2%; overall: 41.6%) and **0%** by NSs (All collocations: 8.21%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate: 96% (25 out of 26).

3.2 Both groups heavily rely on *对不起* (96.2% and 90.5%).

3.3 This type of collocative FSs is one of the three that NNS outnumbers NS production.

4 Conclusions:

4.1.1 Proportionally NNSs used far more FSs of this type. This is due to NNSs' heavy reliance on *对不起* in expressing apologies (see Chapter VI.3.2.2).

4.1.2 NNSs made fewer mistakes than their average level.

4.1.3 Most NNSs FSs are also used by NSs.

* These combinations can be found in Contemporary Chinese Dictionary and do not have spaces between the syllables.

Appendix IV.2.22 Collocation: Verb + complement of result

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
等到 VP	wait until VP	1			1				
定好	fix (appointment) well	1			1				
烦死	annoyed die = deeply annoyed	1	1						
放在心上	put in heart up = bear in mind					2	2		
赶上	rush + manage to do = can meet time line					1	1		
感觉到	feel + with positive result					1	1		
搞到 VP	cause trouble + so that + VP					1		1	
关在门外	lock in + outside the door					1	1		
忽略掉	neglect lose = neglected					1	1		
记错	remember wrongly					2	2		
接到	receive + with positive result					1	1		
记录下来	jot down					2	2		
看错	see wrongly					1	1		
看到	see + with positive result					1	1		
考上	take (a test) + be admitted	1			1				

考完	take (a test) finish = finish taking					1	1		
来到	come arrive = come to					1	1		
念完	read finish					1	1		
认识到	realize + with positive result					1			1
睡过头	sleep overdone = oversleep					3	3		
睡好	sleep well					1	1		
睡觉睡过头	sleep sleep overdone = oversleep					2	1	1	
说出	speak out	1			1				
送到	deliver arrive = send to	2	2						
*听见**	hear + with positive result					1	1		
忘掉	forget lose = forgot					2	1	1	
忘记掉	forget lose = forgot					1	1		
想到	think + with positive result	1			1	1	1		
想起来	remember up = remember suddenly	2	2			4	4		
学到	learn + with positive result					1	1		
约好	fix (appointment) well	2			2				
找到	find + with poistive result	1			1				

准备好	prepare well					1	1		
Total		13	5		8	35	31	3	1

Analysis:

1 FS ratio: **0.37**, with 13 by NNSs and 35 by NSs (All collocations: 0.40; overall: 0.39).

2 Error ratios: **62%** by NNSs (All collocations: 48.2%; overall: 41.6%) and **11%** by NSs (All collocations: 8.21%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate: 23% (3 out of 13).

3.2 NNSs used 10 different verbs and 6 different complements to form 10 different types of FSs, each with 1-2 tokens.

3.3 NSs used 19 different verbs and 14 complements to form 25 types of FSs, each with 1-3 tokens.

4 Conclusions:

4.1.1 Number of NNS FSs is on average.

4.1.2 NNSs made more mistakes than their average level.

4.1.3 NNSs and NSs have very different choices of FSs.

* This combination can be found in Contemporary Chinese Dictionary and does not have space between the two syllables.

Appendix IV.2.23 Collocation: Verb + complement of time

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
熬了几个通宵	endure PRT several unit whole night = didn't sleep for a few days					1	1		
过了 ... 个小时	pass PRT … unit hour = a few minutes past … o'clock	1			1				

想了很长时间	think PRT very long time					1	1		
准备了很长时间	prepare PRT very long time					1	1		
准备很久	prepare + very + long time					1	1		
Total		1			1	4	4		

Analysis:

1 FS ratio: **0.25**, with 1 by NNSs and 4 by NSs (All collocations: 0.40; overall: 0.39)

2 Error ratios: **100%** by NNSs (All collocations: 48.2%; overall: 41.6%) and **0%** by NSs (All collocations: 8.21%; overall: 8.25%).

3 Overlapping rate: 0% (0 out of 1).

4 Due to the limited frequencies, it is difficult to draw valid conclusions.

Appendix IV.2.24 Collocation: Verb + number (+ measure word)

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
有好多好多	have very many very many = there are many many					1	1		
有很多的	have very many PRT = there are many					1	1		
有几个	have several unit	1	1			1	1		
有一些	have some					1	1		
Total		1	1			4	4		

Analysis:

1	FS ratio: **0.25**, with 1 by NNSs and 4 by NSs (All collocations: 0.40; overall: 0.39)
2	Error ratios: **100%** by NNSs (All collocations: 48.2%; overall: 41.6%) and **0%** by NSs (All collocations: 8.21%; overall: 8.25%).
3	Overlapping rate: 100% (1 out of 1).
4	Due to the limited frequencies, it is difficult to draw any conclusions.

Appendix IV.2.25 Collocation: Verb + object (2 syllables)

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs*	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
上课	attend class					2	2		
熬夜	endure + sleepless night					1	1		
帮忙	help with + matter	4	1	2	1				
毕业	finish schoolwork = graduate	1	1			2	2		
出院	discharge hospital	1		1					
打工	do work = work part-time	2	2						
道歉	express apology					2	2		
得病	contract disease	1			1				
堵车	clog traffic	1		1		1	1		
读书	read book = study	1	1						
发烧	have fever	1	1			1	1		
患病	contract disease	1	1						

回国	return homeland	1	1						
回家	return home	1		1					
开车	drive vehicle	1	1						
考试	take test	19	8	8	3	4	2	1	1
来电	come telephone = call me	1			1				
离婚	separate marriage = divorce	1	1						
念书	read book = study					1	1		
排队	line up + queue					1	1		
起床	get off + bed	1		1		2	2		
缺考	miss test					7	7		
认错	admit fault = apologize					1	1		
撒谎	tell lie					2	2		
生病	have sickness	1	1			2	2		
生气	have anger	1			1				
睡觉	have sleep	1			1	7	5	1	1
说话	speak words = speak	6	2		4				
逃课	skip class					1	1		
忘事	forget matter					1	1		
修课	take course					1	1		

有事	have matter = by occupied					1	1		
撞车	bump into + vehicle	1		1					
做事	handle matter	1		1					
作数	be valid, count					1			1
Total		49	21	16	12	41	36	2	3

Analysis:

1 FS ratio: **1.20**, with 49 by NNSs and 41 by NSs (All collocations: 0.40; overall: 0.39)

2 Error ratios: **57%** by NNSs (All collocations: 48.2%; overall: 41.6%) and **12%** by NSs (All collocations: 8.21%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate: 51% (25 out of 49).

3.2 This type of collocative FSs is one of the three that NNS outnumbers NS production.

3.3 FSs of highest frequencies in NNSs' are *考试* (19), *说话* (6) and *帮忙* (4), while in NSs' are *缺考* (7) and *睡觉* (7) and *考试* (4). All others are with 1 - 2 tokens.

3.4 NSSs' FSs are of 22 different types while there are 20 types in NS data.

4 Conclusions:

4.1.1 Proportionally NNSs used far more FSs of this type.

4.1.2 NNSs made more mistakes than their average level.

4.1.3 Half of NNSs FSs are also used by NSs.

4.1.4 Exam-related FSs feature high in both groups (39% in NNS and 31.7% in NS).

* Most combinations can be found in Contemporary Chinese Dictionary and do not have spaces between the pinyin of the syllables.

Appendix IV.2.26 Collocation: Verb + object (3 syllables)

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK

熬通宵	endure + sleepless night					1	1		
编谎言	make lie					1	1		
吃午饭	eat lunch					1			1
出卷子	make + test paper					1			1
处理事	handle matter	1	1						
打电话	make call	2	1	1					
犯错误	make mistake					1	1		
给建议	give advice					1			1
给借口	give excuse	1			1				
给机会	give chance	17	7	10		53	48	4	1
回宿舍	return dormitory					1	1		
讲事情	discuss matter	1	1						
开夜车	burn + midnight oil	1	1						
看医生	see doctor	1		1		1	1		
来电话	come + call = call me	2		1	1				
忙事情	busu with + matter					1		1	
没办法	have not + means = no way out	4	2	1	1	2	2		
没什么	have not + anything	2		2					

说真话	speak true words	1			1				
送医院	send hospital	3	2	1		2	2		
添麻烦	cause trouble	1	1			2	2		
忘了事	forget PRT matter					2	2		
忘事情	forget matter					3	3		
下功夫	put effort					1	1		
下决定	make decision	1			1				
想办法	figure out + solution					1	1		
写声明	write declaration					1	1		
选修课	take course					1	1		
学知识	learn knowledge					1	1		
学中文	learn Chinese	1		1					
有方法	have means					1	1		
有机会	have chance	1		1		1	1		
有考试	have test	5	5						
有情况	have abnormality					1	1		
有时间	have time	1	1						
有事情	have matter	1		1		1	1		
有问题	have problem	1	1						

有原因	have reason					3	3		
有约会	have date	1			1				
约时间	fix appointment	1		1					
找工作	find job	1		1					
找借口	look for + excuses					1	1		
找理由	look for + excuses					2	1		1
找时间	find time					1			1
做访谈	do interview					1	1		
做事情	handle matter					2	2		
做准备	do preparation					1	1		
Total		51	23	22	6	93	82	5	6

Analysis:

1 FS ratio: **0.55**, with 51 by NNSs and 93 by NSs (All collocations: 0.40; overall: 0.39)

2 Error ratios: **55%** by NNSs (All collocations: 48.2%; overall: 41.6%) and **12%** by NSs (All collocations: 8.21%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate: 55% (28 out of 51).

3.2 NSs used 21 different verbs. The two with higest frequencies are *给* (54) and *有* or *没(有)* (9).

3.3 NNSs used 16 different verbs. The two with higest frequencies are *给* (18) and *有* or *没(有)* (17).

3.4 Both groups used only one bisyllabic verb, i.e. the vast majority of verb-object collocations are monosyllabic verb plus bisyllabic object (98% in NNS, 98.9% in NS).

3.5 About 57% of NS FSs are *给机会*, while only 33.3% in NNS are, indicating that NNSs might not have enough proper utterances to ask for one more chance (also see VI.3.1).

3.6 If comparing the disparities between NNSs and NSs in this section and the last two,

it seems appropriate to draw a preliminary conclusion that NNSs can handle 2-syllable verb-object FSs better than 3-syllable FSs, and 3-syllable verb-object FSs better than 4-syllable FSs.

4 Conclusions:

4.1.1 Proportionally NNSs used more FSs of this type.

4.1.2 Both groups made more mistakes than their average level.

4.1.3 About half NNSs FSs are also used by NSs.

4.1.4 In some cases, choice of FSs at word level might inflence the quality of speech acts.

Appendix IV.2.27 Collocation: Verb + object (4 syllables or above)

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
不是问题	not be problem					1	1		
安排补考	arrange + makeup test					8	8		
安排机会	arrange chance					4	4		
安排时间	arrange time	4	2	2		4	3	1	
把握机会	grasp chance					1	1		
表达歉意	express apology					1	1		
不知道怎么办	not know how handle					1	1		
参加补考	sit in + makeup test	3	3			2	1	1	
参加考试	sit in + test	1	1			9	8	1	
处理事情	handle matter	1		1					

出现问题	appear problem					1	1		
错过机会	miss chance					2	2		
错过时间	miss time					2	2		
带来不便	bring inconvenience					1	1		
带来麻烦	bring trouble	1	1						
得到机会	obtain opportunity					1	1		
对不起您	feel sorry to + you					2	2		
发生事件	happen incident					1	1		
发生事情	happen matter					1	1		
复习功课	review schoolwork					1	1		
复习考试	review test					1	1		
改变日期	change date	1			1				
改变时间	change time	1			1				
感到抱歉	feel regret					1	1		
感觉不好意思	feel + not + feel no qualms = sorry					1	1		
检验成果	assess achievement					1	1		
记错日子	remember wrongly + date					1	1		
接到通知	receive notification					1	1		

解释理由	explain reason					1	1		
解释原因	explain reason					1		1	
觉得不舒服	feel + not well	1		1					
考虑事情	consider matter					1	1		
考虑因素	consider factor					1	1		
浪费时间	waste time					1	1		
忙些什么	busy some what = busy with something					1	1		
没有办法	not have + solution = no way out					2	1	1	
没有借口	not have + excuse	4			4				
没有精神	not have + vigore	1	1						
没有损失	not have + loss					1			1
没有问题	not have + problem	1		1					
没有原因	not have + reason					1	1		
弥补错误	remedy fault					1	1		
失去机会	lose chance					1			1
说对不起	say sorry	1	1						
说普通话	speak Putonghua	1		1					
说明情况	explain situation	1		1					

说明原因	explain reason					1	1		
忘了事情	forget PRT matter					6	5	1	
忘记事情	forget matter	1			1				
需要机会	need chance	1	1			1	1		
养成习惯	cultivate habit					1	1		
影响工作	influence work					1	1		
影响学习	influence study					1	1		
有高血压	have + high blood pressure					1	1		
有没有可能	have + not have + possibility = is it possible					1	1		
有没有什么事情	have + not have + what + matter = is there anything					1	1		
有没有事情	have + not have + matter = is there anything					1	1		
有下一次	have next one time					1	1		
造成麻烦	cause trouble					1	1		
知道能力	know ability	1			1				
重视考试	think highly of + test					3	3		

重视学习	think highly of + study	1			1				
Total		*26*	*10*	*7*	*9*	*81*	*73*	*6*	*2*

Analysis:

1 FS ratio: **0.32**, with 26 by NNSs and 81 by NSs (All collocations: 0.40; overall: 0.39)

2 Error ratios: **62%** by NNSs (All collocations: 48.2%; overall: 41.6%) and **10%** by NSs (All collocations: 8.21%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate: 30.1% (8 out of 26).

3.2 NSs used 34 different verbs. The three with higest frequencies are *安排* (16), *参加* (11) and *没有* (7).

3.3 NNSs used 12 different verbs. The three with higest frequencies are *没有* (6), *参加*(4) and *安排*(4).

3.4 Both groups used overwhelmingly bisyllabic verb (100% in NNS, 97.5% in NS).

4 Conclusions:

4.1.1 Proportionally NNSs used fewer FSs of this type.

4.1.2 NNSs made more mistakes than their average level.

4.1.3 Two groups have very different choices of FSs but the ones with highest frequencies are similar.

4.1.4 NNSs rely on smaller varieties of FSs.

Appendix IV.2.28 Collocation: Verb + verb

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
不能控制	not can control					1			1
发现原来是	discover actually is					1	1		
赶回来	rush + come back					1	1		

可以接受	can accept					2	2		
没有办法解释	not have + means + explain = cannot explain					1	1		
应该说	should say	2			2				
有人排队	have + people + line up					1	1		
Total		2	0	0	2	7	6	0	1

Analysis:

1 FS ratio: **0.29**, with 2 by NNSs and 7 by NSs (All collocations: 0.40; overall: 0.39)

2 Error ratios: **100%** by NNSs (All collocations: 48.2%; overall: 41.6%) and **16.7%** by NSs (All collocations: 8.21%; overall: 8.25%).

3 Overlapping rate: 0% (0 out of 2).

4 Due to the limited frequencies, it is difficult to draw any conclusions.

Appendix IV.3.29 Overall distribution of Collocations

COLLOCATIONS	NNS Total	CR (correct)	LK (likely)	ULK (unlikely)	NS Total	CR (correct)	LK (likely)	ULK (unlikely)
Adverb + (adjective/verb) + adverb	6	2	2	2	23	23	0	0
Adverb + verb	22	13	3	6	82	73	0	9
Causative verb + pronoun	12	11	0	1	25	22	0	3
Conjunction + adverb	1	1	0	0	14	14	0	0
Filler	1	1	0	0	27	26	0	1
Juxtaposed noun	2	1	0	1	1	1	0	0

Measure word + noun	44	20	15	9	147	137	7	3
Modifier + noun as head word	29	14	3	12	91	84	3	4
Noun + adverb	5	3	2	0	6	6	0	0
Noun + verb	0	0	0	0	2	2	0	0
Place + direction	1	1	0	0	3	3	0	0
Preposition + noun	0	0	0	0	3	2	1	0
Redundant words	3	1	0	2	8	8	0	0
Repeated words	2	2	0	0	10	10	0	0
Subject + predicate	13	7	3	3	27	27	0	0
Verb + Adverb	0	0	0	0	2	2	0	0
Verb + complement of degree	1	1	0	0	4	4	0	0
Verb + complement of direction	6	6	0	0	30	26	1	3
Verb + complement of movement	7	2	2	3	12	11	0	1
Verb + complement of potential	2	1	0	1	2	1	1	0
Verb + complement of potential (fixed)	26	22	4	0	21	21	0	0
Verb + complement of	13	5	0	8	35	31	3	1

result								
Verb + complement of time	1	0	0	1	4	4	0	0
Verb + number (+ measure word)	1	1	0	0	4	4	0	0
Verb + object (2 syllables)	49	21	16	12	41	36	2	3
Verb + object (3 syllables)	51	23	22	6	93	82	5	6
Verb + object (4 syllables or above)	26	10	7	9	81	73	6	2
Verb + verb	2	0	0	2	7	6	0	1
sub-total (Collocations)	326	169	79	78	805	739	29	37
% of sub-total (Collocations)	100%	51.84%	24.23%	23.93%	100%	91.92%	3.61%	4.60%
% of all non-TSSS FSs	48.7%	25.3%	11.8%	11.7%	47.2%	43.3%	1.7%	2.2%
Total (all non-TSSS FSs)	671	392	127	152	1723	1581	56	86
% of total (overall)	100%	58.4%	18.9%	22.7%	100%	91.8%	3.3%	5.0%

Appendix IV.2.30 Types of Collocations

COLLOCATIONS	NNS Total	NS Total	Total	NNS types	NS types	Common types	Total Types
Adverb + (verb/adjective) + adverb	6	23	29	3	16	2	17
Adverb + verb	22	82	104	17	59	5	71

Conjunction + adverb	1	14	15	1	4	1	4
Causative verb + pronoun	12	25	37	3	8	2	9
Place + direction	1	3	4	1	3	1	3
Filler	1	27	28	1	8	1	8
Juxtaposed noun	2	1	3	2	1	0	3
Measure word + noun	44	147	191	16	38	9	45
Modifier + noun as head word	29	91	120	17	50	4	63
Noun + adverb	5	6	11	2	3	0	5
Noun + verb	0	2	2	0	1	0	1
Preposition + noun	0	3	3	0	2	0	2
Redundant words	3	8	11	3	5	1	7
Repeated words	2	10	12	1	6	1	6
Subject + predicate	13	27	40	13	19	5	27
Verb + Adverb	0	2	2	0	2	0	2
Verb + complement of degree	1	4	5	1	4	0	5
Verb + complement of direction	6	30	36	3	11	2	12
Verb + complement of movement	7	12	19	4	12	3	13
Verb + complement of potential	2	2	4	1	2	0	3

Verb + complement of potential (fixed)	26	21	47	2	3	1	4
Verb + complement of result	13	35	48	10	25	2	33
Verb + complement of time	1	4	5	1	4	0	5
Verb + number (+ measure word)	1	4	5	1	4	1	4
Verb + object (2 syllables)	49	41	90	22	20	7	35
Verb + object (3 syllables)	51	93	144	23	31	7	47
Verb + object (4 syllables or above)	26	81	107	18	50	6	62
Verb + verb	2	7	9	1	6	0	7
sub-total	326	805	1131	167	397	61	503

Appendix IV.3.1 Frame: Adverb + verb frame

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
一个人在 NP	be in NP by oneself					1	1		
一直在 NP	all the way be in NP					1	1		
一直在 VP	all the way VP					3	3		
Total						5	5		

Analysis:

1 FS ratio: **0**, with 0 by NNSs and 3 by NSs (All frames: 0.37; overall: 0.4).

2 Error ratios: **0%** by NSs (All frames: 9.45%; overall: 8.25%).

3 Due to the limited frequencies, it is difficult to draw any valid conclusions.

Appendix IV.3.2 Frame: Adverbial frame

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
比较 Adj 一点儿	comparatively Adj					1	1		
从 NP 到 NP	from NP to NP	1		1					
从 NP 就	began to do it since NP					1	1		
到 NP 才	not until NP					1	1		
对 NP 对 NP 都	both to NP and NP					1			1
可是 NP 却	NP unexpectedly					1	1		
每 NP 都	unexceptionaly all NP	3	3			3	1		2
NP 跟 NP 都	both NP and NP	1			1				
NP 跟 NP 一起	NP together with NP	2			2				
什么 NP 都	all NP without exceptions	4	3		1				
一 VP 之下	as the result of VP					1	1		
一直到 NP 才	as late as until NP					1	1		

再多的 NP 都	even with more NP					1	1		
整个 NP 都	the whole NP without exceptions					1		1	
Total		11	6	1	4	12	8	1	3

Analysis:

1 FS ratio: **0.92**, with 11 by NNSs and 12 by NSs (All frames: 0.37; overall: 0.4).

2 Error ratios: **45%** by NNSs (All frames: 41.84%; overall: 41.6%) and **33.3%** by NSs (All frames: 9.45%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate: 27% (3 out of 11).

3.2 Altogether there are 14 types of FSs of this type. NNSs had production in 5 of them, while NSs had 10. Only 1 type is common for both groups.

3.3 NNSs heavily rely on FSs ending with *都* (8 out of 11).

3.4 NSs' FSs are of 10 different types and distributed more evenly, each with 1 to 3 tokens.

4 Conclusions:

4.1.1 Proportionally NNSs used far more FSs of this type.

4.1.2 NNSs made average mistakes.

4.1.3 Most NNSs FSs are not used by NSs.

4.1.4 NNSs rely on smaller varieties of FSs.

Appendix IV.3.3 Frame: Conjunctive frame

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
不是 VP 是 VP	not mean to VP but because VP					1	1		
不光 VP 而且 VP	not only VP but also VP					1	1		

NP 可以 VP 的话 NP 就会 VP	if NP VP then NP will VP	1		1					
其实 VP 但是 VP	actually VP but VP	1	1			1	1		
如果 VP 的话就 VP	if VP then VP	1	1						
如果 VP 的话那就 VP	if VP then VP					1	1		
如果 VP 就 VP	if VP then VP	1		1		1	1		
如果 VP 那就 VP 了	if VP then VP	1		1					
如果 VP 那 VP	if VP then VP					1		1	
首先 NP VP 其次 NP VP	firstly NP VP secondly NP VP					1	1		
虽然 VP 但是 VP	although VP but VP					2	2		
所以 VP 因为 VP	therefore VP because VP	2	2			1	1		
特地 VP 竟然 VP	specially VP but unexpectedly VP					1	1		
我知道 VP 不过 VP	I know VP but VP	1		1					
我知道 VP 但 VP	I know VP but VP					1	1		

我知道 VP 可是 VP	I know VP but VP	1			1				
应该 VP 而不是 VP	should VP and not VP					1	1		
因 故 VP 所 以 VP	somehow VP therefore VP					1		1	
因为 VP 就 VP	because VP then VP					1	1		
因为 VP 就知道 VP	because VP so have to VP					1	1		
因为 VP 所以呢 VP	because VP therefore VP	1			1				
因为 VP 所以说 VP	because VP therefore VP					1			1
因 为 VP 所 以 VP	because VP therefore VP	10	10			13	12	0	1
因为 VP 又因为 VP	because VP and also because VP					1	1		
由于 (NP) VP 所以 (NP) VP	because (NP) VP therefore (NP) VP					3	2	1	
只是 VP 就 VP	just because VP then VP					4			4
Total		20	14	4	2	38	29	3	6

Analysis:

1 FS ratio: **0.53**, with 20 by NNSs and 38 by NSs (All frames: 0.37; overall: 0.4).

2 Error ratios: **30%** by NNSs (All frames: 41.84%; overall: 41.6%) and **24%** by NSs (All frames: 9.45%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate: 70% (14 out of 20).

3.2 There are 26 types of FSs of this category. NNSs have 10 and NSs have 20. Only 4 types are common.

3.3 The frame with highest frequency is *因为 VP 所以 VP*, accounted for half of NNS and one third of NS production. Distributions of other FSs are quite even.

4 Conclusions:

4.1.1 Proportionally NNSs used more FSs of this type.

4.1.2 NNSs made fewer mistakes than their average level.

4.1.3 Most NNSs FSs are also used by NSs.

4.1.4 NNSs rely on smaller varieties of FSs.

Appendix IV.3.4 Frame: Noun frame

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
NP 的时间	time for NP	1			1				
NP 等等的东西	things like NP etc.					1	1		
NP 和 NP 的决赛	final match between NP and NP					1	1		
Nu 点多钟	a few minutes past Nu.					1	1		
VP 的程度	degree of VP					1	1		
VP 的次数	frequency of VP					1	1		
VP 的错误	fault of VP					1	1		
VP 的东西	matter about VP					1			1

VP 的过程	process of VP					1			1
VP 的建议	suggestion that (you) VP					1	1		
VP 的机会	opportunity to VP	1		1		16	15	1	
VP 的经验	experience of VP					1			1
VP 的考试	test of VP					1			1
VP 的事	matter about VP	1			1				
VP 的时候	time when VP					1			1
VP 的时间	time for VP	2	1	1					
VP 的时期	period for VP	1			1				
VP 的事情	matter about VP	1			1	6	6		
VP 的通知	notification of VP					1	1		
VP 的知识	knowledge that VP					1	1		
VP 的状况	condition of VP					1	1		
VP 的状态	state of VP					1			1
VP 机会	opportunity to VP					1	1		
VP 时间	time to VP					2	2		
VP 这样的问题	problems such as VP					1	1		
Total		7	1	2	4	42	35	1	6
Analysis:									

1 FS ratio: **0.17**, with 7 by NNSs and 42 by NSs (All frames: 0.37; overall: 0.4).

2 Error ratios: **57%** by NNSs (All frames: 41.84%; overall: 41.6%) and **17%** by NSs (All frames: 9.45%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate: 29% (2 out of 7).

3.2 Vast majority of FSs begin with a VP, followed by a PRT *的*, which is optioned and omitted by 3 tokens by NSs.

3.3 There are 25 types identified (6 types used by NNSs, 21 used by NSs).

3.4 Distributions are very even, except for *VP 的机会* (16 tokens) and *VP 的事情* (6 tokens) in NS data.

4 Conclusions:

4.1.1 Proportionally NNSs used fewer FSs of this type.

4.1.2 NNSs made more mistakes than their average level.

4.1.3 Most NNSs FSs are not used by NSs.

4.1.4 This kind of FSs is of great importance but seemingly not well mastered by NNSs.

Appendix IV.3.5 Frame: Particle + particle frame

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
VP 啊 VP 啊 VP	VP and VP and VP					1	1	0	0
Total						1	1		

Analysis:

1 FS ratio: **0**, with 0 by NNSs and 1 by NSs (All frames: 0.37; overall: 0.4).

2 Error ratios: **0%** by NSs (All frames: 9.45%; overall: 8.25%).

3 Due to the limited frequencies, it is difficult to draw any conclusions.

Appendix IV.3.6 Frame: Position and direction frame

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
从 NP 中	from the midst of NP					1			1
在 NP 里	inside NP	1	1						
在 NP 里面	inside NP					2	2		
在 NP 面前	in front of NP					1			1
在 NP 上	in the scope of NP					4	4		
在 NP 外	outside NP					1	1		
Total		1	1	0	0	9	7	0	2

Analysis:

1 FS ratio: **0.11**, with 1 by NNSs and 9 by NSs (All frames: 0.37; overall: 0.4).

2 Error ratios: **0%** by NNSs (All frames: 41.84%; overall: 41.6%) and **22%** by NSs (All frames: 9.45%; overall: 8.25%).

3 Overlapping rate: 0% (0 out of 1).

4 Among the 6 types identified, 5 begin with *在* and 1 with *从*.

5 While difficult to draw valid conclusions due to the limited frequency in NNS, 9 tokens in NS data shows that this type of considerable significance.

Appendix IV.3.7 Frame: Preposition + verb frame

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
NP 把 NP 错过了	NP missed NP					1	1		

NP 把NP 定了	NP fixed NP					1	1		
NP 把 NP 给错过了	NP missed NP					1	1		
NP 把 NP 给忽略了	NP neglected NP					1	1		
NP 把 NP 给记错了	NP remember NP mistakenly					1	1		
NP 把NP 给忘了	NP forgot about NP					4	4		
NP 把NP 给忘记掉了	NP forgot about NP					1	1		
NP 把 NP 给忘记了	NP forgot about NP					2	2		
NP 把 NP 解释一下	NP explained NP a bit					1	1		
NP 把 NP 记录下来	NP jotted down NP					1	1		
NP 把NP 送到NP	NP sent NP to NP	2	2						
NP 把 NP 忘得一干二净	NP completely forgets about NP					1	1		
NP 把NP 忘了	NP forgets about NP					4	4		
NP 被关在 NP 了	NP was locked					1	1		

	in NP								
NP 比 NP 少 Nu M	NP is less than NP by Nu M					1	1		
NP 从 NP 中醒来	NP woke up from NP					1			1
NP 对 NP 抱歉	NP felt sorry for NP	1	1						
NP 对 NP 不重视	NP does not think highly of NP					2	2		
NP 对 NP 不尊重	NP does not respect NP					1	1		
NP 对 NP 带来 NP	NP brings NP to NP					1	1		
NP 对 NP 来讲重要	NP is important to NP					1	1		
NP 对 NP 来说重要	NP is important to NP					2	2		
NP 对 NP 没有损失	NP does not cause any loss to NP					1			1
NP 对 NP 说	NP tells NP that					3	3		
NP 对 NP 说一句	NP says something to NP					1	1		

NP 对 NP 重视	NP thinks highly of NP					1	1		
NP 对 NP 重要	NP is important to NP	1		1		6	6		
NP 对 VP 表示VP	NP shows VP to VP					1	1		
NP 给 NP 安排	NP arranged for NP	1	1			6	6		
NP 给 NP 打电话	NP gave NP a call	1	1						
NP 给 NP 带来麻烦	NP brought troubles to NP	1	1						
NP 给 NP 添麻烦	NP made trouble to NP	1	1			2	2		
NP 给 NP 造成麻烦	NP made trouble to NP					1	1		
NP 跟NP 吵架	NP quarrelled with NP	1			1				
NP 跟NP 讲	NP said to NP	1	1						
NP 跟NP 解释	NP explained to NP					2	2		
NP 跟NP 撒谎	NP lied to NP					1	1		
NP 跟NP 说	NP said to NP	1	1			2	1		1
NP 跟NP 说话	NP talked to NP	4	2		2				

NP 跟NP 一样	NP is the same as NP	1		1					
NP 跟NP 约	NP made appointment with NP	1			1				
NP 为 NP 安排	NP arranged for NP					3	3		
NP 为 NP 准备	NP prepared for NP					1	1		
NP 为了 NP 准备	NP prepared for NP					2	1	1	
NP 为了 NP 做准备	NP prepared for NP					1	1		
NP 向 NP 保证	NP vows to NP					1	1		
NP 向 NP 表达 NP	NP expresses NP to NP					1	1		
NP 向 NP 道歉	NP apologizes to NP					1	1		
NP 向 NP 认错	NP apologizes to NP					1	1		
NP 向 NP 学	NP learns from NP					1	1		
Total		17	11	2	4	68	64	1	3

Analysis:

1 FS ratio: **0.25**, with 17 by NNSs and 68 by NSs (All frames: 0.37; overall: 0.4).

2 Error ratios: **35%** by NNSs (All frames: 41.84%; overall: 41.6%) and **5.8%** by NSs (All frames: 9.45%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate: 23.5% (4 out of 17).

3.2 Among the 50 types identified (13 types used by NNSs, 41 used by NSs), only 4 are used by both.

3.3 The prepositions used in the FSs are *把* (13 types), *对* (12), *跟* (8), *给* (5), *向* (5), *为(了)* (4), *被* (1), *比* (1) and *从* (1). While NS used all prepositions, NNSs only used 6 types with *跟*, 4 types with *给*, 2 types with *对*, and only 1 type with *把*. The greatest disparities lie in *把* [190] and *向* and *为(了)*.

3.4 FSs with highest frequencies are *NP 给 NP 安排* (6) and *NP 对 NP 重要* (6) in NS and *NP 跟 NP 说话* in NNS (4).

4 Conclusions:

4.1.1 Proportionally NNSs used much fewer FSs of this type.

4.1.2 NNSs made fewer mistakes than their average level.

4.1.3 Most NNSs FSs are not used by NSs.

4.1.4 NNSs rely on FSs using limited varieties of prepositions.

Appendix IV.3.8 Frame: Sentence core frame

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
把 NP 给 VP 掉	VP NP (with negative result)					1	1		
把 NP 给 VP 掉了	VP NP (with negative result)					1	1		
把 NP 给 VP 了	VP NP (with negative result)					8	8		
把 NP VP 了	VP NP	2	2			15	15		

[190] *把* construction has been reported the most challenging for NNSs even though it has been introduced and sufficiently practiced in elementary levels.

把 NP 再 VP 一下	again VP NP a bit					1	1		
被 VP 了	to be VP					1	1		
不 Adj 了	not Adj any more	2			2				
不会 VP 的	surely will not VP					2	2		
不会 VP 了	will not VP any more					1	1		
不会再 VP 了	will not VP any more					4	3		1
不是故意 VP 的	did not VP deliberately					1	1		
不是 NP 的错	not NP's fault	1	1						
不是 VP 的	not meaned to VP	2	2			7	7		
不是有意 VP 的	did not VP deliberately					1		1	
不 VP 了	not VP any more					1			1
不想 VP 了	do not want to VP any more					1			1
不用再 VP 了	no need to VP again					1	1		
当然 VP 吧	surely will VP	1			1				
都是 NP 的错	all are NP's faults	1	1						
刚刚 VP 了	just finished VP	1			1				
刚刚 VP 完	just finished VP					1	1		
给您 VP 了	sorry for making you VP					1	1		

给 VP 掉	VP (with negative result)					1	1		
给 VP 掉了	VP (with negative result)					1	1		
给 VP 了	VP (with negative result)	1	1			13	13	0	0
很早就 VP 了	VP long ago					1	1		
会 VP 的	surely will VP					5	5		
忽然 VP 了	suddenly VP					1	1		
竟然 VP 错	unexpectedly VP mistakenly					1	1		
竟然 VP 掉	unexpectedly VP (with nagetive result)					1	1		
就马上 VP 了	immediately VP then	1	1						
就 VP 了	then VP	1	1			2	2		
就 VP 嘛	then just VP, you know	1			1				
可以 VP 了	can VP then					1	1		
快要 VP 了	will soon VP	1	1						
另外 VP 一个 NP	VP another NP	1	1						
麻烦 NP 了	really need NP's help					2	2		
没有什么 NP	there was no any NP					1	1		
没有 VP 什么 NP	did not VP any NP					1			1

那就 VP 了	then VP	1		1		1	1		
什么 NP 都可以	any NP will do	1	1						
什么 NP 都没有	no NP of any kind	1			1				
是 Adj 的	is really Adj					2	1		1
是 Adj 的状态	in a Adj state of mind					1			1
是 keyi VP 的	actually can VP					1	1		
是 NP 吧	is NP I supposed					1			1
是 NP 的	is like NP					1	1		
是 NP 的不对	is NP's fault					1	1		
是 NP 的错	is NP's fault	4	3	1		1	1		
是 NP 的错误	is NP's fault					1	1		
是 VP 的	actually VP					5	5		
是 VP 的好机会	is a great opportunity to VP					1	1		
太 Adj 了	too Adj	2	2			5	5		
太 VP 了	VP excessively					2	2		
特别 VP 一次	specially VP once					2	2		
突然 VP 了	suddenly VP	1			1	2	2		
突然间 VP 了	suddenly VP					1	1		
为什么会 VP	why would VP					1			1

要VP 了	really need to VP					2	2		
已经 VP 过了	already VP					2		2	
已经 VP 了	already VP	3	2	1		17	14	2	1
一直VP 着 NP	all the way VP NP	0	0	0	0	1	1		
应该 VP 的	should have VP					1	1		
有 VP 一次 NP	once again VP NP					1			1
再 VP 第二次	VP again for the second time					2	2		
再VP 个NP	VP another NP					1	1		
再 VP 一次	VP one more time	2	2			36	36		
再 VP 一个NP	VP another NP	5	4	1		11	10	1	
再 VP 一下	VP a bit more					2	2		
真 Adj 了	really so Adj	1	1						
只能说 VP 了	can only request to VP					1	1		
最后再VP 一下	VP one more time as a final check					1	1		
Total		37	26	4	7	186	170	6	10

Analysis:

1 FS ratio: **0.21**, with 35 by NNSs and 170 by NSs (All frames: 0.37; overall: 0.4).

2 Error ratios: **31%** by NNSs (All frames: 41.84%; overall: 41.6%) and **9.4%** by NSs (All frames: 9.45%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate: 62.8% (22 out of 35).

3.2 Among the 72 types identified (23 types used by NNSs, 60 by NSs), only 11 are used by both.

3.3 FSs with highest frequencies in NS are *再 VP 一次* (36), *已经 VP 了* (17), *给 VP 了* (13) and *再 VP 一个 NP* (11). The top three in NNS data are *再 VP 一个 NP* (5), *是 NP 的 错* (4) and *已经 VP 了*(3).

3.4 The greatest disparities in FSs lie in *再 VP 一次*(2 vs. 36) and *给 VP 了*(1 vs. 13), and those beginning with *把* (0 vs 11). It seems that 把 constructions are indeed very challenging for them (also see IV.3.6).

4 Conclusions:

4.1.1 Proportionally NNSs used fewer FSs of this type.

4.1.2 NNSs made fewer mistakes than their average level.

4.1.3 Most NNSs FSs are also used by NSs.

4.1.4 NNSs are weak in some important FSs.

Appendix IV.3.9 Frame: Sentence crown frame

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
不管怎么样	no matter what happened					1	1		
不知道为什么	I don't know why but					3	3		
从 NP 来说	as far as NP is concerned					1	1		
等 VP 之后	after VP					1	1		
对 NP 来讲	as far as NP is concerned					5	5		
对于 NP 来说	as far as NP is concerned	1	1			1	1		

对于这个问题	as far as this matter is concerned					1	1		
就像 NP 那样	just like NP					1	1		
那个时候	at that moment	1			1				
您记得	you remember	1	1						
您说吧	just tell me	1			1				
您也知道	you also know that					2	2		
您知道	you know	2	1		1	5	3		2
请问	may I know	1	1						
求求您	please					1	1		
如果可能的话	if possible					1	1		
如果可以的话	if possible					1	1		
如果您能 VP	if you can VP					1	1		
如果您 VP 呢	if you VP					1			1
如果 VP 的话	if VP	4	4			3	3		
谁知道	to my dismay					1	1		
所以说	therefore	1	1						
VP 的路上	on the way to VP					1	1		
VP 的时候	in the process of VP	11	3	1	7	7	6		1

VP 了以后	after finishing VP	2	2						
VP 完 NP 以后	after finishing VP NP					1	1		
我不知道为什么	I don't know why but					1	1		
我跟您说了	I have mentioned to you earlier					1	1		
我觉得	I think	1			1	14	13		1
我求求您	please	1	1						
我认为	I think	1			1				
我想	I think	1			1	7	6		1
我想说	I want to say that					1	1		
我想问您	I want to ask you	1	1						
我也觉得	wo also feel that					1	1		
我自己觉得	I feel that					1	1		
无论如何	no matter what happened					6	6		
因为 Adj 的原因	because of Adj reason					1	1		
由于一些 NP	because of some NP					1	1		
在这一点上	at this point					1	1		
准备 VP 的时	when about to VP	1	1						

候									
Total		31	17	1	13	74	68		6

Analysis:

1 FS ratio: **0.42**, with 31 by NNSs and 74 by NSs (All frames: 0.37; overall: 0.4).

2 Error ratios: **45%** by NNSs (All frames: 41.84%; overall: 41.6%) and **8.1%** by NSs (All frames: 9.45%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate: 64.5% (20 out of 31).

3.2 Among the 41 types identified (16 types used by NNSs, 31 by NSs), only 6 types are used by both.

3.3 FSs with highest frequencies in NS are *我觉得* (14), *VP 的时候* (7), *我想* (7) and *无论如何* (6). The top four in NNS data are *VP 的时候* (11), *如果 VP 的话* (4), *VP 了以后* (2) and *您知道* (2). Only one type is common.

3.4 Regretfully *VP 的时候*, the top frequency FS in NNS, has a high error rate of 8 out of 11.

3.5 The greatest disparities lie in *我觉得* (1 vs. 14), *我想* (1 vs. 7) and *无论如何* (0 vs. 6).

4 Conclusions:

4.1.1 Proportionally NNSs used slightly more FSs of this type.

4.1.2 NNSs made slightly more mistakes than their average level.

4.1.3 Most NNSs FSs are also used by NSs.

4.1.4 NNSs are weak on most important FSs used by NSs.

Appendix IV.3.10 Frame: Sentence head frame

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
拜托您 VP	beg you to VP					1	1		
不知道可不可以 VP	don't know if it's possible to VP					1	1		

不知道能不能 VP	don't know if it’s possible to VP					1	1		
不知道 NP VP	don't know if NP VP	1		1					
错误在于 NP	it is NP's fault					1	1		
看 您 能 不 能 VP	would like to see if you can VP					1	1		
可不可以 VP	is it possible to VP	8	7	1		4	4		
可不可以 VP 呢	is it possible to VP					1	1		
可能 NP 会觉得 NP VP	maybe NP will think NP VP					1	1		
恳请您 VP	beg you to VP					1	1		
可不可以 VP	is it possible to VP	1	1						
没 有 人 可 以 VP	nobody can VP	1			1				
能不能 VP	is it possible to VP	1	1			2	1		1
能不能 VP 呢	is it possible to VP					3	3		
您会不会 VP	will you VP or not	1	1						
您 看 能 不 能 VP	do you think it is possible to VP					1	1		
您 可 不 可 以 VP	is it possible that you VP	1		1		1	1		

您可以 VP 吗	is it possible that you VP	1			1	1	1		
您要 VP	please VP					1	1		
请老师 VP	please VP, teacher	2	2						
请您 VP	please VP	14	11	3		6	6		
请您 VP 啊	please VP	1			1				
请您 VP 吧	please VP					1	1		
请 您 一 定 要 VP	please please VP					2	2		
请 VP 吧	please VP					1	1		
请求您 VP	please VP					3	3		
其 实 是 因 为 NP VP	actually it is because NP VP					1	1		
是不是能 VP	is it possible to VP					1	1		
是否可以 VP	is it possible to VP	1	1						
是否可以 VP 呢	is it possible to VP	1	1						
首先我得 VP	firstly I need to VP					1	1		
首先我要 VP	firstly I need to VP					1	1		
我必须VP	I have to VP					1	1		
我不会 VP	I will not VP	3		3					

我不可能 VP	I can't possible VP	2			2				
我不想 VP	I don't want to VP					2	2		
我不希望 VP	I don't want to VP					1			1
我不要 VP	I don't want to VP	1			1				
我不应该 VP	I should not VP	1			1				
我不是 NP 嘛	I am not NP, you know	1			1				
我不知道 NP VP	I don't know NP VP					2	2		
我不知道您是不是 VP	I don’t know if you VP					1	1		
我得 VP	I need to VP	1	1						
我非常希望 VP	I really want to VP	1	1						
我还要 VP	I still want to VP	1			1				
我还是想 VP	I still want to VP					1	1		
我还是希望 VP	I still want to VP					1	1		
我还是要 VP	I still want to VP					1	1		
我很想 VP	I really want to VP	1	1						
我很希望 VP	I really want to VP					1	1		
我就会 VP	then I will VP	1		1					

我就是不知道 NP VP	I simply don't know NP VP					1	1		
我就是想 VP	I just want to VP					1	1		
我觉得 VP	I feel VP	1	1			1	1		
我 可 不 可 以 VP	is it possible that I VP	1	1						
我可以 VP	I can VP	2			2				
我能不能 VP	is it possible that I VP					1	1		
我能不能 VP 呢	is it possible that I VP					1	1		
我请您 VP	I beg you to VP					1	1		
我求您 VP	I beg you to VP	1	1			1	1		
我确实 VP	I indeed VP					1	1		
我想 VP	I want to VP	5	2	3		4	4		
我 相 信 NP VP	I believe NP VP					1	1		
我希望能 VP	I hope (I) can VP					1	1		
我希望您可以 VP	I hope you can VP					1	1		
我 希 望 您 能 VP	I hope you can VP					3	3		
我希望您能够	I hope you can VP					1	1		

VP									
我希望您 VP	I hope you will VP					1	1		
我希望 VP	I hope NP VP					1	1		
我需要 VP	I need to VP	2	1	1					
我要VP	I want to VP	3	1		2	1	1		
我要先 VP	firstly I want to VP	1			1				
我也不能VP	neither can I VP					1			1
我 也 不 知 道 NP VP	neither do I know NP VP					2	2		
我也不知道为什么 NP VP	neither do I know why NP VP					1	1		
我也没有想到 NP VP	neither did I expect NP VP					1	1		
我也想 VP	I also want to VP					1	1		
我 也 希 望 您 VP	I also hope you can VP					1	1		
我也应该 VP	I also should VP					1	1		
我一定 VP	I definitely will VP	4	3		1	9	9		
我一定 VP 吧	I most likely will VP	1			1				
我应该 VP	I should VP	2	1		1	2	2		
我 知 道 NP	I know NP VP	5	4		1	11	11		

VP									
我只是想 VP	I just want to VP					1	1		
希望能够 VP	hope (NP) can VP					1	1		
希望您会 VP	hope you will VP	1	1						
希 望 您 可 以 VP	hope you can VP					1	1		
希望您能 VP	hope you can VP					8	8		
希 望 您 能 够 VP	hope you can VP					2	2		
希望您 VP	hope you can VP					1	1		
需不需要 VP	is it a must to VP	1			1				
原因就是 NP VP	the reason is that NP VP					1	1		
责任在于 NP	NP should be responisible for this					1	1		
Total		*77*	*44*	*14*	*19*	*117*	*114*	*0*	*3*

Analysis:

1 FS ratio: **0.66**, with 77 by NNSs and 117 by NSs (All frames: 0.37; overall: 0.4).

2 Error ratios: **42.8%** by NNSs (All frames: 41.84%; overall: 41.6%) and **2.5%** by NSs (All frames: 9.45%; overall: 8.25%).

3 average.

4 Other findings:

4.1 Overlapping rate: 59.7% (46 out of 77).

4.2 Among the 93 types identified (37 types used by NNSs, 68 by NSs), only 12 types are used by both.

4.3 FSs with highest frequencies in NS are *我知道 NP VP* (11), *我一定 VP* (9), *希望您能 VP* (8) and *请您 VP* (6). The top four in NNS data are *请您 VP* (14), *可不可以 VP* (8), *我想 VP* (5) and *我知道 NP VP* (5). Two types are common.

4.4 The greatest disparities in FSs lie in *希望 (您) 能(够) VP* (0 vs. 11).

5 Conclusions:

5.1.1 Proportionally NNSs used more FSs of this type.

5.1.2 Most NNSs FSs are also used by NSs.

Appendix IV.3.11 Frame: Sentence tag frame

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
可以吗	is it okay	2	2						
行不行	is it okay	1			1				
Total		3	2		1				

Analysis:

1 FS ratio: **infinitive**, with 3 by NNSs and 0 by NSs (All frames: 0.37; overall: 0.4).

2 Error ratios: **33.3%** by NNSs (All frames: 41.84%; overall: 41.6%).

3 Due to the limited frequencies, it is difficult to draw any valid conclusions.

Appendix IV.3.12 Frame: Verb + object frame

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK

安排 VP 的机会	arrange a chance to VP					4	4		
安排 VP 的时间	arrange a time to VP	1	1						
赶上 VP 的时间	catch the time to VP					1	1		
给机会 VP	give opportunities to VP					1	1		
给 VP 的 机会	give opportunities to VP	1		1		9	8	1	
会说 NP 话	can speak NP language	1		1					
接到 VP 的通知	receive a notification to VP					1	1		
没办法 VP	have no means to VP	2	1		1				
没有办法 VP	have no means to VP	1	1			1		1	
没有什么 NP	have not any NP	1			1				
没有 VP 的经验	have no experience to VP					1			1
NP 帮 NP 忙	NP gives NP a hand	2	1		1				
NP 给 NP 建议	NP gives NP an advice					1	1		
NP 给 NP 借口	NP gives NP an excuse	1			1				
NP 给 NP 机会	NP gives NP a chance	15	5	10		51	47	4	
是 NP 的失误	is NP's fault					1	1		
是 NP 的问题	is NP's problem	1	1						
是 VP 的时期	is the period to VP	1			1				
忘记 VP 的事情	forget things about VP					2	2		

想办法 VP	figure out a way to VP					1	1		
需要 VP 的机会	need opportunities to VP					1	1		
养成 Adj 的习惯	cultivate an Adj habit					1	1		
影响将来的 NP	influence NP in the future					1	1		
有 Adj 的原因	have Adj reason					1	1		
有机会 VP	have chances to VP					1	1		
有时间 VP	have time to VP	1	1						
有事情 VP	have something to VP					1	1		
有 VP 的时候	have a time to VP					1			1
找 Adj 的理由	find an Adj excuse					1	1		
找理由 VP	find an excuse to VP					1			1
Total		28	11	12	5	83	74	6	3

Analysis:

1 FS ratio: **0.34**, with 28 by NNSs and 83 by NSs (All frames: 0.37; overall: 0.4).

2 Error ratios: **60.7%** by NNSs (All frames: 41.84%; overall: 41.6%) and **10.8%** by NSs (All frames: 9.45%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate: 60.7% (17 out of 28).

3.2 Among the 30 types identified (12 types used by NNSs, 21 by NSs), 9 types are used by both.

3.3 FSs with highest frequencies in NS are *NP 给 NP 机会* (51), *给 VP 的机会* (9) and *安排 VP 的机会* (4). The top three in NNS data are *NP 给 NP 机会* (15), *NP 帮 NP 忙* (2) and *没办法 VP* (2). Only the top type is common.

3.4 The top three in NS (totally 64) are actually synonymous FSs. They are all the indispensable components of 'pleading for another chance' speech act. On the other hand, NNSs only have 16 token of such FSs.

4 Conclusions:

4.1.1 Proportionally NNSs used average number of FSs.

4.1.2 NNSs made far more mistakes than their average level.

4.1.3 Most NNSs FSs are also used by NSs.

4.1.4 NNSs obviously did not produce sufficient amount of FSs for ceitain speech act.

Appendix IV.3.13 Frame: Verb + verb frame

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
NP 听见 NP 响	NP heard NP making a sound					1	1		
等着 NP 去做	waiting for NP to do					1	1		
赶回 NP 来	rush and come back to NP					1	1		
给 机 会 参 加 NP	give opportunities to participate NP					1		1	
给机会来 VP	give opportunities to VP					1	1		
给 NP 机会来 VP	give NP opportunities to VP					2	1	1	
记得要 VP	remember (that you) need to VP	1	1						
NP 帮 NP 安排	NP help NP arrange					1	1		
NP 帮 NP 安排时间	NP help NP arrange a time					1	1		
NP 帮 NP 照顾	NP help NP take care of NP	1	1						

NP									
NP 帮助 NP 学习	NP help NP study	1			1				
NP 带NP 去NP	NP take NP to (go to) NP					1	1		
NP 到 NP 去	NP go to NP (go)	1	1			1	1		
NP 陪 NP 等	NP accompany NP to go to NP					2	2		
NP 陪NP 去NP	NP accompany NP to wait					1	1		
NP 送 NP 到NP	NP escort NP to NP					1	1		
NP 送NP 到NP 去	NP escort NP to NP					1	1		
NP 送NP 回NP	NP escort NP to return to NP					1	1		
NP 送NP 去NP	NP escort NP to NP					1	1		
送到NP 去	deliver to NP (go)	2	2						
VP 来想一想	VP to think about					1	1		
有 NP 等着 NP 去 VP	there is NP waiting for NP to VP					1	1		
有NP 发生	there is NP happening					2	2		
有NP 需要VP	there is NP needing VP	1	1						
Total		7	6		1	22	20	2	

Analysis:

1 FS ratio: **0.32**, with 7 by NNSs and 22 by NSs (All frames: 0.37; overall: 0.4).

2 Error ratios: **14%** by NNSs (All frames: 41.84%; overall: 41.6%) and **9%** by NSs (All

frames: 9.45%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate: 14% (1 out of 7).

3.2 Among the 24 types identified (6 types used by NNSs, 19 by NSs), only 1 type is used by both.

3.3 Distributions in both NNS and NS data are very even, only with 1 – 2 tokens each.

3.4 In this kind of FSs, each contains 2 to 3 verbs. The first verbs with higher frequencies are *送* (5 tokens), *帮 (助)* (4), *给* (3), *有* (3) and *陪* (2) etc.

3.5 The second verbs are *去* (6), *到* (3), *来* (3), *回* (2), *等* (2) and *安排* (2) etc. Among them, 2 *去* and 2 *来* are used between two VPs to indicate that the former is the means and the latter is the purpose (*有 NP 等着 NP 去 VP, 等着 NP 去做, 给机会 来 VP, 给 NP 机会 来 VP*).

3.6 Other than *到* and *等*, verbs used in the first place are different from those in the second place.

4 Conclusions:

4.1.1 Proportionally NNSs used average number of FSs.

4.1.2 NNSs made fewer mistakes than their average level.

4.1.3 Most NNSs FSs are not used by NSs.

Appendix IV.3.14 Overall distribution of Frames

FRAMES	NNS Total	CR (correct)	LK (likely)	ULK (unlikely)	NS Total	CR (correct)	LK (likely)	ULK (unlikely)
Adverb + verb	0	0	0	0	5	5	0	0
Adverbial	11	6	1	4	12	8	1	3
Conjunction	20	14	4	2	38	29	3	6
Noun	7	1	2	4	42	35	1	6
Position and direction	1	1	0	0	9	7	0	2
Particle + particle	0	0	0	0	1	1	0	0
Preposition + verb	17	11	2	4	68	64	1	3

Sentence core	37	26	4	7	186	170	6	10
Sentence crown	31	17	1	13	74	68	0	6
Sentence head	77	44	14	19	117	114	0	3
Sentence tag	3	2	0	1	0	0	0	0
Verb + object	28	11	12	5	83	74	6	3
Verb + verb	7	6	0	1	22	20	2	0
sub-total	239	139	40	60	657	595	20	42
% of sub-total	100%	58.16%	16.88%	25.32%	100%	90.56%	3.04%	6.39%
% of total (frame)	35.7%	20.8%	6.0%	9.0%	38.5%	34.9%	1.2%	2.5%
total	11	6	1	4	12	8	1	3
% of total (overall)	0	0	0	0	5	5	0	0

Appendix IV.3.15 Types of Frames

FRAMES	NNS Total	NS Total	Total	NNS types	NS types	Common types	Total Types
Adverb	11	12	23	5	10	1	14
Adverb + verb	0	5	5	0	3	0	3
Conjunction	20	38	58	10	20	4	26
Noun	7	42	49	6	21	2	25
Particle + particle	0	1	1	0	1	0	1

Position and direction	1	9	10	1	5	0	6
Preposition + verb	17	68	85	13	41	4	50
Sentence core	37	186	222	23	60	11	72
Sentence crown	31	74	105	16	31	6	41
Sentence head	77	117	194	37	68	12	93
Sentence tag	3	0	3	2	0	0	2
Verb + object	28	83	111	12	21	3	30
Verb + verb	7	22	29	6	19	1	24
sub-total	239	657	895	131	300	44	387

Appendix IV.4.1 Polywords: Adjectives

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
不好意思	not + feel no qualms = sorry	6	6			19	19		
不舒服	not well = feel sick	2	1	1					
不太舒服	not very well = feel sick					1	1		
各种各样	various type various shape = all kinds of					1	1		

糊里糊涂	muddled	2			2				
绝佳	extremely good					1		1	
无意识	have not + consciousness					1			1
一干二净	extremely clean					1	1		
有礼貌	have manner = polite	1			1				
有意义	have meaning = meaningful					1	1		
Total		11	7	1	3	24	22	1	1

Analysis:

1 FS ratio: **0.46**, with 11 by NNSs and 24 by NSs (All polywords: 0.41; overall: 0.4).

2 Error ratios: **36.3%** by NNSs (All polywords: 20.75%; overall: 41.6%) and **8.3%** by NSs (All polywords: 5.34%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate: 54.5% (6 out of 11).

3.2 Among the 9 types identified (4 types used by NNSs, 6 by NSs), only 1 type is used by both (the one with highest frequencies in both groups: *不好意思*).

3.3 Besides *不好意思*, distributions of other FSs in both NNS and NS data are very even, only with 1 – 2 tokens each.

3.4 As mentioned in IV.2.21, when expressing apologies, NNSs used many *对不起* and few others, including *不好意思*, which is of the same abundance as *对不起* in NS data.

3.5 Besides the errorless *不好意思*, other FSs of this type have high error rates in both groups (80% in NNS and 40% in NS).

3.6 In terms of structure, FSs of this type are either degreed form of certain adjectives[191] (*糊里糊涂* derived from *糊涂; 一干二净* derived from *干净*), or composed of an adverb and an adjective/verb (*绝佳, 不舒服, 不太舒服, 不好意*

[191] As 'rock-hard' and 'red-faced' in English, this kind of adjectives in Chinese cannot be modified by adverbs because semantically they already have an inbuilt degree or modifier. For example, while *白* means 'white', *雪白* means 'snow white' (as white as snow); while *糊涂* means 'muddled', *糊里糊涂* means 'so muddled'; while *干净* means 'clean', *一干二净* means 'absolutely clean'. They cannot be preceded by adverbs like *很* (very) or 非常 (very).

思) or a verb and a noun (*无意识，有礼貌，有意义*).

4 Conclusions:

4.1.1 Proportionally NNSs used average FSs of this type.

4.1.2 NNSs made more mistakes than their average level.

4.1.3 Half of NNSs FSs are also used by NSs.

Appendix IV.4.2 Polywords: Adverbs

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
诚心诚意	sincerely sincerely					1	1		
的确也是	indeed (also be)*					1	1		
都是	both (be)					1			1
反正是	anyway (be)					1	1		
非常的	very (PRT)	2	2			9	9		
故意的	deliberately (PRT)	3	1	2		5	5		
其实是	actually (be)					1	1		
其实也是	actually (also be)					1	1		
确实是	indeed (be)					5	5		
实际上	actually (up)					2	2		
实在是	indeed (be)					7	7		

也许是	probably (be)					1	1		
一个人	one unit person = alone					1	1		
一急之下	one hurry's down = in a great hurry					1	1		
已经是	already (be)					2	1	1	
应该是	supposingly (be)					1	1		
有一些	have + a bit = quite					1	1		
有的时候	some times					1	1		
有心的	deliberately (PRT)					1	1		
再一次	again + once	1			1	1	1		
真的	really (PRT)	18	18			26	26		
真的是	really (PRT be)					18	18		
真是	really (be)					2	2		
真心诚意	sincerely sincerely					1	1		
主要是	mainly (be)					1	1		
Total		24	21	2	1	92	90	1	1

Analysis:

1 FS ratio: **0.26**, with 24 by NNSs and 92 by NSs (All polywords: 0.41; overall: 0.4).

2 Error ratios: **12.5%** by NNSs (All polywords: 20.75%; overall: 41.6%) and **2.1%** by NSs (All polywords: 5.34%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate: 100% (24 out of 24).

3.2 Among the 25 types identified, only 4 were used by NNSs while NSs used all.

3.3 Distributions are very uneven in both groups. In NNS, *真的* along accounts for 75% of occurances. In NS, the 6 types with highest frequencies account for 76%.

3.4 In NS, the 6 types with highest frequencies are *真的* (26), *真的是* (18), *非常的* (9), *实在是* (7), *确实是* (5) and *故意的* (5). The top 5 types are semantically very similar: very/indeed/really. The top 3 types in NNS are also among this group.

3.5 In terms of structure, 13 types of FSs have a verb 'to be' *是* (or 'also be' *也是*) and 4 have a PRT *de*, added to an adverb. It seems that these are the main ways to create polyword adverbs.

4 Conclusions:

4.1.1 Proportionally NNSs used fewer FSs of this type.

4.1.2 NNSs also made fewer mistakes than their average level.

4.1.3 All NNSs FSs are also used by NSs.

4.1.4 NNSs rely on very small varieties of FSs.

* The components in the brackets do not semantically contribute to the meaning of the FSs

Appendix IV.4.3 Polywords: Conjunctions

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
还有	still have = moreover	1	1						
或者是	or (be)					1	1		
所以说	therefore (say)					1			1
甚至是	even (be)					1	1		
Total		1	1			3	2		1

Analysis:

1 FS ratio: **0.33**, with 1 by NNSs and 3 by NSs (All polywords: 0.41; overall: 0.4)

2	Error ratios: **0%** by NNSs (All polywords: 20.75%; overall: 41.6%) and **33.3%** by NSs (All polywords: 5.34%; overall: 8.25%).
3	Overlapping rate: 0% (0 out of 1).
4	As shown by the 3 FSs by NSs, adding a verb 'to be' *是* or a verb *说* to a conjunctive word seems to be how Chinese polyword adverbs are made.
5	Due to the limited frequencies, it is difficult to draw other conclusions.

Appendix IV.4.4 Polywords: Nouns & pronouns

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
别的时候	other time	2			2				
出勤情况	attendance situation = attendance rate					1	1		
第二次	number two time = second time					3	3		
第一次	number one time = first time	1		1		2	1		1
电话号码	telephone number	1	1						
冠军杯比赛	champion cup final = championship					1	1		
交通事故	traffic accident	1		1					
今早	today morning					1	1		
今天上午	today morning	2	2			10	9		1
今天下午	today afternoon					1		1	

今天早上	today early morning	12	12			11	10	1	
另外一个	other one M	2	1		1	1	1		
那天早上	that day morning	1			1				
那种	that type	1			1				
千错万错	thousand faults ten-thousand faults = all faults					1	1		
前几天	front few day = last few days					1	1		
上次	up time = last time					3	2	1	
上一次	up one time = last time	2	2			2	2		
什么时候	what time	4	3	1		1	1		
什么事情	what matter					1	1		
突发事件	suddenly happen incident = urgent matter					1	1		
下次	down time = next time	1	1			2	2		
下个	down unit = next one					1	1		
下一次	down one time = next time	1	1			3	3		
下一个	down one unit = next one					1	1		
一大半	one majority = majority					1	1		
一开始	one beginning = the very beginning					1	1		

再一次	again one time = another time	3			3	2	1		1
再一个	again one unit = another	1			1				
这次	this time					21	19	1	1
这段	this period					2	2		
这几天	this few day = these days					1	1		
这件	this M					4	3	1	
这门	this M					1	1		
这样子	this way (of doing thing)					1	1		
这一次	this one time = this time	1	1			7	7		
这一点	this one point = this point					1	1		
这一个	this one unit = this one					1	1		
这一门	this one division = this course					3	3		
这种	this type	1		1		3	3		
最后一次	last one time					1	1		
最后一个	last one unit	2	2						
昨晚	yesterday evening					1	1		
昨天晚上	yesterday evening					8	8		
昨天早上	yesterday morning	1			1				
Total		40	26	4	10	108	99	5	4

Analysis:

1 FS ratio: **0.37**, with 40 by NNSs and 108 by NSs (All polywords: 0.41; overall: 0.4)

2 Error ratios: **35%** by NNSs (All polywords: 20.75%; overall: 41.6%) and **8.3%** by NSs (All polywords: 5.34%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate: 75% (30 out of 40).

3.2 Among the 46 types identified (19 types used by NNSs, 38 by NSs), 11 types are used by both.

3.3 In NS, the 5 types with highest frequencies are *这次* (21), *今天早上* (11), *今天上午* (10), *昨天晚上* (8) and *这一次* (7). The first and the fifth are synonyms and the other 3 are time words.

3.4 Two of the top 3 types in NNS are also among this group: *今天早上* (12), *什么时候* (4) *and* *再一次* (3).

4 Conclusions:

4.1.1 Proportionally NNSs used average number of FSs of this type.

4.1.2 NNSs made significantly more mistakes than average.

4.1.3 Most NNSs FSs are also used by NSs.

Appendix IV.4.5 Polywords: Verbs

(CR: correct; LK: likely; ULK: unlikely)

		NNS				NS			
FSs	Literal English	TOTAL	CR	LK	ULK	TOTAL	CR	LK	ULK
病了	sick PRT = got sick	1	1						
错了	wrong PRT = wronged					1	1		
错过了	miss PRT = missed					1	1		
加班加点	add time add hour = work extra hours					1	1		
事出有因	things happened with reason = excusable					1	1		

实话实说	true words true say = tell the truth					1	1		
死了	die PRT = died	1	1						
忘了	forget PRT = forgot	17	16	1		18	18		
忘记了	forget PRT = forgot	11	11			10	10		
想着	think PRT = thinking					1	1		
Total		30	29	1		34	34		

Analysis:

1 FS ratio: **0.88**, with 30 by NNSs and 34 by NSs (All polywords: 0.41; overall: 0.4)

2 Error ratios: **3.3%** by NNSs (All polywords: 20.75%; overall: 41.6%) and **0%** by NSs (All polywords: 5.34%; overall: 8.25%).

3 Other findings:

3.1 Overlapping rate: 93.3 (28 out of 30).

3.2 Among the 10 types identified (4 types used by NNSs, 8 by NSs), only 2 types are used by both (the two with highest frequencies in both groups: *忘了* and 忘记了).

3.3 Besides *忘了* and 忘记了, distributions of other FSs in both NNS and NS data are extremely even, with 1 token each.

3.4 In terms of structure, FSs of this type can fall into two categories: verb + PRT (e.g. *病了*; 想着); fixed idioms with 4 characters (e.g. *加班加点; 实话实说*).

3.5 Both NNSs and NSs used *忘了* and 忘记了 many times, as a natural consequence of the language task being investigated.

4 Conclusions:

4.1.1 Proportionally NNSs used far more FSs of this type.

4.1.2 NNSs made far fewer mistakes than their average level.

4.1.3 Both groups had very similar choices of FSs.

Appendix IV.4.6 Overall distribution of Polywords

POLYWORDS	NNS Total	CR (correct)	LK (likely)	ULK (unlikely)	NS Total	CR (correct)	LK (likely)	ULK (unlikely)

Adjective	11	7	1	3	24	22	1	1
Adverb	24	21	2	1	92	90	1	1
Conjunction	1	1	0	0	3	2	0	1
Noun & pronoun	40	26	4	10	108	99	5	4
Verb	30	29	1	0	34	34	0	0
sub-total	106	84	8	14	261	247	7	7
% of sub-total	100%	79.25%	7.55%	13.21%	100%	94.64%	2.68%	2.68%
% of total (polyword)	15.8%	12.6%	1.2%	2.1%	15.3%	14.5%	0.4%	0.4%
total	671	392	127	152	1723	1581	56	86
% of total (overall)	100%	58.4%	18.9%	22.7%	100%	91.8%	3.3%	5.0%

Appendix IV.4.7 Types of Polywords

POLYWORDS	NNS Total	NS Total	Total	NNS types	NS types	Commom types	Total Types
Adjective	11	24	35	4	6	1	9
Adverb	24	92	116	4	25	4	25
Conjunction	1	3	4	1	3	0	4
Noun & pronoun	40	108	148	19	38	11	46
Verb	30	34	64	4	8	2	10
sub-total	106	261	367	32	80	18	94