

A Study on Correlation between Chinese Compound Sentence and Constituting Clause Based on the Menzerath-Altmann Law

Renkui Hou^{a,b}, ChuRen Huang^b, Hue San Do^b & Hongchao Liu^b

^aSchool of Literature, Ludong University, Yantai, Shandong, China

^bDepartment of Chinese and Bilingual Studies, The Hong Kong Polytechnic University,
 HongKong

Abstract: The Menzerath-Altmann law (henceforth the MA law), uncovered by Menzerath and formulated by Altmann, describes the correlation between the language structure and its immediate components with respect to all linguistic levels. This paper examined the correlation between Chinese compound sentences and its components, clauses, in different style texts based on the MA law. The results show that the correlation is described by the MA law only in the formal written texts.

Keywords: Menzerath-Altmann law, Chinese sentence length, Different registers

1. Introduction

Quantitative linguistics is one of the linguistic branches which combine linguistic and mathematic methods. The MA law, as one of the most influential quantitative linguistic laws, originates from the fact that the length of construct influences the length of its constituents. Paul Menzerath, a German phonetician and psychologist, summarized it that “The larger the whole the smaller the parts” after he detect the dependence of syllable length on word length (Menzerath 1954: 101). Based on the works of Menzerath, Altmann generalized this hypothesis with respect to all levels of linguistic analysis. He formulated it that “The longer a language construct the shorter its components” (Altmann 1980).

Hřebíček (1992, 1995, 1997) showed that the whole hierarchy of textual levels is based on this dependence and called it the Menzerath-Altmann’s law.

The law was interpreted by different ways, for example Köhler (1989). He shows that mechanism of shortening is a consequence of restrictions of the memory: the longer the construct, the more place must be reserved for the structural information between the constituents, thus the size of the constituents must be reduced.

Altmann (1980) gave the theoretical derivation and the corresponding differential equation of MA law in his seminal “Prolegomena to Menzerath’s Law”.

$$\frac{y'}{y} = -c + \frac{b}{x}$$

The solution to this differential equation is the function:

$$y = ax^b e^{-cx} \quad (1)$$

Where y is the mean size of the immediate constituents and dependent variable, x is the size of the construct and independent variable, a , b and c are parameters. For example, word length and average syllable length which constitutes word are x and y respectively.

It was proved that one of the two parameters of function, b or c , can be set to 0. Then a simplified form is obtained, either $y=ax^b$ or $y=ae^{-cx}$. A large number of observations showed that

parameter b is close to zero when lower levels of language were investigated whereas higher levels lead to very small values of parameter b ; only on intermediate levels, such as word length in morphs and morph length in syllables the full formula is needed (Köhler 2012).

This paper aims to study the relationship between the Chinese compound sentence and clauses which constitute the sentence based on MA law. The stylistic factor is taken into account in the study.

1.1 Literary review

Quantitative linguistics, which the MA law belongs to, is one of main streams of Mathematical linguistics. The first serious attempts to bring a quantitative aspect into the linguistics started to rise at the end of 19th and at the beginning of 20th century. In that time mathematics started to be used within other scientific disciplines, therefore it found its way into the linguistics as well. Quantitative analysis applied to linguistic research was first adopted by the English mathematician De Morgan in 1851.

Deeper insights to linguistics need the sharper instrument and improved conceptual means. The history of all sciences shows that the best conceptual instrument or analytical means are just the mathematical method Altmann (1997). Köhler (2012) also pointed out successes of modern natural sciences all derive from their instruments and their modern models. This implies that these instruments and models, of which the quantitative parts of mathematics (probability theory and statistics, function theory, differential equation) are indispensable ingredients, are worth integrating into linguistics.

In Chinese, mathematical and quantitative aspects of language was observed and considered. For example, Huang (1989) pointed out that the study of mathematical properties of natural languages has both theoretical and application implication. Feng (2012) and Liu & Huang (2012) discussed the theory and methodology of quantitative linguistics, especially the application of this law to the Chinese language.

As the important of quantitative linguistics law, the MA law has attracted considerable scientific interest for its existence in linguistic organizations at various levels.

For example, Gerlach (1982) investigated whether the correlation between word length and the constituting morpheme length measured in phonemes follows Menzerath's law. The investigations of Rothe (1983) showed that Menzerath's law is also valid in the semantic context.

The investigations of Hřebíček (1989, 1995) showed that the relation between the length of texts and the length of the text forming sentences measured in words can be described by the MA law.

The MA law has been tested on data from many languages and on various levels of linguistic investigation. On the sentence level, however, not too many studies have been done and the investigated results are not always comparable (Köhler 2012).

Köhler (1982) conducted the first empirical test of the MA law on the sentence level as far as we know. In this investigation, the highest constructs are sentences, their length being measured in the number of their constituents (i.e. clauses). The mean length of the clauses of a sentence was calculated as the number of words of the given sentences divided by the number of clauses. The tests on the data confirmed the validity of the law with high significance.

Teupenhayn & Altmann (1984) studied the relationship between the length of sentences and its clauses measured in words, which showed that their empirical data meets Menzerath's law.

Another study (Heups 1983) evaluates sentences from texts separated with respect to text genre. Her results confirmed this law also with high significance.

Up to now the following particular cases have been examined (see Table 1)¹.

Table 1: Constructs and components examined up to now

Construct	Measured in number of	Dependent variable
Hreb	Sentences	Sentence length
Sentence	Clauses	Clause length
Sentence	Words	Word length
Sentence	Syllables	Syllable duration
Sentence	Syllables	Syllable length
Rhythmic unit	Syllables	Syllable duration
Word	Morphs	Morph length
Word	Syllables	Syllable length
Word	Sounds	Sound duration
Syllable	Sounds	Sound duration
Word and syllable	Syllable and Sounds	Syllable duration
Chinese characters	Graphemes	Grapheme complexity

In the analysis of German literary prose texts, Arens observed the phenomena in which word length increases along with an increase of sentence length (Arens 1965). By way of a solution, Altmann pointed out that the MA law is likely to hold true only when one is concerned with the immediate components of a given construct (Altmann 1983).

Except the application of the MA law to different language levels, some research studied the theory and formula per se. For example, Cramer (2005) and Baixeries et al (2013) discussed the parameters of the formula respectively.

Until now, most of works focused on the application of the MA law to Indo-European language, while only a few works applied the MA law to Chinese. An important reason to lead to this phenomenon is that it is difficult to define the sentence and to determine the components of the sentence in Chinese. To our knowledge, first work applied the MA law to Chinese is Bohn (1998, 2002). He analyzed the relationship between Chinese characters and the complexity of composing graphemes, length of words and simplicity of characters, clause length and word length, sentence length and clause length. In his study relationship between sentence length and clause length only in news texts was examined. Tereza Motalová et al (2013, 2014) and Ščigulinská & Schusterová (2014) verified the validity of the MA law applied to contemporary written and spoken Chinese respectively.

Different from them, this study divided the sentence into two kinds which are compound sentence and simple sentence. Based on the precondition, this study explored whether the mutual relationship between compound sentence and clauses conforms to the MA law and attempt to explain them from the perspective of theoretical linguistics. As Huang (2000) pointed out that the contribution of empirical approaches can't be theoretically significant unless the quantitative data can be generalized and interpreted as a qualitative account of the linguistic facts.

¹ Quoted from http://lql.uni-trier.de/index.php/Hierarchic_relations

In the mean time, stylistic factors are taken into account in this study. Style is the speech variant of language and the usage of language means is the material basis of style (Wang and Chen 2000). Feng (2010) thinks that style is a polarized opposite continuum, with formal written style and daily colloquial style in two poles, and other styles lying in between. There are some different style category schemes, the three styles are considered in this paper.

There are lots of studies for comparing the different styles. Many studies gave evidence to show that the distributed differences of linguistic structures in different styles are objective (Biber 1988, Stamatatos et al., 2001). Biber (1995) pointed out that the use of statistical methods in style processing has proved to be a reliable approach. Huang (2014) pointed out that there are usage differences of light verbs between Chinese language variation in mainland China and Taiwan. This paper, taking style factor into account, explores whether there are differences of correlation between compound sentence and constituting clauses in the different stylistic texts.

1.2 Research question and methodology

This paper examine whether the relationship of the compound sentence and the constituting clauses conforms to MA law in different stylistic texts.

Indo-European languages, sentence is defined as a language fragment with subject and predicate, or the language unit of NP+VP. Different from Indo-European languages, a Chinese sentence is only defined from the perspective of speech (Lu 1993). Chao (1968) defined a sentence as an utterance with pauses at the beginning and at the end. For Zhu (1982) a sentence expresses relatively intact meanings with intonation and pause at the beginning and the end.

Considering the complexity of Chinese sentence and the aim of this paper, the operational concept of sentence was defined from the perspective of written form: the sentence is defined the word sequence separated by periods, question marks and exclamation points. Because the two word sequences separated by colons are complete sentence forms, so it is considered the sentence marks. Sentence is composed of single sentence and compound sentence. Sentences in which there are no pauses, i.e. commas, are called as single sentences. Otherwise they are called compound sentences. The components of compound sentence are called clauses.

The length of compound sentence was measured by the number of constituting clauses, and the length of clause was measured by the number of constituting words.

The frequency distribution of compound sentence length measured by the number of constituting clauses was established. Regression analysis was used to fit the correlation between compound sentence length and clause length.

Sentence is the higher linguistic level, function $y = ax^b e^{-cx}$ was selected to fit the correlation between compound sentence length and clause length.

R language was used to realize all the works of data analysis.

2. Corpus Establishment and Preprocessing

The texts of “*News Co-Broadcasting*”, the comedy “*I Love My Family*”, and “*Behind the Headlines with Wentao*” were selected to represent the written formal, TV conversational and daily conversational styles respectively.

The program of CCTV (Central China TV), *News Co-Broadcasting*, mainly gives a brief introduction to important state policies and events taking place at home and abroad. It characterized by the formal, serious and solemn use of language.

Behind the headline with Wentao, as a program of Phoenix satellite TV, the host discusses some current hot issues together with guest in TV. They talk freely, chatting so as to deliver recreational information, create fun and discriminate truth from falsehood, not focusing on the “right answers” to the issues. They do not read the scripts edited ahead of time. It can be represent TV conversational style.

The comedy taking place in family, *I Love My Family*, tells the story through the dialogue. So it can be seen as representation of daily conversation.

The textual materials of *News Co-broadcasting* are obtained from the Language Resources Monitoring and Research Centre. Textual materials of *Behind the Headlines with Wentao* are collected from the website of Phoenix Satellite TV. The texts of My Love My Family were downloaded from the Internet, each one of the 120 episodes was treated as a text.

The Chinese lexical analysis system created by Institute of Computing Technology of Chinese Academy of Science (ICTCLAS) was used for word segmentation and part-of-speech tagging. ICTCLAS has been acknowledged with a high accuracy of 97.58%, a recall rate of over 90% for the recognition of unknown words based on role tagging, and a recall rate of approximate 98% for the recognition of Chinese names.

The segmented and tagging texts were manually checked in order to correct the errors of lexical analysis. For example, words in the bracket pair were deleted if they were explanatory notes in “*Behind the Headlines with Wentao*” because we do not consider explanatory notes as the parts of texts of different style.

The scales of these three stylistic texts are about 450,000 words respectively, total of which are 130,0000 words.

The ratio of the number of simple and compound sentences are the follows, as shown in table 2.

Table 2: Number and Percentage of Two Kinds of Sentences in the Texts

	Simple sentences	Compound sentences	total
<i>News Co-Broadcasting</i>	6295 (31.96%)	13400 (68.04%)	19695
<i>Behind the Headlines with Wentao</i>	4430 (32.39%)	9248 (67.61%)	13678
<i>I Love My Family</i>	13030 (47.83%)	14210 (52.17%)	27240

3. Results and discussion

3.1 Frequency distribution of compound sentence length

Figure 1 and Figure 2 showed the frequency distribution and relative cumulative frequency of compound sentence length. From these two figures, we can see that the frequencies of compound sentences with different lengths are very asymmetrical.

There are similar decreasing tendency of compound sentences frequencies along with the increase of the compound sentence length.

The relative frequencies of compound sentences composed of 2 clauses are 43.5% and 37.2% in daily conversation and news co-broadcasting respectively. The relative frequency of compound sentences of same length in TV conversation is minimum, 23.4%.

The frequent interaction in daily conversation is the reason of more relative frequency of compound sentences of 2 clauses. In part 2, same reason causes the high ratio of simple sentences in daily conversation.

In News Co-broadcasting, people hope to convey more information in short time. So compound sentences lengths are small and concentrate more on 2, 3, 4 clauses. From Figure 2, the length of 90% of compound sentences is less than 5 clauses.

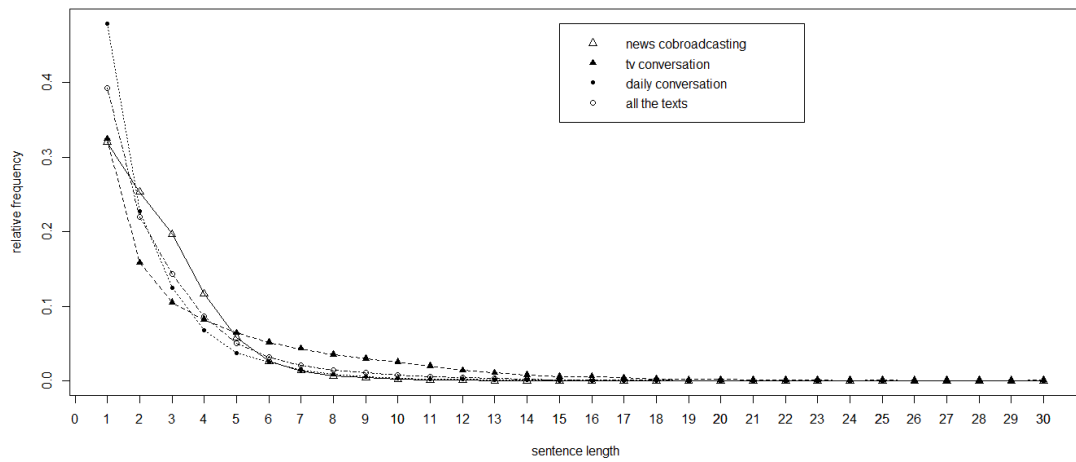
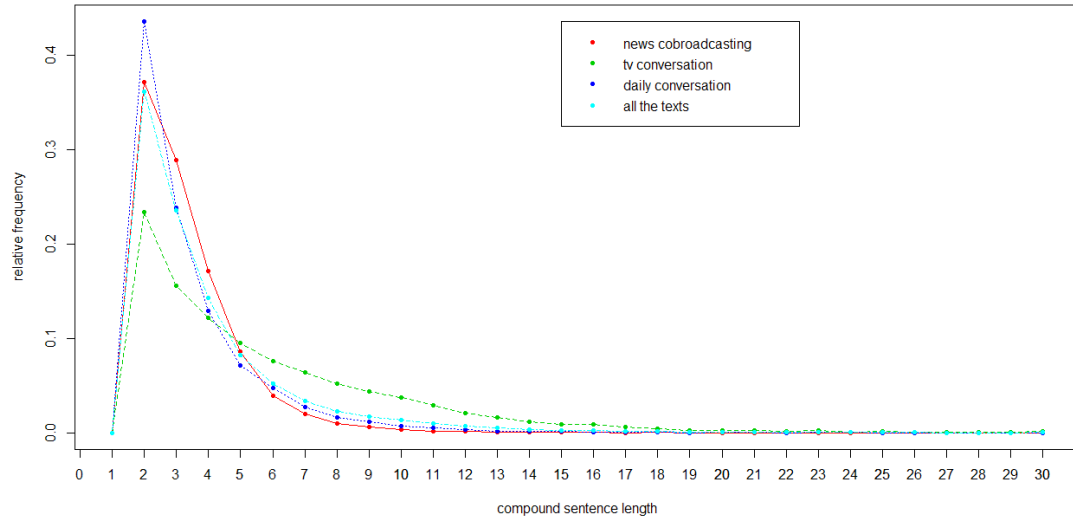


Figure 1: Frequency Distribution of compound sentence length

R 中 fflen 表示以所包含分句个数表示的句子长度

There is a quite difference in the frequency distributions of compound sentence length in TV conversation and daily conversation although they are conversations. In daily conversation, the length of 80% of compound sentences is less than 4 clauses. Meanwhile the compound sentence length needs to be 8 and 12 clauses when the ratio of them reaches 80% and 90%.

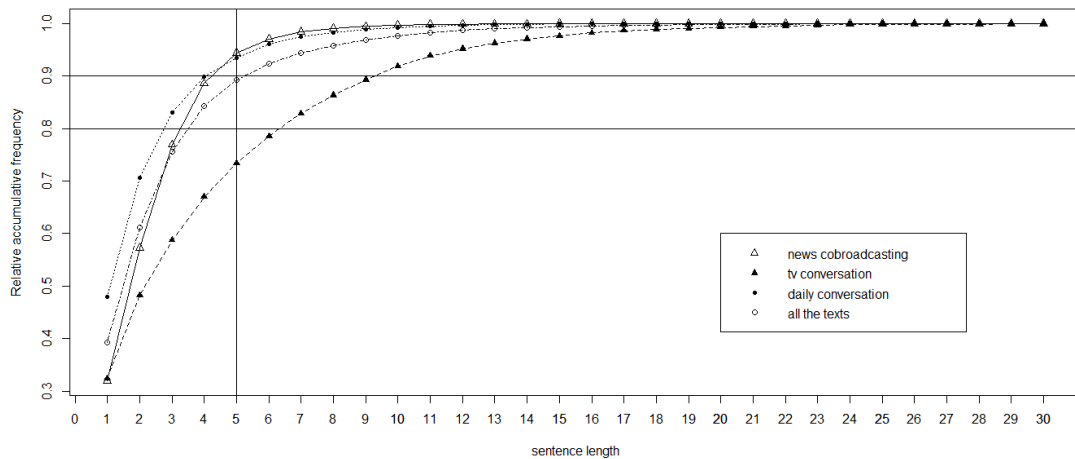
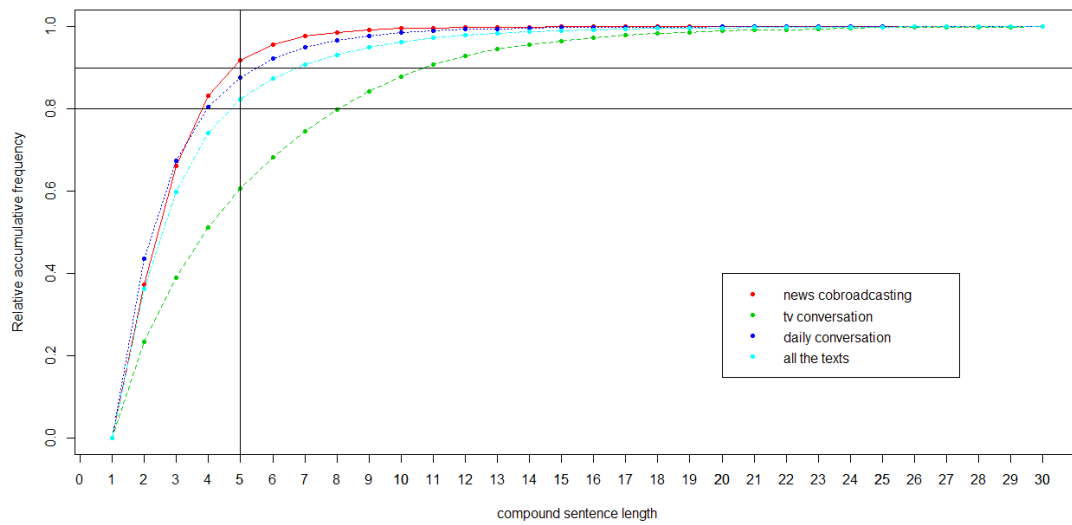


Figure 2: Relative cumulative frequency curves of compound sentences Length

3.2 Relationship of compound sentence and clauses

Köhler (1982) determined the number of clauses by counting the number of finite verbs in a sentence. Different from Indo-European language, Chinese clauses are marked by colons included in the compound sentence.

Since it is unnecessary to determine the lengths of the individual clause, the mean length of the clauses of a sentence was calculated as the number of words of the given sentence divided by the number of clauses.

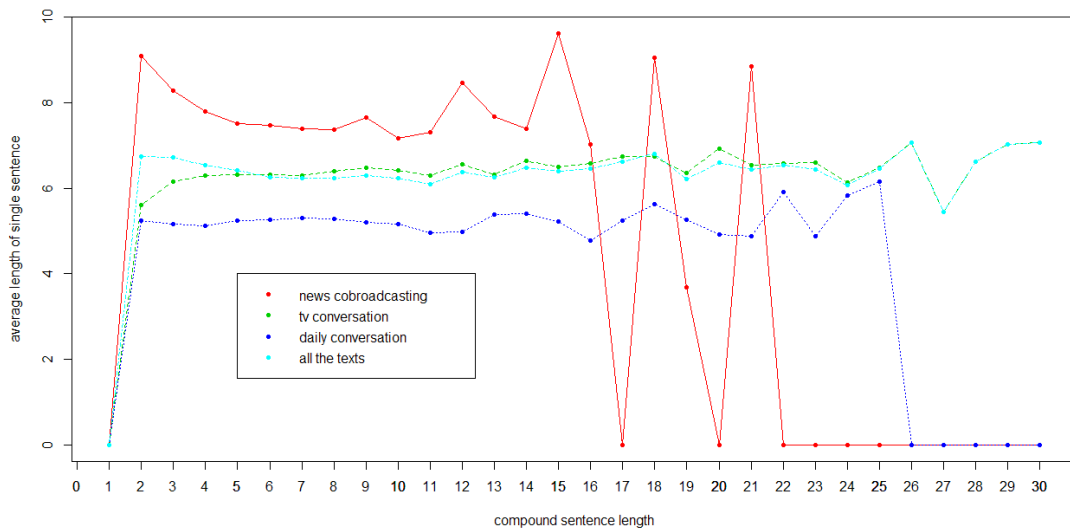


Figure 3: Correspondence of average clause length and compound sentence length

The correspondence between mean length of clauses and compound sentence length was established, as shown in Figure 3. The mean clause length is the dependent variant and the corresponding compound sentence is the independent variant.

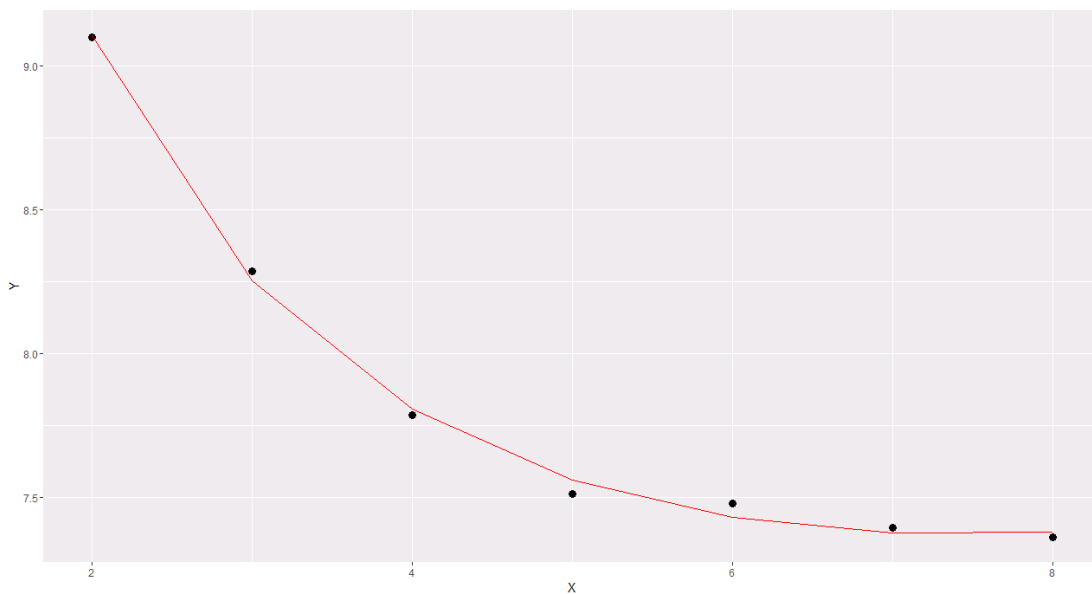


Figure 4: Fitted Result of Average Clause Lengths and Compound Sentence Length in News Co-Broadcasting ($A=10.6231$, $b=-0.3631$, $c=-0.0488$, Residual Standard error=0.0145)

Figure 3 shows that the change of the mean clause length is not regular along with the increase of compound sentence lengths. From Figure 1, we know that the frequencies of compound sentences are decreasing along with the increase of compound sentence length. When relative cumulative frequency of compound sentences reach 95%, the compound sentences length of these texts (*News Co-BroadCasting*, *TV Conversation*, *Daily Conversation*, total) are 6, 13, 7, 9 clauses respectively. We think that these compound sentences can represent all the compound sentences. Whether the relationship of compound sentences and clauses conforms for MA law was

examined.

Function $y = ax^b e^{-cx}$ was used to fit the correspondence of the average clauses lengths and compound sentences lengths of *News-CoBroadcasting*. The fitted result is shown in Figure 4.

From the fitted result, we can see that the average length of clauses is decrease along with the increase of compound sentence lengths. Good fitted result, residual standard error is 0.0145, showed that the relationship between compound sentence and constituting clauses, conforms to the MA law in *News Co-BroadCastring*. So the clauses can be seen as the immediate components of the compound sentences.

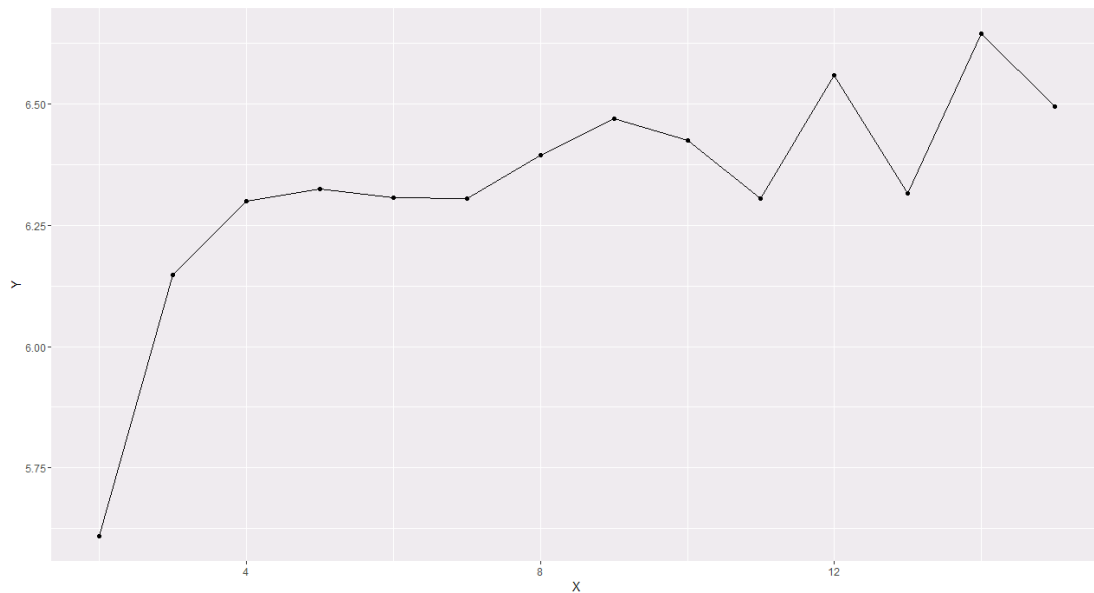


Figure 5: Correspondence between average clauses lengths and compound sentence lengths in TV conversation

From figure 3, we can see that the average length of clauses is increasing along with the increase of the compound sentence length in TV conversation. To display this relationship more clearly, this relation line chart was highlighted separately in Figure 5. From figure 5, there is not consistent change tendency of average clauses lengths in compound sentences whose lengths are from 10 to 15 clauses. In compound sentences of length with 2 to 10 clauses the average clause lengths are increasing along with the increase of compound sentences lengths. This relationship does not conform to MA law obviously. From figure 6, the smooth curve of figure 5, we also can see that the average clause lengths are increasing along with the increase of compound sentence length.

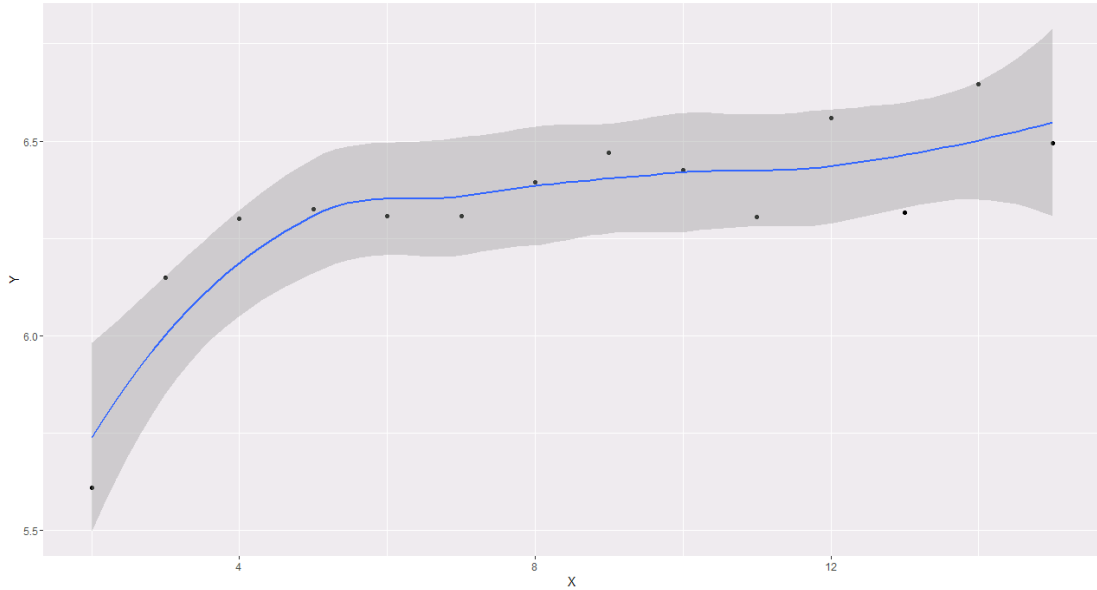


Figure 6: Smooth curve of relationship between average clauses lengths and compound sentence lengths in TV conversation

The same method was used to examine the relationship between the average clause lengths and compound sentence lengths in daily conversation. The line chart and smooth curve of this correspondence are shown in figure 7.

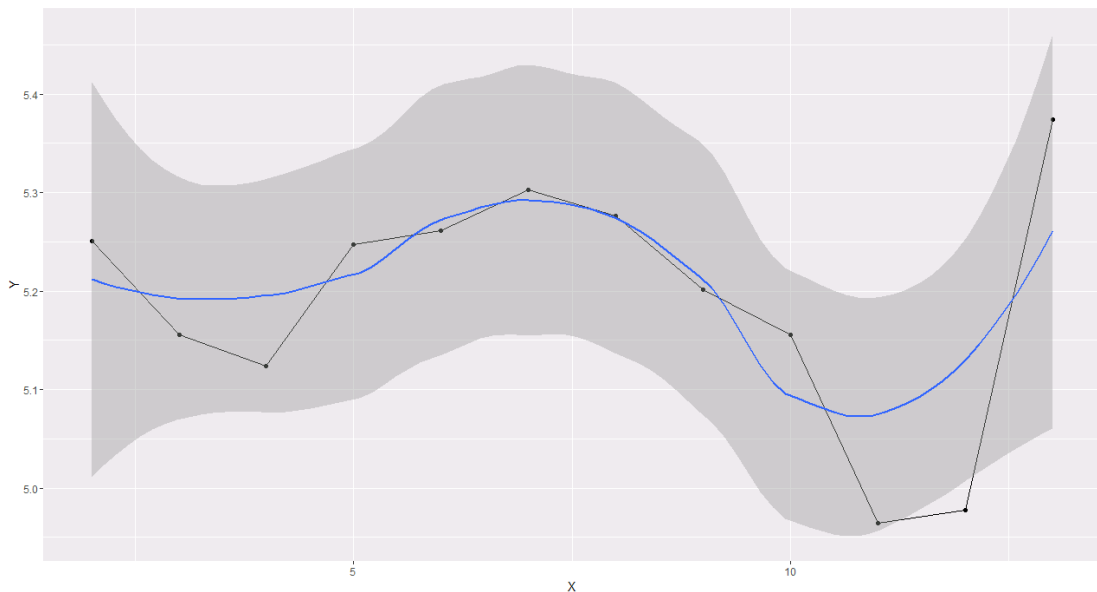


Figure 7: Correspondence between average clause lengths and compound sentence lengths in daily conversation

In Figure 7, the blue line stands for the smooth curve of the correspondence. From this figure, we can see that the relationship between average clause lengths and compound sentence lengths does not conform to the MA law in daily conversation too.

The similar analysis for all the texts has shown that the relationship between average clauses lengths and compound sentence lengths does not conform to the MA law too.

From the above analysis, we can see that relationship between average clause lengths and compound sentence lengths conforms to the MA law only in *News Co-Broadcasting* which represents written formal language, which should be further proved.

Although TV conversation and daily conversation are different in some aspects, they have similarity and are both the spoken languages. Thus it can be said that the relationship between clause lengths and compound lengths does not conform to the MA law in spoken language.

According to Chao (1968), “Chinese sentences may be classified into full and minor sentences. A full sentence is made up of two parts, a subject and a predicate, and is the common type in connected discourse. A minor sentence is not in the subject-predicate form. It occurs more frequently in two-way conversation and in speech interposed or accompanied by action than it does in connected discourse. The full sentence is more likely to be used in declarative sentences and questions, while a minor sentence is more likely to be used for commands, vocatives and responses, and exclamations”. Although minor sentences have no subject or predicate, they can express a whole meaning. Because a succession of minor sentences does not necessarily make a full sentence, a Chinese discourse is often made up of a series of “flowing sentences” as defined by Lv (1979). Flowing sentences causes that the relationship between compound sentence lengths and clauses does not conform to the MA law in spoken language. So the clauses are not the immediate components of compound sentences in spoken language. Presumably, the increasing tendency of the average clause length along with the increase of compound sentence length was caused by more quickly speaking.

3.3 Test of Hypothesis

This part attempts to further prove the above conclusion on the Lancaster Corpus of Mandarin Chinese (LCMC) that became available in 2003 (McEnery and Xiao 2004). The corpus includes 500 random samples, from 15 written registers, of 2000 word tokens, totaling 1,000,000 words which were taken from publications from mainland China between 1988 and 1992.

We are not sure that the texts of some registers, for example, science fiction and romance fiction, all belong to written formal stylistics. Thus the texts of News editorials, News reviews, Religion, Skills, trades and hobbies, Science (academic prose), Reports and official documents were selected which are considered the written formal style. These texts include 400,000 words. In LCMC, the sentences are marked by the punctuations which were tagged by “ew”.

Table 3: The observed and theoretical value of mean clause length

Sentence length	Observed values	Theoretical values (2 parameters, $y=ax^b$)	Theoretical values (3 parameters, $y=ax^be^{-cx}$)
2	8.10	8.07	8.09
3	7.59	7.61	7.60
4	7.22	7.30	7.28
5	7.09	7.07	7.06
6	6.98	6.89	6.89
7	6.69	6.74	7.76

From Table 3, the mean clause length decreases along with the increase of sentence length. The correlation between compound sentence length and clause length was fitted using $y=ax^b$ and

$y=ax^be^{-cx}$ respectively. The fitted results were shown in figure 9 and figure 10. In fitted results using $y=ax^b$, the adjusted R-squared and residual standard error of the fitted results are 0.9788 and 0.009854 respectively.

In texts, the lengths of the 97.13% of compound sentences are less than 8 clause, we think that the relationship between compound sentence and clause are be observed in these 97.13% of compound sentences. The observed mean clause length and the length of corresponding compound sentence are listed in Table 3.

From the values of residual standard error of these two fitted results, the fitted result using $y=ax^b$ are better. However, Table 3 showed that the theoretical values of mean clause length of fitted result using $y=ax^be^{-cx}$ are more similar to observed values.

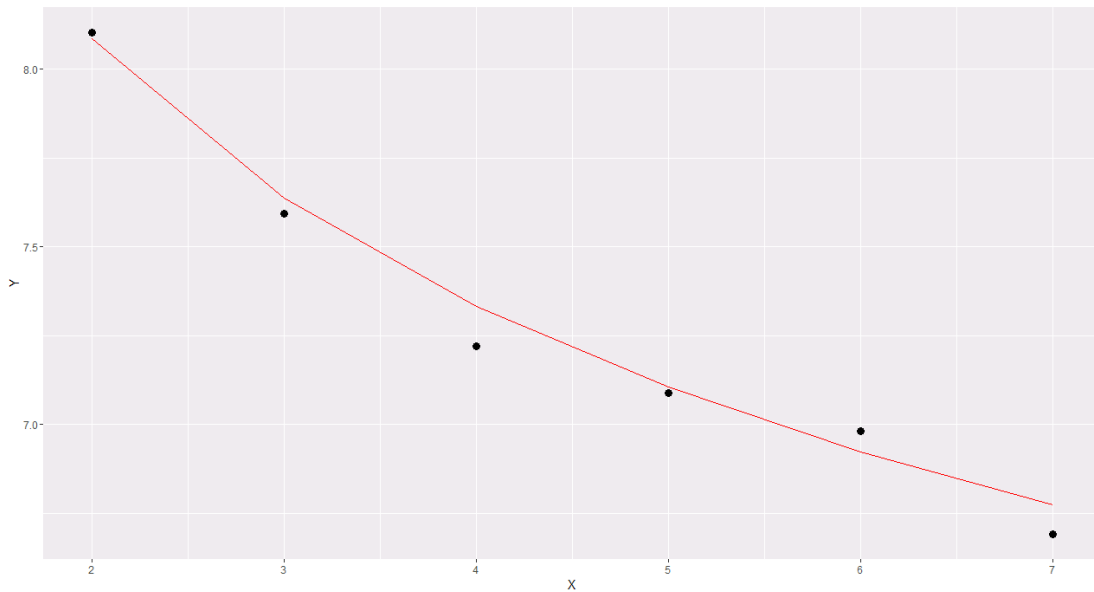


Figure 9: The fitted results of mean clause length using $y=ax^b$ ($a=8.92$, $b=-0.1441$)

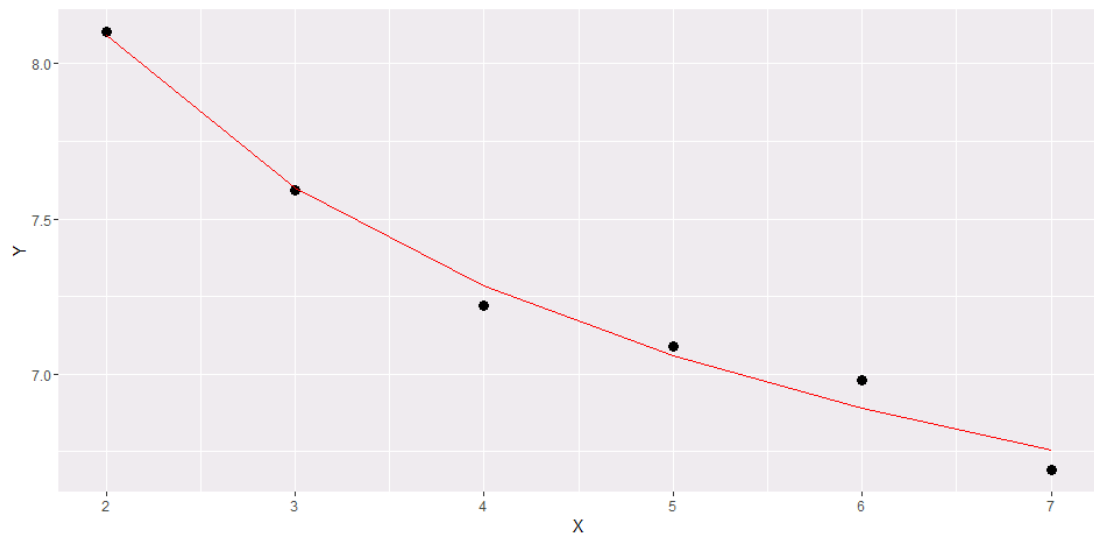


Figure 10: The fitted result of mean clause length by $y=ax^be^{-cx}$, ($a=8.99$, $b=-0.17$, $c=-0.01$, residual standard error= 0.0774)

Next, the correlation between compound sentence and clauses of Science (academic prose)

was fitted using $y=ax^b$. Table 4 showed the observed values and theoretical values of mean clause length. Figure 11 showed the fitted curve of mean clause length. The adjusted R-squared and residual standard error of the fitted results are 0.9909 and 0.00612 respectively, which means that the fitted result is better.

Table 4: The observe values and theoretical values of mean clause length in science texts

Sentence length	Observed values	Theoretical values (2 parameters, $y=ax^b$)
2	8.67	8.65
3	8.14	8.17
4	7.84	7.86
5	7.63	7.62
6	7.50	7.44
7	7.24	7.28

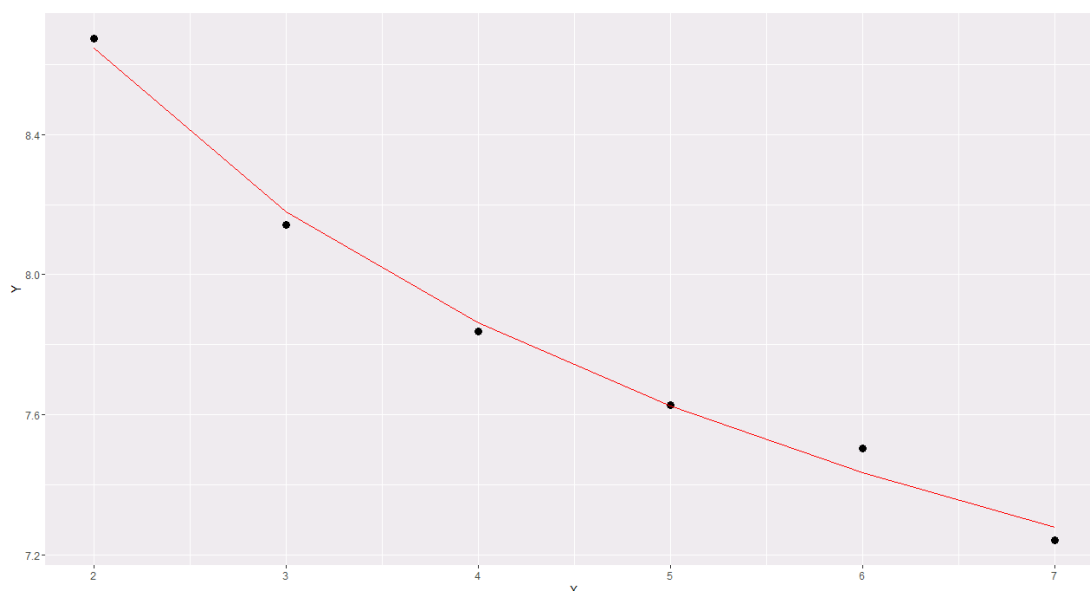


Figure 11: The fitted curve of mean clause length in science texts ($a=9.5087$, $b=-0.1371$)

From these two fitted results, the latter is better than the former because the texts of the latter experiment are only the science texts presumably.

Above all, it was confirmed that the correlation between Chinese compound sentence and clauses conforms to the MA law only in written formal stylistics.

So, we can't discuss whether it conforms to the MA law generally about the relationship between compound sentence and constituting clauses. We should take the stylistic factor into account. Only so, we can get more detailed knowledge and avoid the wrong conclusion about language.

4 Conclusions

Because of the uniqueness of Chinese language, Chinese sentence can't be given an explicit definition just as Indo-European language. Operational definition of Chinese sentence was the premise of this paper and was defined as a word sequence between two punctuations which are

periods, question marks, exclamation marks and colons.

Chinese sentence was divided into compound sentence and simple sentence in this study in order to examine whether the correlation of compound sentence length and mean constituting clause length conformed the MA law.

The MA law pointed out that “the longer a language construct the shorter its components”. This paper analyzed the correlation of Chinese compound sentence and the clauses included in compound sentences based on the MA law. The results showed that the relationship does not conform to the MA law when all the texts of different styles are considered as a whole.

Previous researches provided the theoretical basis for taking style factor into account in this study. For example, Lv (1979) pointed out that restraining force of Chinese grammars are not strong because the style factor is not taken into account when we summarized them. Tao (1999, 2007) and Zhang (2007, 2012) both pointed out that there are different Chinese grammar in different style and Chinese grammar should be studied according to style. Then stylistic factor can help to study the Chinese language and should be taken into account in this study.

The results showed that the relationship does not conform to the MA law in spoken language, while it conforms to the law in the written formal stylistic texts, for example *News CoBroadcasting*. The clauses, minor sentences, can be regarded as the immediate components of compound sentences in written formal texts. Rigorous and standard expression needs some clauses to express a complete meaning in written formal language. In spoken language, most of compound sentences are composed of minor sentences, with complete meaning and used independently. So the correlation between compound sentence and constituting clauses does not conform to the MA law and clauses are not the immediate components of compound sentences.

Reference

- Altmann, G. (1980). Prolegomena to Menzerath's law. *Glottometrika* 2, 1-10.
- Altmann, G. (1983). H. Arens' «Verborgene Ordnung» und das Menzerathsche Gesetz. In: M. Faust et al. (Eds.), *Allgemeine Sprachwissenschaft, Sprachtypologie und Textlinguistik*: 31-39. Tübingen: Narr.
- Arens, H. (1965). *Verborgene Ordnung. Die Beziehungen zwischen Satzlänge und Wortlänge in deutscher Erzählprosa vom Barock bis heute*. Düsseldorf: Pädagogischer Verlag Schwann.
- Bohn, H. (1998). *Quantitative Untersuchungen der modernen chinesischen Sprache und Schrift*. Hamburg: Kovač.
- Bohn, H. (2002). Untersuchungen zur chinesischen Sprache und Schrift. In: Köhler, R. (ed.), *Korpuslinguistische Untersuchungen in die quantitative und systemtheoretische Linguistik*: 127-177. <http://ubt.opus.hbz-nrw.de/volltexte/2004/279/>
- Chao, Yuen Ren. (1968). *A Grammar of Spoken Chinese*. Berkeley and Los Angeles: University of California Press.
- DeChun Wang, RuiDuan Chen. YUTIXUE[M]. Guangxi Education Press. 2000.12:6
- Feng Shengli. (2006). On the properties and pedagogy of Chinese written formal style. *TEACHING IN THE WORLD*, 4 (98), 106.
- Feng Shengli. On mechanisms of Register System and its grammatical property [J]. *Studies of the Chinese Language*. 2010(5):400-412
- Feng Zhiwei. (2012). 用计量方法研究语言. *Foreign Language Teaching and Research*, 44(2),

256-269.

Gabriel Altmann (1997) The Art of Quantitative Linguistics, *Journal of Quantitative Linguistics*, 4:1-3, 13-22.

Gerlach, R. (1982). Zur Überprüfung des Menzerathschen Gesetzes im Bereich der Morphologie. *Glottometrika*, 4, 95-102.

Heups, G. (1983). Untersuchungen zum Verhältnis von Satzlänge zu Clauselänge am Beispiel deutscher Texte verschiedener Textklassen. In R. Köhler & J. Boy (Eds.), *Glottometrika* 5 (pp. 113 – 133). Bochum: Brockmeyer.

Hřebíček, L. (1989). Menzerath-Altman's law on the semantic level. In L. Hřebíček (Ed.), *Glottometrika* 11 (pp. 47 – 56). Bochum: Brockmeyer.

Hřebíček, L. (1992). Text in communication: Supra-sentence structure. Bochum, Brockmeyer.

Hřebíček, L. (1995). Text levels: Language constructs, constituents and Menzerath-Altman law. Trier: WVT.

Hřebíček, L. (1997). Lectures on text theory. Prague: Oriental Institute.

Huang, C. R., Chen, K. J., & Gao, Z. M. (1998). Noun class extraction from a corpus-based collocation dictionary: An integration of computational and qualitative approaches. *Quantitative and Computational Studies of Chinese Linguistics*, 339-352.

Huang, C. R. (2000). From quantitative to qualitative studies: Developments in Chinese computational and corpus linguistics. *漢學研究*, 第十八卷特刊, 473-509.

Irene Cramer. (2005). The Parameters of the Altmann-Menzerath Law, *Journal of Quantitative Linguistics*, 12:1, 41-52.

Huang, C. R., Lin, J., Jiang, M., & Xu, H. (2014). Corpus-based Study and Identification of Mandarin Chinese Light Verb Variations. *COLING 2014*, 1.

Jume Baixeries , Antoni Hernández-Fernández , Núria Fornés & Ramon Ferrer-i-Cancho (2013) The Parameters of the Menzerath-Altman Law in Genomes, *Journal of Quantitative Linguistics*, 20:2, 94-104.

Köhler, R. (1982). Das Menzerathsche Gesetz auf Satzebene. In W. Lehfeldt & U. Straus (Eds.), *Glottometrika* 4 (pp. 103 – 113). Bochum: Brockmeyer.

Köhler, R. (1989). Das Menzerathschen Gesetz als Resultat des Sprachverarbeitungsmechanismus. In: Altmann, Schwibbe (1989): 108-112.

Köhler, R. (2012). *Quantitative syntax analysis* (Vol. 65). Walter de Gruyter.

Liu Haitao & Huang Wei. (2012). Quantitative Linguistics: State of the Art, Theories and Methods. *Journal of Zhejiang University (Humanities and Social Sciences)*. 42(2). 178-192.

Lu Jianming. (1993). The features of Chinese sentences. *Chinese Language Learning*. No.1, 1-6.

Lv Shuxiang. (1979). Chinese parsing problem. Commercial Press.

McEnery, Tony & Richard Xiao. 2004. The Lancaster corpus of Mandarin Chinese: A corpus for monolingual and contrastive language study. In M. Lino, M. Xavier, F. Ferreire, R. Costa, R. Silva (eds.), *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC) 2004*, pp. 1175 – 1178. Lisbon, May 24 – 30, 2004.

Menzerath, P. (1954). *Die Architektonik des deutschen Wortschatzes* (Vol. 3). F. Dümmler.

Rothe, U. (1983). Wortlänge und Bedeutungsmenge. Eine Untersuchung zum Menzerathschen Gesetz an drei romanischen Sprachen. *Glottometrika*, 5, 101-112.

Ščigulinská & Schusterová (2014) “*An Application of the Menzerath-Altman Law to Contemporary Spoken Chinese*”. Palacký University in Olomouc. First Published 2014.

- Stamatatos, E., Fakotakis, N. and Kokkinakis, G. (2001). Automatic text categorization in terms of genre and author. *Computational Linguistics*, 26(4): 471–95.
- Tao Hongyin. (1999). Grammar Meaning of Style Category. *Contemporary Linguistics*, (3), 15-24.
- Tao Hongyin. (2007). The discourse pragmatics of argument structure in Chinese procedural discourse. *Chinese Language*. (1). 3-13.
- Tereza Motalová et al. (2013, 2014). An application of Menzerath-Altmann law to contemporary written Chinese.
- Teupenhayn, R., & Altmann, G. (1984). Clause length and Menzerath's law. *Glottometrika*, 6, 127-138.
- Zhang Bojiang. (2007). Stylistic differences and grammar rule. *Rhetorical study*, 2, 1-9.
- Zhang Bojiang. (2012). Research on style for the purpose of syntax interpretation. *Modern Rhetoric*, 6.
- Zhu Dexi. (1982). *Grammar handouts*. Commercial Press.