university of groningen

University Medical Center Groningen

University of Groningen

# The Dynamics of English Writing Development in Advanced Chinese Learners

Hou, Junping

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*
Publisher's PDF, also known as Version of record

*Publication date:*
2017

Link to publication in University of Groningen/UMCG research database

*Citation for published version (APA):*
Hou, J. (2017). *The Dynamics of English Writing Development in Advanced Chinese Learners*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.

**Copyright**
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

**Take-down policy**
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

Download date: 05-02-2023

# Chapter 4

## Chunk use and development in advanced Chinese L2 learners of English

This chapter is a slightly edited version of Hou, J., Loerts, H., & Verspoor, M. H. (2016). Chunk use and development in advanced Chinese L2 learners of English. *Language Teaching Research*, 1-21. https://doi.org/10.1177/

# Chapter 4 | 64

## 4.1 Introduction

Most second language (L2) teaching and research has focused on grammar and the lexicon, and some linguists claim that the lexicon may be more important than grammar in the L2 developmental process (e.g., Widdowson, 1989; Ding & Qi, 2005). Moreover, the lexicon is argued to be far more than a collection of isolated words. If L2 learners want to be fluent, accurate and authentic, they must also know how words are combined (Eyckmans, Boers, & Stengers, 2007; Pawley & Syder, 1983). Increasingly, studies suggest that the use of 'ready-made chunks or preferred sequences of words plays a significant part in language acquisition and production' (Wood, 2010:38). Ready-made chunks or preferred sequences of words (also referred to as multi-word units, formulaic sequences or prefabs) is a well-known phenomenon in everyday language learning, processing and production. It is estimated that the proportion of chunks in native-like written language is about 52.3% (Erman & Warren, 2000) and it tends to be even higher in native-like spoken language (Ellis, 2008:4). The importance of lexical chunks has been demonstrated as they are considered to be a characteristic feature of an L1 speaker's spoken and written repertoire (Sinclair, 1991; Nattinger & DeCarrico, 1992; Wray, 2002; Schmitt & Carter, 2004) and 'are required for acceptance into the target community' (Appel & Wood, 2016:55). Therefore, we may assume that chunks are a highly relevant variable in tracing the development of L2 as the use of chunks is assumed to be related to the learner's fluency, accuracy and authenticity in the L2 (Pawley & Syder, 1983; Nattinger & DeCarrico, 1992; Perera, 2001; Eyckmans et al., 2007; etc.). Also Housen and Kuiken (2009) argue that the use of chunks is interrelated with the three dimensions of language proficiency—complexity, accuracy and fluency—but few studies so far have looked at how. The goal of the chapter is to see to what extent chunks relate to the construct of proficiency and whether the use of chunks changes over the course of 18 months of English classes in advanced Chinese learners of English.

## 4.2 Theoretical approaches and empirical studies

Lexical chunks can be defined according to grammatical structure, function, or degree of conventionalization or a combination thereof.

They are usually defined as groups of multiword units of language which are stored in long-term memory as if they were single lexical words such as *for example*, *depend on*, and *pay attention to*. Other names for the same phenomenon abound in the literature: 'prefabricated patterns' (Hakuta, 1974), 'prefabs or lexical phrases' (Nattinger & DeCarrico, 1992), 'ready-made (complex) units' (Cowie, 1992), 'lexical chunks' (Lewis, 1993), 'multi-word items' (Moon, 1997:43), 'formulaic sequences' (Wray, 2002), 'lexical items' (Nation & Meara, 2002), and so on. All definitions point to a basic characteristic of lexical chunks: they can be fixed or semi-fixed lexical phrases, but usually units longer than a single word.

Some of the differences in definitions of lexical chunks are due to the different theoretical perspectives that researchers take. Alexander (1984) classified lexical chunks from the perspective of formal degree and function and divides them into idioms, discourse markers, proverbs, catchphrases as well as quotations and allusions. The classification by Nattinger and DeCarrico (1992) consisted of four categories: poly words, institutionalized expressions, phrasal constraints and sentence builders. A similar classification was created by Lewis (1997), who identified four basic groups of lexical items: words and poly words, collocations, institutionalized utterance frames and heads. Moon (1997), on the other hand, argued that lexical chunks consist of five kinds of items: compounds, verbs, idioms, fixed phrases and prefabs or ready-made units, which were defined into three criteria: institutionalization, fixedness and non-compositionality.

The different frameworks and approaches can be reconciled by the fact that, as Wray (2002) pointed out, the definition and classification of chunks very much depend on the chunks actually identified in the target texts and methodological choices that have to be made depending on the theoretical positions of the researchers and goals of the study.

The current study will take a usage-based perspective because usage-based theory 'can account for the more problematic structural properties of chunks such as continua, nesting, overlaps, and chunking at different levels; the place of chunks in the language system, their acquisition, and development over time' (Smiskova-Gustafsson, 2013:16). Usage-based approaches see language as part of human cognition, as a tool for making meaning in a social context; language learning is the learning of

constructions during usage events by mapping form onto meaning (Ellis & Cadierno, 2009; Tomasello, 2003). Since usage-based approaches do not draw strict boundaries between traditional language subsystems such as grammar and lexicon (Römer, 2009), chunks can be seen as conventionalized form-meaning mappings, the result of repeated use of certain linguistic units and specificity, from two-word collocations, to collostructions (i.e., the interaction between collocations and constructions), to clause collocations, and so on. 'Observed chunk categories (i.e., different chunk subtypes) are also emergent, which means that characteristic features of chunks form continua, rather than closed categories' (Smiskova-Gustafsson, 2013:21), i.e., it may be difficult to classify chunks according to clear definitions. Thus, from a usage-based perspective, a chunk is defined as 'a conventionalized word sequence expressing a certain concept, a combination of two or more orthographic words, which may also include variable slots, expressing an idea (concept) in a conventionalized way' (Smiskova-Gustafsson, 2013:28), a definition in line with Langacker's (2008:84) notion of 'units representing normal ways of saying things'. From a usage-based perspective, 'formulaic' can be regarded as 'conventionalized', thus, 'all units of language are formulaic' (Wray, 2002:245) and conventionalized form-meaning mappings are at all levels of language, which may cause difficulties in identifying and classifying chunks. A usage-based perspective takes frequency of occurrence as the driving force in both L1 and L2 development. Therefore, it is assumed that proficient language users will be able to recognize conventionalized sequences: their cognitive systems are tuned into frequencies of occurrence, so they have strong intuitions about how things are normally said.

Jespersen was the first to describe phraseology as an 'indispensable' (2013:16-17) dimension of language competence. Becker (1975) also pointed out the importance of lexical phrases and suggested that ready-made frameworks are used to hang the expression of our ideas on. Lewis (1993) advocated a lexical chunks approach to teaching and argued that language consists of grammatical lexis, not lexicalized grammar, and lexical chunks, as a kind of grammatical lexis, play a vital role in language acquisition in all aspects of listening, speaking, reading and writing. In line with this reasoning, Nattinger and DeCarrico (1992) conducted a study on lexical

phrases and concluded that the input of these lexical phrases can help English as a Foreign Language (EFL) learners to express themselves well in writing. Data collected by Perera (2001) suggested that chunks not only aid in acquisition because they help a learner achieve fluency, but also because they can help a learner internalize grammatical rules. Boers, Eyckmans, Kappel, Stengers and Demecheleer, (2006:257) found that advanced L2 speakers' use of formulaic sequences with a 'noticing' instructional method over the course of 22 teaching hours was effective and concluded that 'the use of formulaic sequences was shown to be especially beneficial to perceptions of learners' fluency and range of expressions.' The influence of the intensity in instructional hours on the acquisition of formulaic sequences by various adult EFL learners was recently investigated by Serrano, Stengers and Housen (2015). They compared the performance of EFL adult-learners with low, intermediate, and high FL proficiency and compared the effects of 110 hours concentrated in one month versus 110 hours of instruction spread over seven months. The results showed that learners with an advanced level do not seem to benefit from intensive instruction to the same extent as lower-proficiency learners do, and students at the intermediate level are the ones that benefit the most from intensive instruction in terms of production of formulaic sequences. AlHassan and Wood (2015) studied the effectiveness of focused instruction of formulaic sequences in augmenting L2 learners timed academic writing skills over a 10 week period; they found that an explicit instructional approach to formulaic sequences can enhance their subsequent acquisition and promote L2 learners' tendency to integrate this language phenomenon in their writing. Appel and Wood (2016) studied how recurrent word combinations can be used to identify differences between L2 English writers with different levels of proficiency. They found that lower-level writers relied more heavily on repeated word sequences in the discourse they produce, whereas high level writers use more referential expressions.

There are also two studies from a usage-based perspective on chunks in L2 development. In a cross-sectional study, Verspoor et al. (2012) studied the number and types of chunks in written texts at five different proficiency levels (beginner to intermediate) of Dutch learners of English as an L2 and found that the total number of chunks was one of the strongest

predictors of proficiency levels. Furthermore, the types of chunks used were different among the text levels and high intermediate learners used more words with target-like collocations and more lexical chunks, in particular particles and compounds. In a similar population, Smiskova-Gustafsson (2013) investigated the relationship between the L2 learners' use of chunks and the amount and kind of input they are receiving. She concluded that L2 learners in the high-input conditions were more successful in their development of authentic sounding L1 chunks than the low-input learners. The difference was significant, not so much in the number or types of chunks the learners used, but in the so-called 'chunk coverage', the ratio of the total number of words used in chunks and the total number of words in the text, a measure that will also be examined in the current study.

Recently, lexical chunks have also attracted Chinese researchers, who now realize the importance of lexical chunks in English as a second or foreign language (ESL/EFL) teaching and learning. Most Chinese papers on this topic review the history of research on lexical chunks, the development of the definitions, and the application of the classifications (e.g., Duan, 2008; Wu & Wang, 2002; Yao, 2002). A critical review of the development, problems and future directions of chunk studies in China was recently written by Yin (2013), who concluded that most studies are corpus-based and show little about development over time. A few researchers have carried out experimental studies to verify the positive effects of chunks on students' language learning and acquisition. Ding and Qi (2005), for example, analyzed the relationship between the use of chunks and writing proficiency and showed that there is a significant positive correlation between learners' use of chunks and their writing proficiency. Ma (2009) studied the use of 191 high-frequency chunks in expository writings of Chinese university students following English as an L2 major and concludes that chunks with past tense, nouns with particles and chunks with clauses have a low frequency in Chinese students' writings. Hsu (2007) found a significant positive correlation between frequency of lexical collocations and oral proficiency scores for native Chinese participating in an impromptu speech contest in English in Taiwan. All of these studies share the view that chunks are highly correlated with L2 proficiency, and advanced learners produce more chunks than lower level learners in L2 spoken and written English.

Even though chunks are important to sound fluent in the L2, it also appears to be difficult for adult learners to pick them up. As Wray points out 'the formulaic sequences used by native speakers are not easy for learners to identify and master; their absence greatly contributes to learners not sounding idiomatic' (2002: 176). Zhao (2009) found a correlation between the use of lexical chunks and proficient language production as measured by a writing test in native Chinese speakers learning English and found that Chinese speakers had quite a poor knowledge of English lexical chunks overall.

The current study will look at the development of chunks use in the texts written by highly advanced Chinese learners of English over the course of 18 months. In the previous study of Chapter 3 (also see Hou, Verspoor & Loerts, 2016) the same texts were holistically rated on proficiency level and analyzed on a great number of syntactic complexity measures, none of which showed much development in this group of advanced learners. However, rather than assuming that no development took place at this level, we based ourselves on Verspoor et al. (2012) and hypothesized that syntactic complexity must have reached somewhat of a ceiling effect and the growth was to be found more in the lexicon, specifically in the use of chunks. The main research questions are as follows:

(1) To what extent does chunk use correlate with holistic proficiency ratings at advanced levels?
(2) To what extent does chunk use increase in advanced learners during 18 months of English instruction?

### 4.3 Methodology

The current study has a pre-post design. The first two and last two texts of a group of advanced Chinese students of English enrolled in an 18-month CET-6 course (a course that prepares for the highest level English exam for students in China who do not major in English), were first holistically scored on a number of features that are assumed to collectively represent proficiency (details follow below). After that the texts were scrutinized for L1 target like chunks, which were categorized and counted in a number of ways based on the Verspoor et al. (2012) classification scheme to assess potential relationships with proficiency levels and to ascertain whether there

are any differences in these measurements between pre- and posttest texts.

### 4.3.1 Participants

The 18 Chinese participants were enrolled in the most advanced course for non-majors in English. After entering a prestigious university, highly selective in itself, the participants took a proficiency exam that placed them in the higher level class, aimed at passing the CET-6 exam (the highest level exam for non-majors). They were enrolled in an 18-month English program at the university.

### 4.3.2 Procedures

The students were exposed to English input with professional teachers with a high L2 proficiency and materials and assignments appropriate for their level as part of their training for the exam in an 18-month course. Each week they were exposed to English for about 4-5 hours (totaling about 196 hours). Even though English is not their major, these students have to spend a relatively larger amount of time than their peers who are studying for the lower CET-4 exam to obtain the higher English proficiency certificate, which is thought to be of great benefit for future employment.

The teachers at this university apply a learning-by-writing theory and spend time on writing practice following CET directions. They use model essays, sometimes those written by class members, to be discussed in class. In addition to grades, the teachers give rather detailed feedback on errors and brief evaluations and suggestions for improvement. Sometimes students are asked to give each other peer feedback. Most writing assignments have argumentative topics.

For the current study, the first two and last two texts for each student were selected as pre- and posttest texts, all of which were argumentative in nature. There was no time pressure as the essays were written after class with possible access to dictionaries and other resources, such as internet. The assignments followed the demands and purposes of written exercises as described in each curriculum unit, and students were asked to practice what they had learned in that specific unit. It is possible that students copied chunks from other sources, but the fact that teachers did not emphasize the use of chunks and the learners were not aware that their texts would be examined for chunks makes it highly unlikely that they focused on using

more or longer chunks. There was no word limit either and the text length varies from 96 to 399 words.

### 4.3.3 Holistic proficiency scores

In the study of Chapter 3, all texts were holistically rated on general proficiency. A rubric consisting of 5 general indices of L2 proficiency--complexity, accuracy, fluency, idiomaticity and coherence (CAFIC)--was created and refined during a pilot assessment procedure. After a two-hour training session with the rubric, 196 texts (including the 72 texts of the present study) at three different levels of proficiency obtained at the beginning and end of the students' respective courses were randomly mixed and holistically rated by 8 trained raters (5 L1 and 3 L2 speakers of English) in groups of 3. The raters were not aware of the data collection background nor of the specific measures that were to be examined in this study.

Each text was scored on a Likert scale from 1 to 5 for each of the five CAFIC rubrics, with a possible total score of 25 points. The procedure yielded several rating results: for each rater there were scores on each rubric and a total for each text. All scores were statistically analyzed with SPSS (Version 23) for their reliability, correlations and differences within and between groups. The statistical analyses (see Chapter 3) revealed excellent interrater reliability on the overall CAFIC score (ICC(1,3) = .799), high interrater reliability on the sub-scores complexity, fluency, idiomaticity, and coherence (all ICCs(1,3) > .636) and fair agreement among raters for accuracy (ICC(1,3) = .494).

Further analysis within each of the proficiency groups of learners (beginning, intermediate and advanced; see Chapter 3) showed that the advanced group, the group of interest in the current study, showed no significant development in the total average score, the consensus score nor on any of the five CAFIC rubrics over the 18 months of study. In a subsequent analysis of more objective and specific complexity measures using the Synlex Analyzer (Lu, 2012), the learners showed significant development in a few isolated variables, such as verb variation-II (types of verbs divided by the number of lexical words), adjective variation, modifier variation, coordinate phrase per clause, and coordinate phrase per T-unit, which are mainly lexical in nature.

In the present study, the total rating score and CAFIC sub-measures are used to examine more closely how these measures correlate with the use of chunks within this group of advanced learners.

### 4.3.4 Chunk classification

In the literature there is no general consensus on what constitutes a chunk, and chunks are notoriously difficult to identify (cf. Eyckmans et al., 2007). Therefore, it is important to have clear, delineated definitions (even though they may not be exhaustive), and coding has to be done consistently within the study. One important choice that has to be made is whether all frequently occurring chunks are counted, even though they are not fully L1 target-like chunks, or only chunks that are fully L1 target-like. For example, in our example text there is a string of words 'to power up the brain'. Here we counted 'to power up' as a particle, but 'to power up the brain' as a whole is not considered an L1 chunk as it does occur frequently, unlike the phrase 'power up your brain'. In our study, we chose to focus on L1 target-like chunks as one of the goals is to compare the findings with those in Verspoor et al. (2012). Therefore, we will also draw on their identification and coding schemas and distinguish between two general kinds: partially schematic and fixed chunks. This classification was originally based on the assumption that schematic (grammatical) chunks may become productive once the pattern was recognized, whereas fixed (lexical) chunks would have to be learnt one by one. As Table 4-1 exemplifies, the former category involves flexible, highly institutionalized, compositional grammatical structures named structures and complements that act as a unit but have open slots that may be filled with different vocabulary items. These types of chunks can be taught explicitly and practiced in the classroom. These also showed significant differences at the lower levels of proficiency in the Verspoor et al. (2012) study. The latter category includes words which collocate at a lexical level, like compounds, collocations, chunks with particles, fixed phrases and discourse chunks. These lexical categories are all based on their surface forms, except for the discourse chunks, which are categorized according to function. The reason to count these separately is that we would expect more advanced writers to make more use of these. Some discourse chunks may

**Table 4-1.** Chunk coding schema showing the two categories containing grammatical and lexical chunks.

| Label | Definition | Examples |
|---|---|---|
| ***Grammatical (schematic) chunks*** | | |
| **Structures** | Chunks with a fixed structure with slot-fillers | *It is …. (adj.) for (pron./n.) to …. (v.), not only …, but also …, etc.* |
| **Complements** | Chunks with verbs and complements such as infinitives, gerunds, nominal sentences, or reflexives | *decide to + infinitive, be able to + infinitive, instead of + gerund, think/know that CLAUSE, etc.* |
| ***Lexical (fixed) chunks*** | | |
| **Compounds** | Fixed combinations of nouns, adjectives, prepositions, or particles (sometimes written as one word) | *college graduates, high school, homework, science fiction, lifestyle, etc.* |
| **Particles** | Chunks with verbs (including phrasal verbs) or nouns with prepositions or particles, | *depend on, be valuable to, in the evening, a group of, because of, addicted to, etc.* |
| **Collocations** | Collocating nouns, adjectives, verbs and also adverbs, prepositions, pronouns | *choose wisely, a positive attitude, keep calm, enjoy ourselves, pay attention, bright future, etc.* |
| **Fixed phrases** | Highly conventional word combinations, often idiomatic, consisting mainly of more than two words | *ups and downs, practice makes perfect, all at once, what the future may have in store, no pain no gain, etc.* |
| **Discourse** | A chunk (regardless of its form) with a discourse function | *for example, in my opinion, such as, as we know, etc.* |

*Note* that Verspoor et al. (2012) used the terms schematic and fixed chunks to refer to grammatical and lexical chunks respectively.

also have slots, but they are grouped as discourse chunks because of their specific function. For lexical chunks, there is usually no known motivation for the specific combinations and they are not usually focused on systematically in traditional teaching approaches because of their great numbers. Therefore, they are assumed to be acquired mainly through exposure, memorization, and repetition.

As coding the texts consistently according to the very specific definitions given in Table 4-1 was of the utmost importance, the following procedure was used. First, two of the authors worked carefully through 20

texts together, identifying and classifying all possible chunks encountered, refining the descriptions. As the detailed example coding scheme in Appendix 6 shows, one of the main problems is how to classify embedded chunks as in 'pay attention to'. To count types, we decided to count each chunk as a separate type: 'pay attention' is a collocation and 'attention to' is a particle. After agreeing on the classification scheme and adding specific examples from the current study, one of the authors coded the remainder of the texts and checked with her co-authors when questions arose. When a chunk was identified, an automatic search, based on formulas and constructions, was made through the remainder of the texts to see if other texts contained the chunk, too. The author then transferred all chunks from the word file to an excel file with their classification and the number of words in the chunk and checked with the co-authors on any possible inconsistencies. All chunks from all texts and participants were sorted and again checked for consistency. Any inconsistencies in grouping and counting were taken out before the data was statistically analyzed.

### 4.3.5 Measures

Not only is identifying and classifying chunks notoriously difficult, but how to count them is problematic as well. Some chunks consist of only one or two words (e.g., a compound) and some chunks consist of four or more words (e.g., the fixed phrase 'once and for all'). Moreover, chunks may be concatenated or embedded as the 'pay attention to' example showed. Because we expect more advanced learners to use longer chunks (e.g., 15 3-word chunks rather than 15 2- word chunks), we calculated average chunk length. In the case of concatenated or embedded chunks, the longest chunk was counted, so in the case of 'pay attention to' the length was 3 words.

Another problem is that using longer chunks may result in relatively fewer chunks (e.g., 15 two-word chunks versus 10 three-word chunks). Also, a more advanced learner may use relatively more chunks (e.g., 20 two-word chunks versus 15 two-word chunks). To capture an increase in both length and relative frequency of chunks, chunk coverage was calculated. Chunk coverage is the total number of words occurring in chunks in the text (each word only counted once) divided by the total number of words in the text.

**Table 4-2.** Measures used in analyses

| **Text length:** | **total number of words in the text** |
|---|---|
| **Number of words in chunks** | total number of unique words in chunks |
| ***Chunks count*** | |
| **Total chunks count** | absolute number of chunks in the text |
| **Grammatical count** | absolute number of grammatical chunks in the text |
| **Lexical count** | absolute number of lexical chunks in the text |
| **structures, complements, compounds, particles, collocations, fixed phrases, discourse** | absolute numbers per type |
| ***Chunk length*** | |
| **Total chunks length** | average number of words per unique chunk (based on total-length) |
| **Grammatical length** | average number of words per unique grammatical chunk (based on total-length) |
| **Lexical length** | average number of words per unique lexical chunk (based on total-length) |
| **structures, complements, compounds, particles, collocations, fixed phrases, discourse** | average length per type |
| ***Chunk ratio*** | |
| **Total chunks** | number of chunks / square root (text length) |
| **Grammatical chunks** | number of grammatical chunks / square root (text length) |
| **Lexical chunks** | number of lexical chunks / square root (text length) |
| **structures, complements, compounds, particles, collocations, fixed phrases, discourse** | number of types of chunks / square root (text length) |
| ***Chunk coverage*** | number of words in chunks / text length |

Finally, the number of chunks need to be corrected for text length. Originally two *chunk ratio* calculations were compared in our preliminary analyses: a plain type token ratio (TTR) and a root type token ratio **(**R-TTR). Because the R-TTR showed several strong correlations with other measures and the plain TTR did not, we decided to use the R-TTR. We assume the effect of the root calculation is that it shows more subtle differences within the middle of a distribution curve (Wachal & Spreen 1973). The root chunk ratio is calculated by dividing the number of chunks by the square root of the total number of words. Table 4-2 gives a summary of the measures used in the analyses.

### 4.3.6 Analyses

To see to what extent chunk measures correlate to general proficiency measures statistical analyses were first carried out on all texts (pre- and posttest texts together). To see if there were differences between the pre- and posttest texts, the analyses were done on these datasets separately, too.

Pearson *r* correlations were carried out between the CAFIC measures, number of words in the text, chunk ratio, chunk length, and chunk coverage. The results of Spearman Rank-Order correlations are reported in case of violations of the normality assumption. Paired Samples *t*-tests and, if required due to deviations from normality, Wilcoxon Signed-Rank Tests were conducted to compare the use of chunks in the pre- and posttest texts.

## 4.4 Results

### 4.4.1 Correlational analyses

The analyses in this part aim to answer the first research question: to what extent does chunk use correlate with holistic proficiency ratings? Table 4-3 shows correlations between CAFIC scores, length of text, root chunk ratios of all chunks together and grammatical and lexical chunks separately, and chunk coverage of all 72 pre- and posttest texts together.

As shown in Chapter 3, the CAFIC sub-measures are highly and significantly positively correlated with each other and with the total CAFIC score. As far as chunk ratios is concerned, all proficiency measures show strong positive correlations with both the overall and lexical root chunk ratio measure. The grammatical chunk ratio does not correlate significantly with any of the CAFIC measures, nor is it related to overall and lexical root chunk ratio or chunk coverage (all *r*s < .239; all *p*s > .16). Text length is strongly and positively correlated with all CAFIC proficiency measures, overall as well as lexical chunk ratio, but again not with grammatical chunk ratio ($r = .216$; $p = .205$). Chunk coverage shows moderately strong positive correlations with overall and lexical root chunk ratio (*r*s > .574; *p*s < .001) and weak positive correlations with idiomaticity and coherence (*r*s > .369; *p*s < .05). Weak positive correlations were found between chunk coverage and accuracy ($r = .327$; $p = .051$) and between coverage and overall CAFIC score ($r = .307$; $p = .064$), but these did not reach significance. Correlations with

**Table 4-3.** Correlations between CAFIC scores, root chunk ratios, chunk coverage, and text length of all 72 texts.

| | **Proficiency measures** | | | | **Root Chunk ratio measures** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | accuracy | Fluency | idiomaticity | coherence | CAFIC overall | Chunk ratio | Grammatical chunk ratio | Lexical chunk ratio | Length of text[#] | Average chunk length | Chunk coverage |
| complexity[#] | .664** | .728** | .741** | .813** | .903** | .681** | .085 | .658** | .698** | .238 | .263 |
| accuracy | 1 | .563** | .726** | .768** | .817** | .638** | .029 | .629** | .464** | .019 | .327 |
| fluency | | 1 | .712** | .743** | .875** | .722** | .112 | .682** | .838** | .116 | .097 |
| idiomaticity | | | 1 | .892** | .908** | .680** | .074 | .659** | .525** | .213 | .369* |
| coherence | | | | 1 | .943** | .732** | -.004 | .741** | .555** | .155 | .380* |
| CAFIC | | | | | 1 | .788** | .054 | .773** | .729** | .166 | .307 |
| Chunk ratio | | | | | | 1 | .239 | .910** | .662** | .037 | .574** |
| Grammatical chunk ratio | | | | | | | 1 | -.183 | .216 | -.212 | -.022 |
| Lexical chunk ratio | | | | | | | | 1 | .560** | .134 | .591** |
| Length of text[#] | | | | | | | | | 1 | .121 | -.015 |
| Average chunk length | | | | | | | | | | 1 | .314 |
| Chunk coverage | | | | | | | | | | | 1 |

[#] As these data did not follow a normal distribution (*SW*s < .923, *df* = 18, all *p*s < .05), Spearman rho correlations are provided instead of Pearson *r*.

**. Correlation is significant at the 0.01 level (2-tailed); *. Correlation is significant at the 0.05 level (2-tailed).

**Table 4-4.** Correlations between CAFIC measures, root chunk ratios, text length and chunk coverage in the 36 pretest texts.

| | **Proficiency measures** | | | | | **Root Chunk ratio measures** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | accuracy[#] | fluency | Idiomaticity | coherence | CAFIC | Chunk ratio | Grammatical chunk ratio | Lexical chunk ratio | Length of text | Average chunk length | Chunk coverage |
| Complexity | .608** | .736** | .760** | .789** | .907** | .838** | .265 | .777** | .774** | .533* | .628** |
| Accuracy[#] | 1 | .422 | .495** | .644** | .700** | .631** | .259 | .528* | .344 | .086 | .384 |
| Fluency | | 1 | .655** | .639** | .815** | .720** | .178 | .684** | .832** | .391 | .367 |
| Idiomaticity | | | 1 | .904** | .910** | .676** | .165 | .649** | .486* | .408 | .642** |
| Coherence | | | | 1 | .932** | .748** | .215 | .697** | .513* | .266 | .617** |
| CAFIC | | | | | 1 | .823** | .219 | .777** | .690** | .400 | .625** |
| Chunk ratio | | | | | | 1 | .323 | .920** | .756** | .271 | .717** |
| Gram chunk ratio | | | | | | | 1 | -.071 | .263 | -.263 | .142 |
| Lexical chunk ratio | | | | | | | | 1 | .692** | .403 | .698** |
| Length of text | | | | | | | | | 1 | .419 | .235 |
| Average chunk length | | | | | | | | | | 1 | .469* |
| Chunk coverage | | | | | | | | | | | 1 |

[#] As these data did not follow a normal distribution (*SW*s < .858, *df* = 18, all *p*s < .05), Spearman rho correlations are provided instead of Pearson *r*.

**. Correlation is significant at the 0.01level (2-tailed); *. Correlation is significant at the 0.05 level (2-tailed).

**Table 4-5.** Correlations between CAFIC measures, root chunk ratios, text length and chunk coverage in the 36 posttest texts.

| | **Proficiency measures** | | | | | **Root Chunk ratio measures** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | accuracy | fluency | Idiomaticity | coherence | CAFIC[#] | Chunk ratio | Grammatical chunk ratio | Lexical chunk ratio[#] | Length of text[#] | Chunk coverage |
| Complexity[#] | .721** | .746** | .706** | .854** | .904** | .766** | -.015 | .697** | .705** | .219 |
| accuracy | 1 | .799** | .793** | .834** | .836** | .750** | -.079 | .553* | .576** | .036 |
| fluency | | 1 | .763** | .825** | .888** | .836** | .067 | .532* | .829** | -.186 |
| idiomaticity | | | 1 | .889** | .865** | .736** | -.006 | .537* | .594** | .109 |
| coherence | | | | 1 | .930** | .819** | -.107 | .689** | .580* | .215 |
| CAFIC[#] | | | | | 1 | .773** | .121 | .651** | .778** | .145 |
| Chunk ratio | | | | | | 1 | .096 | .843** | .742** | .207 |
| Gram chunk ratio | | | | | | | 1 | -.445 | .284 | -.362 |
| Lexical chunk ratio[#] | | | | | | | | 1 | .478* | .606** |
| Length of text[#] | | | | | | | | | 1 | -.086 |
| Chunk coverage | | | | | | | | | | 1 |

[#] As these data did not follow a normal distribution (*SW*s < .895, *df* = 18, all *p*s < .05), Spearman rho correlations are provided instead of Pearson *r*.

**. Correlation is significant at the 0.01level (2-tailed); *. Correlation is significant at the 0.05 level (2-tailed).

overall chunk length (grammatical chunk length and lexical chunk length are not provided in table) were non-significant (all *r*s < .290, all *p*s > .09), except for a marginally significant positive correlation between chunk length and chunk coverage (*r* = .314; *p* = .062).

Tables 4-4 and 4-5 show the same correlations as in Table 4-3, but now between the various measurements in the pre- and posttest texts separately. The tables for the posttest data do not include information on the correlations with overall chunk length, grammatical chunk length, and lexical chunk length, because, as in the overall dataset, these did not reach significance in the posttest data (all *r*s < .261, all *p*s > .228). Text length did correlate with overall chunk length in the pretest data (*r* = .420; *p* < .05) and marginally so with length of grammatical chunks in the posttest data (*r* = .450; *p* = .06). Additionally, significant positive correlations were found between complexity and overall chunk length in the pretest data (*r* = .533; *p* < . 05) and between chunk coverage and overall chunk length in the pretest data (*r* = .469; *p* < . 05). Separate analyses revealed that, within the pretest data, accuracy was weakly positively correlated with both lexical and grammatical chunk length, but these correlations did not reach significance (*r* = .438; *p* = .068 and *r* = .404; *p* = . 09 respectively). Idiomaticity correlated positively with lexical chunk length in the pretest (*r* = .535; *p* < . 05), but the relationship between overall chunk length and overall CAFIC score failed to reach significance in the pretest data (*r* = .400; *p* = .099).

The correlations between proficiency measures and chunk ratio measures appear to be similar in the pretest data (Table 4-4) and posttest data (Table 4-5) with strong positive correlations for overall and lexical chunk ratio, but no significant correlations of grammatical chunk ratio with any of the other measures. As was visible in the overall dataset, no clear relationship is found between lexical and grammatical chunk ratio in the pretest (*r* = -.071), but a weak to moderate negative relationship between these two measures was present when only looking at the posttest data (*r* = -.445; *p* = .066). While chunk coverage only correlated positively with lexical chunk ratio in the posttest data (*r* = .606; *p* < .001), strong positive correlations between chunk coverage and several proficiency measures were additionally found in the pretest data, where only the correlations between accuracy and fluency and chunk coverage failed to reach significance (*r*

= .384; $p$ = .12 and $r$ = .367; $p$ = .13 respectively).

### 4.4.2 Differences in use of chunk types

Our second analysis aimed to answer the second research question, to what extent does chunk use increase after 18 months of English instruction? We explored the types of chunks used and then tested if there are differences over time in types of chunks, chunk ratio, and chunk coverage. As far as absolute numbers of lexical and grammatical chunks is concerned (not corrected for text length) learners on average used more lexical chunks ($M$ = 58.6; $SD$ = 18.1) than grammatical chunks ($M$ = 15.1; $SD$ = 5.3) in their texts. Learners used more chunks in the posttest texts (Median = 74.5) as compared to the pretest texts (Median = 67.5), but a Wilcoxon Signed Ranks Tests indicated that this increase did not reach significance for the overall score ($Z$ = 1.79; $p$ = 0.07) and was non-significant when testing the grammatical and lexical chunks counts separately ($p$s > .11).

As far as absolute numbers in types of chunks is concerned, Figure 4-1 shows that the use of most types of chunks, but especially complements and collocations, appear to increase slightly.

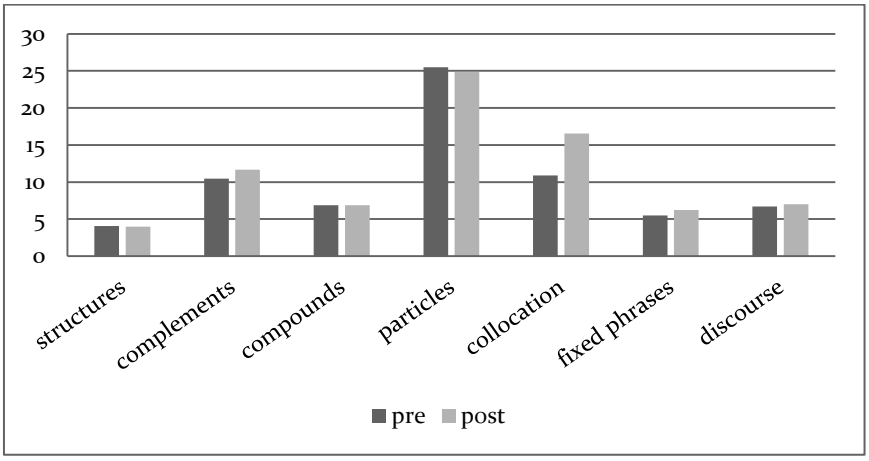


**Figure 4-1**. Overview of the number of different chunk types used in the pretest and posttest texts (average number of chunks per type)

As far as scores corrected for text length was concerned there were some clear differences. As Table 4-6 shows, overall root chunk ratio was

significantly higher in the posttest texts ($M$ = 3.73; $SD$ = 0.5) as compared to the pretest texts ($M$ = 3.38; $SD$ = 0.54); ($t(17) = 2.67$; $p < .05$). An analysis of grammatical and lexical chunk ratio measures revealed that this result was mainly driven by a significant increase in lexical root chunk ratio ($t(17) = 2.31$; $p < .05$), but not in grammatical root chunk ratio ($t(17) = 0.65$; $p = .53$). When looking at the different types of chunks in the pretest and posttest texts, particles are used most often overall ($M$ = 25.2; $SD$ = 8) followed by collocations ($M$ = 13.7; $SD$ = 7.2) and complements ($M$ = 11.1; $SD$ = 4.5). Paired Samples *t*-tests (Table 4-6) on the root chunk ratios of the different types revealed relatively more use of collocations at posttest ($M$ = 0.79; $SD$ = 0.29) than at pretest time ($M$ = 0.52; $SD$ = 0.21). This increase was significant ($t(17) = 3.92$; $p < .01$). The other structures did not significantly decrease or increase over time (see Table 4-6).

**Table 4-6.** Differences between the use of chunk types between pre- and posttest writings.

| | Mean Difference | *P*-value |
|---|---|---|
| Structures | 0.002 | .949 |
| Complements | .061 | .418 |
| Compounds | 0.008 | .873 |
| Particles | 0.032 | .678 |
| Collocations | .267 | .001** |
| Fixed phrases | .039 | .367 |
| Discourse | .021 | .490 |
| Grammar chunks | .058 | .525 |
| Lexical chunks | .285 | .033* |
| Chunks ratio | .375 | .016* |
| Average chunk length | -.030 | .532 |
| Chunk coverage | .039 | .006** |

**. Significant at p < .01; *. Significant at p < .05.

A Paired Samples *t-test* also revealed an increase in chunk coverage ($t(17) = 3.18$; $p < .01$), and a Wilcoxon Signed-Rank Tests showed that the number of words learners wrote in the pretest texts and posttest texts was very comparable ($Z = 0.06$; $p = 0.95$). The average length of the chunks did not

differ between pre- and posttest texts ($t$ (17) = 0.64; $p$ = .53), nor was there an increase when looking at the length of grammatical and lexical chunks separately ($p$s > .45).

### 4.5 Discussion and conclusion

The present study aimed at investigating to what extent chunk use correlates with holistic proficiency ratings and to what extent the use of chunks changes over the course of 18 months of English classes in advanced Chinese learners of English. Four texts written by 18 students who took an 18-month course in advanced English were collected, two at the beginning of the course and two at the end of the course. These texts were scored holistically on complexity, accuracy, fluency, idiomaticity and coherence by a team of raters. In Chapter 3 study, there appeared to be no development made whatsoever by this group, despite their 18 months' English course. As this result was difficult to believe, the authors further explored the data on specific variables that may not have been taken into account in the holistic scores. These subsequent analyses indeed revealed that there was a significant increase in some very specific lexical measures.

Because these changes were mainly at the lexical level and the use of chunks is also mainly a lexical phenomenon, the goal of the present study was to see if the use of chunks correlated with the holistic proficiency scores and whether over time, chunks had increased in number, had diversified or had become relatively longer over time. The chunk classification scheme developed by Verspoor et al. (2012) was used and several chunk measures were devised. To calculate an increase in number that allows for differences in text length, a root chunk ratio (the number of chunks in the text / the root of the number of the words in the text) was calculated. To calculate average chunk length, the number of words in chunks was divided by the number of chunks in total. To calculate chunk coverage the number of words that appeared in chunks was divided by the number of words in the text.

To see to what extent the use of chunks is related to proficiency levels, a number of correlation analyses were conducted. First, correlations were calculated with the previously established holistic CAFIC scores (sub scores and summed up scores) for all texts together and for the pre- and posttest texts separately. When all texts were taken together, chunk ratio and lexical

chunk ratio appeared to be significantly positively related with all CAFIC measures. Chunk coverage correlated significantly (but weakly) with only idiomaticity and coherence. Grammatical chunks, however, did not correlate with any of the CAFIC measures. The absence of any significant relationships here may be explained by the low number of occurrences of grammatical chunks. On the whole grammatical chunks occur less frequently than lexical chunks and with such low numbers, it may be difficult to obtain any significant correlations. The fact that lexical chunks correlated with the CAFIC measures when all texts were taken together is very much in line with previous findings (Ding & Qi, 2005; Hsu, 2007; Zhao, 2009; Verspoor et al., 2012). The more proficient writer uses more chunks; in other words, more use of chunks is related to a higher level of proficiency.

Also when the pre and posttests were looked at separately, both the overall and lexical chunk ratios correlated significantly with all CAFIC measures significance. In chunk coverage, however, there was a difference between the pre and posttests. In the pretest, chunk coverage correlated significantly with the totaled CAFIC score, three sub scores (complexity, idiomaticity and coherence), and average chunk length. In the posttest , however, there was no relation found between chunk coverage and any CAFIC measures, nor average chunk length. The positive correlations of chunk coverage scores in the analyses containing all of the texts was therefore mainly due to the texts written at the beginning of the program, not at the end. Apparently, the pretest chunk coverage was related more to chunk length, but at the end chunk coverage was related more to the relative number of chunks used. As we can see from the pre-post analysis, the number of collocations, which are two-word chunks, had increased significantly. This can be explained by the fact that chunk coverage is a hybrid measure that is related to both chunk length and the relative number of chunks used.

In addition to investigating possible relationships between chunk use and proficiency measures, we wanted to find out if the advanced Chinese learners of English had changed their use of chunks over time. Despite the fact that there was no increase in holistic scores, we hoped to find some differences in specific measures and expected lexical chunks to have increased. Indeed, there were several significant increases in the use of

chunks. At the end of their course, the learners used more lexical chunks, specifically collocations, and used relatively more chunks as indicated by the increase in chunk coverage. This is to a certain extent in line with Smiskova-Gustafsson (2013), who also found a significant difference in chunk coverage between low and high input learners. However, there is also a clear difference. Smiskova-Gustafsson (2013) attributed the difference in coverage to longer concatenated or embedded chunks such as 'one thing I know for sure'. In our study, however, the average chunk length did not increase at all, so the increased coverage must be due to relatively more chunks in the text. The fact that there was also such a strong correlation with the chunk ratio suggests that our learners did not embed or concatenate chunks as much as the Dutch learners in the high input condition. They just used relatively more short chunks.

Our findings have some implications for research and teaching. For research we would argue that in determining proficiency levels or in discovering developmental patterns of learners over time, especially at the advanced levels, two measures are useful,

- a chunk ratio which includes counting overlapping chunks as in the 'pay attention to' example as three separate types and
- chunk coverage (an indirect measure of the relative number of chunks or chunk length as it counts the words in overlapping chunks only once).

Interestingly, average chunk length, did not increase in this population.

For research into development over time, it means that if we want to see development at all in this advanced group, we may have to look especially at lexical measures. Our previous study had already shown significant development in three specific lexical measures, and in this study collocations and chunk coverage showed significant development. We might be able to explain the relative little development with the claim that relatively more development takes place earlier on than later (MacKay, 1982; Newell & Rosenbloom, 1981; Rosenbloom & Newell, 1987; Serrano et al., 2015). However, a complementary explanation in line with a dynamic view of development could be that, at different levels of proficiency, different

sub-systems of the language develop (Verspoor et al., 2012). At this advanced level, the linguistic system seems to have settled in general, and only rather subtle lexical changes still seem to take place.

For teaching, especially in the Chinese context, it shows that advanced students at the most advanced levels do develop, but in more subtle areas that are mainly lexical and specifically in knowing the words and 'the company these words keep' (Firth, 1957:11).

This study also has its limitations. We investigated only a small group of learners in China who were following an advanced course in English. We should not generalize beyond this group. Another limitation is the elusive construct of chunks. There are many ways to identify and operationalize them, and we decided to limit ourselves to a number of categories that may not be a chunk in every one's view. However, we would argue that we at least tried to keep the coding consistent. Still, another coding schema may have yielded different results.