RESEARCH ARTICLE

How does language change as a lexical network? An investigation based on written Chinese word co-occurrence networks

Heng Chen¹, Xinying Chen², Haitao Liu^{1,3,4*}

1 Center for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, Guangzhou, China, **2** School of Foreign Studies, Xi'an Jiaotong University, Xi'an, China, **3** Department of Linguistics, Zhejiang University, Hangzhou, China, **4** Ningbo Institute of Technology, Zhejiang University, Ningbo, China

* lhtzju@yeah.net

Abstract

Language is a complex adaptive system, but how does it change? For investigating this process, four diachronic Chinese word co-occurrence networks have been built based on texts that were written during the last 2,000 years. By comparing the network indicators that are associated with the hierarchical features in language networks, we learn that the hierarchy of Chinese lexical networks has indeed evolved over time at three different levels. The connections of words at the micro level are continually weakening; the number of words in the meso-level communities has increased significantly; and the network is expanding at the macro level. This means that more and more words tend to be connected to medium-central words and form different communities. Meanwhile, fewer high-central words link these communities into a highly efficient small-world network. Understanding this process may be crucial for understanding the increasing structural complexity of the language system.

OPEN ACCESS

Citation: Chen H, Chen X, Liu H (2018) How does language change as a lexical network? An investigation based on written Chinese word co-occurrence networks. PLoS ONE 13(2): e0192545. <https://doi.org/10.1371/journal.pone.0192545>

Editor: Irene Sendiña-Nadal, Universidad Rey Juan Carlos, SPAIN

Received: September 1, 2017

Accepted: January 25, 2018

Published: February 28, 2018

Copyright: © 2018 Chen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by the National Social Science Foundation of China (Grant No. 12&ZD224) to HC, Fundamental Research Funds for the Central Universities Program of Big Data PLUS Language Universals and Cognition, Zhejiang University, and the National Social Science Foundation of China (Grant No. 17AYY021) to HL, and the MOE Project of the Center for Linguistics and Applied Linguistics, Guangdong University of

Introduction

Language is the main evolutionary contribution of humans [1, 2], and is intimately interwoven with cultural evolution and our superior mental functions [3]. Many studies show that language is a complex adaptive system (CAS) [4, 5, 6, 7], possessing features such as hierarchical structures in syntactic organizations [8, 9], scale-free and small-world properties in language networks [10], and the well-known “principle of least effort” of Zipf’s law [11]. As complex systems, “all languages are complex to some degree” [12, 13], and “the older a language, the more complexity it has” [14]. The question we would like to address here is how a language as a complex system, to be specific, lexical network, changes over time.

Hierarchy is considered to be an important feature of complex systems [15], and the understanding of the relationship between hierarchies has been a key concern in complex system research. In the latest approach—namely, complex network analysis—hierarchies are even considered one of the key elements that lead to the complexity of network systems, such as small-world and scale-free features [16]. Simon [17] proclaims that 1) the complex system

Foreign Studies to HC and HL. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

must be hierarchical; 2) the origin of the hierarchical structure is the interaction between agents; and 3) adaptation and hierarchy are implicitly coupled features.

Modern linguists also believe that the hierarchical nature of language is one of the essential attributes of languages [8, 18]. Existing studies discussing the language hierarchies are multifarious and are mainly concerned with problems that occur in at a certain hierarchical level, such as understanding the patterns of combining words to phrases. In contrast, the CAS theory recognizes the relativity and complexity of language hierarchies, thereby bringing fresh insights to the question. It can address the macro and micro features simultaneously, and balance the discussions on both sides as well [19, 20]. Moreover, the CAS theory treats the states of the system as parts or results of an evolving process, which provides a new way for us to understand the adaptability and complexity of the language system. Therefore, the evolution of the language hierarchy feature can be seen, also, as the process of language adaptation. However, it needs to be noticed that this hierarchy feature is not equivalent to the syntactic hierarchy feature in Linguistics. The hierarchy feature we explored here is from the perspective of local connectivity, modular community structure, and the global connectivity of a network, e.g., as applied in [21, 22, 23]. Moreover, the classical hierarchy feature of Degree Dependent Clustering Coefficient in language [24] is also observed.

Complex network analysis is a very effective tool for studying complex systems [19], and it has been increasingly applied in linguistic studies (for a review, see [10, 25]). A complex network is a model of a system [26] and dynamic network constitutes a large part of CAS theory [27]. In the current study, complex network analysis offers a feasible way of discussing the hierarchy question in languages from a more systematic way, which describes the macroscopic and microscopic levels of the language system simultaneously [20].

Applying complex network models to study the dynamic features of language systems has become popular in recent years in fields such as language evolution/change and language development [2, 28, 29], quantitative linguistics [30, 31, 32, 33, 34], and language acquisition [21, 35, 36]. With regard to word co-occurrence networks, there have been many studies [29, 30, 35, 37–39]. Among them, Ferrer-i-Cancho, Solé and Köhler [24] verified the existence of hierarchical features in the English syntactic network and Liu and Cong [40] found that hierarchical features also exist in the Chinese syntactic and semantic networks; in addition, they also found out that there are differences between the subsystems of these two kinds of networks.

However, research on the change of language hierarchy has been still absent until now. To explore how language evolves, we need to investigate a manageable number of reliable diachronic texts. Although there is usually a gap between written and spoken language, especially in Sinitic languages, most of our findings about languages, even in modern times, depend much on written records [41, 42]. In this study, we reason that written Chinese is suitable for our exploration [43, 44, 45], since it is one of the most archaic living languages for which we have a more or less continuous written record. This study constructs and analyzes four diachronic Chinese word co-occurrence networks. It aims to discover the macroscopic change patterns appearing in the Chinese language system by observing and comparing the hierarchical characteristics of four diachronic networks. Furthermore, we will combine our results with common quantitative linguistic measurements such as utterance length distribution and word length distribution and attempt to reveal and explain the causes and internal mechanism of the origin and the change of language hierarchies.

The questions that will be discussed in the current study include:

1. Has the Chinese word co-occurrence network always been a complex network? If so, then how does it change as a complex network?

2. How has the Chinese word co-occurrence network changed from the perspective of network hierarchy? And what is the linguistic significance of this change?
3. How does the length distribution of clauses and words change in written Chinese? What is the significance of this change with regard to the network change?
4. What is the most likely motivation and mechanism of the change of the Chinese word co-occurrence network?

The study is organized as follows: First, the construction method of the word co-occurrence network and the acquisition and processing of the language network parameters are introduced in Section 2. Then, the parameters of the four diachronic language networks and some related language units are compared in Section 3. Section 4 is devoted to a discussion of the results. The patterns of change and the mechanism behind it will be explored in depth in this section. The study is concluded in Section 5.

Data and methods

Data selection and preparation

For ensuring that this study reflects the real dynamic change of the Chinese language, we have chosen authentic texts that are relatively closer to the spoken language during different times in history (for details on the selected texts, see Table 1). It is critical to choose appropriate ancient texts since there is an evident distinction between two kinds of Chinese written texts: the mandarin Chinese and the classic Chinese [46]. The former is much closer to spoken language, and the latter is an independent system that did not change much since pre-Qin dynasty (about from B.C. 3 to Pre-A.D. 20). This study includes short narrative texts of mandarin Chinese from four historical periods, which are representative for language used in their times [42, 45, 47]. Specifically, the narrative texts from Time Period 1 to 3 are from <http://www.gushiwen.org/>, and the Xinwen Lianbo texts in Time Period 4 are from <http://tv.cctv.com/lm/xwlb/>. To ensure the comparability, the text length has been fixed at 10,000 tokens, which means the four networks may have different sizes. Even so, we attributed the parameter differences between four networks to the change of Written Chinese per se, to be specific, the birth, use and death of words in language change [48].

Genre is also an essential factor of texts. We only select short narrative stories from the same register for this study. Since there is no ready-made word segmentation tool for ancient Chinese, we commissioned experts on ancient Chinese to manually segment the texts. The chief problem of manual segmentation is selecting the criteria for wordhood, especially when it comes to the distinction between words and phrases [49]. Through manual proofreading and correction of the segmentation results, we ensure that the accuracy rate of word segmentation is at least 98%. The total amount of Chinese characters in each of these texts is different

Table 1. Selected texts in the corpus.

Corpus information	Corpus 1	Corpus 2	Corpus 3	Corpus 4
	Ancient	Middle Ancient	Modern Times	Modern
Name of the text	孟子 <i>mengzi</i>	颜氏家训书 <i>yanshi jiaxun shu</i>	宋元白话小说 <i>song yuan baihua xiaoshuo</i>	新闻联播 <i>xinwen lianbo</i>
Size (words)	10,000	10,000	10,000	10,000
Total of word types	1,162	1,945	1,421	1,370
Total of characters	1,165	12,640	13,457	17,854
Time Period	B.C. 5–B.C. 3	A.D. 4–A.D. 6	A.D. 12–A.D. 14	A.D. 21

<https://doi.org/10.1371/journal.pone.0192545.t001>

due to the change of word length over time. However, words are the actual basic units of language usage, not characters. Thus, it does not matter whether they are mono syllable/character words in ancient Chinese (e.g., 道, *dao*, means road, morality, doctrine, or method) or two syllable/character words in modern Chinese (e.g., 电话, *dianhua*, means telephone); both of them are the basic units of language used in their times.

Word co-occurrence networks

We construct the networks based on the linear relation of words in the authentic texts. It is worth noting that before building the networks, we must deal with the punctuation problem. Punctuation such as “,” “!” “.” “...” “:” and “;” are treated as signals to stop, while other punctuation such as “《》” ““”” “.” “『』” and “【】”, are dealt with according on the context. We refer to the words sequences that are separated by punctuation as “linguistic blocks” in this study. The loops, the arcs between two tokens that share the same word type, are allowed in the networks.

Why do we choose the “word co-occurrence” networks? Miyagawa, Berwick and Okanoya argued that the language system is composed of two modules—namely, L(lexical) and E(xpression)—and expressions are the combinations of vocabulary/words [9]. Meara defined the co-occurrence relationship of words as the “co-citation links of words in a text” [50]. Furthermore, she claimed that word clustering, which is caused by word co-occurrence relationships, is perhaps the embodiment of this emerging feature. In other words, the co-occurrence relationships of words reflect the nature of a text, and they can generate some very interesting and rather unusual mappings of text structures. Furthermore, word co-occurrence models have also been applied to authorship attribution [51], semantic analyses [52] and word clustering [53]. It is a widely used text feature from the linguistic point.

We run java codes, 2-gramWordCount, for obtaining the adjacent word pairs or 2-grams. For illustration, an English network sample based on a simple English paragraph (S1 File) is given in Fig 1. The nodes in Fig 1 are word forms. However, since there is no distinction between word forms and word types in Chinese, there is only one word type for a Chinese text.

We used Cytoscape and Pajek to obtain the basic parameters. There are 34 nodes and 44 arcs in the network. The more in-depth parameters are displayed below.

Network parameters

The network parameters that we are most interested in are those closely related to language hierarchical features, such as clustering coefficient ($\langle C \rangle$), degree-dependent clustering coefficient $\bar{C}(k)$, average path length ($\langle l \rangle$), degree (Q), and modularity. Some common network parameters, such as, the power rate index of the degree distribution, average degree of nearest neighbors $\bar{K}_{nn}(k)$, diameter (D), density, centrality, etc. are also investigated for comparison. For better understanding of the results, the non-network parameters, mean utterance length (MUL) and average word length are also applied in current study. All the network parameters are calculated by network analysis software Cytoscape and Pajek.

Degree distribution and average degree of nearest neighbors. The degree k of a node is the number of its neighbors. For example, the degree of the node “she” in Fig 1 is 6, and the degree of the node “woman” is 2. If we distinguish the directions of the arcs between nodes, we can then obtain in-degrees and out-degrees. In-degree is the number of arcs that a node receives, and out-degree is the number of arcs that a node sends out. For example, the in-degree of the node “she” is 3, and its out-degree is 3 also; meanwhile, the node “woman” has

Small-world networks. Another crucial aspect of complex networks is the small-world feature. What is the small-world feature? According to Watts & Strogatz, it is defined by two complex network parameters, the clustering coefficient ($\langle C \rangle$) and the average path length ($\langle l \rangle$) [54]. The clustering coefficient is an index used to measure the tendency of aggregation or the tendency of forming clusters. It measures the possibility that a node's neighbors are also connected to each other. Specifically, if k_i is the degree of node i (i.e., node i has k_i neighbors), and E_i is regarded as the edges between these k_i nodes, then these k_i nodes can maximally have $k_i(k_i - 1) / 2$ edges and the clustering coefficient C_i of the node i is $2E_i / k_i(k_i - 1)$. The clustering coefficient $\langle C \rangle$ of the whole network is therefore the mathematical mean of the clustering coefficients of all nodes. For example, as mentioned, the degree of the node “she” in Fig 1 is 6. Its six neighbors are *was*, *because*, *could*, *but*, *jumped*, and *and*. There is an edge between these six neighbors, *jumped-and*. Therefore, E_i is 1, and the maximum number of edges that can exist between these six neighbors is 15, so the clustering coefficient C_i of node “she” is approximately equal to 0.0667. It can be seen that the smaller the clustering coefficient, the sparser the network.

The average path length $\langle l \rangle$ measures the mean of all paths that connect any two nodes in the network. For a given pair of nodes i, j in the network and the distance L_{ij} between them is the number of edges of the shortest path between them. For example, the shortest path between nodes “cat” and “she” is 2 in Fig 1. The longest path among all minimum paths represents the diameter of the network. For example, the diameter of the network in Fig 1 is 13—that is, the path between “an” and “jumped.”

If a network has a small Average Path Length $\langle l \rangle$ ($\langle l \rangle \sim \langle l_{rnd} \rangle$), and a large Clustering Coefficient (C), then we can call it a Small World network. However, this definition is rather qualitative, because it is difficult to decide whether the observed value is small or large. In general, it requires comparing the observed $\langle l \rangle$ and (C) to that of a stochastic network. The stochastic network model used in this paper is the ER model proposed by Erdős and Renyi [55]. Humphries & Gurney defined a quantitative parameter, the S-Small-worldness, for measuring the Small World feature [56]. The formula is as follows:

$$S = \frac{C/C_{rnd}}{l/l_{rnd}}$$

In this formula, C denotes the network Clustering Coefficient, C_{rnd} represents the Clustering Coefficient of the ER random network with the same number of nodes and the same Average Degree; l denotes the Average Minimum Path, and l_{rnd} represents the Average Path Length of the ER random network. When the S is greater than 1, the network can be considered as a Small World network.

Degree-dependent clustering coefficient and modularity. The degree-dependent clustering coefficient $\overline{C}(k)$ actually reflects the influence of the node to its neighboring nodes' clustering coefficients. $\overline{C}(k)$ is the mean of the clustering coefficients of all nodes with degree k . If the degree k is represented by the X-axis, $\overline{C}(k)$ denotes the Y-axis, and the data distribution conforms to the power law, then the network is a hierarchical network. Again, as with many networks in the real world, language complex networks' $\overline{C}(k)$ distributions usually meet this condition, i.e., $\overline{C}(k) \sim k^{-\gamma}$. This means that language complex networks tend to present as hierarchical structures. In short, the so-called hierarchical structure refers to such a network model: nodes with low degrees tend to form relatively dense sub-networks, the connections between these sub-networks are rather sparse, and then the nodes with high degrees will connect these sub-networks into a large connected network [16].

Modularity is a parameter that is closely related to the hierarchy feature of networks. It is generally believed that hierarchical networks will show modular characteristics and that these two are interrelated [57]. Modularity, in fact, is a parameter that concerns the clusters or communities in networks. However, a widely recognized, unique definition of the “community structure” within a network is still absent. The commonly used definition is based on the relative frequency of links: the nodes in a network can be divided into groups, and the links within a group are thought to be dense while the links between groups are sparser. In this definition, the criteria for “dense” and “sparse” are unclear, thus limiting the usefulness of this definition. Therefore, researchers have tried to offer a more quantitative definition of this concept. Newman proposed the Q-modularity index for measuring the modularity feature of networks, and this is the most commonly accepted method nowadays [58]. It is also the method we use for analyzing modularity of language networks in this study.

Other indicators. The Mean Length of Utterance (MLU), which is often treated as one of network parameters, is also observed. The definition of utterance is in Section “Word co-occurrence networks.”

The Average Degree k is the average of all nodes’ degrees.

Density refers to the ratio of the number of existing edges to theoretical maximum number of edges that one network, with the same number of nodes, can have. It reflects how dense a network is. The four observed language networks in this study are rather sparse networks [59], which indicates the low probability of connections between two random words.

The Centrality of a node can reveal its importance in the network. We use the most common centrality indicators in this work: Degree Centrality and Betweenness Centrality. Degree Centrality is the degree of a node. Obviously, this indicator only considers the direct neighbors of a node, so it reflects only the local importance of the node, while the Betweenness Centrality of a node is more focusing on the macro importance. If a node were on many shortest paths of other node-pairs, it would occupy a relatively central position in the network. Since the Betweenness Centrality considers not only the direct neighbors of a node, but also the other nodes in the network, it can reflect the overall importance of a node in the network better.

The Centrality of the whole network is generally called Centralization. Degree Centralization reflects how strong the local hubs (central nodes) are in a network. In language networks, the Degree Centralization reveals the strength of the local hubs (high-degree nodes). The greater the Degree Centralization of a network, the stronger the combination abilities of high-degree nodes/words in the network.

Results

Hierarchy is also recognized to be a result of other features. It is generally believed that hierarchical networks must first share some characteristics with complex networks (such as being small-world and scale-free); furthermore, the hierarchicality is likely to be closely related to the origin of these complex network features. For investigating the evolution process of Chinese networks, we will observe and describe the topological characteristics (including small-world and scale-free), the modularity, and the hierarchicality of four diachronic Chinese networks. The relationships between these networks characteristics and some common linguistic features, such as the frequency and length distributions of linguistic units, are also discussed. The observed network parameters, n = number of nodes; $\langle k \rangle$ = average degree; P = network density; $\langle C_{rnd} \rangle$ = clustering coefficient of the random network counterpart; $\langle l \rangle$ = average path length; $\langle l_{rnd} \rangle$ = average path length of the random network counterpart; D = diameter; $\langle C \rangle$ = clustering coefficient; γ = exponent of the power law that best fits to the degree distribution; R_l^2 = determination coefficient of the power law with exponent γ ; β = exponent of the

power law that best fits to $\overline{K}_{nn}(k)$; R_2^2 = determination coefficient of the powerlaw with exponent β ; α = exponent of the power law that best fits to degree dependent clustering coefficient $\langle \overline{C}(k) \rangle$; δ = coefficient of the the power law that best fits to $\langle \overline{C}(k) \rangle$; R_3^2 = determination coefficient of the power law with exponent α ; NC = network centralization, are shown in Table 2. Corresponding random network parameters as a baseline are also presented in Table 2 with a tag '*rnd*'. To be specific, the random networks share the same numbers of nodes and average degrees with the observed networks. The γ , β , and R^2 parameters are absent for the random networks since their degree distributions and average degree of nearest neighbors distributions do not fit to power law functions.

A MANOVA test (Model as independent variable, Time as concomitant variable, and P , $\langle l \rangle$, D , $\langle C \rangle$, NC , Q -modularity measure, δ , α , R_3^2 as dependent variables) shows that there is a significant difference between two models with $P \approx 0.033 < 0.05$. We discuss the details on change of co-occurrence networks from several different aspects below.

Evolution of degree distribution

Overall, in a co-occurrence network, the degree of a node/word reflects how many other words it can be adjacent to in sentences. Meara has referred to this co-occurrence relationship as co-citation [50]. Linguistically speaking, the degree of a word can be explained, in a way, as a word's "valence," which usually means the syntactic ability of a word—usually a verb though sometimes an adjective or a noun—to combine with complements or adjuncts [60].

In this paper, the interpretation of the valence is based on PVP theory (Probabilistic Valency Pattern Theory) [61, 62]. Unlike the common definition of the term, the distinction between complements and adjuncts [60] is ignored in PVP; therefore, valence is regarded as a general syntactic capacity of a linguistic unit (word) combining with other linguistic units (words). In language networks, the valence of a word is equal to the sum of its centripetal (input) and centrifugal (output) degrees. The centripetal degree refers to the ability of a word to be dominated by other words, while the centrifugal degree refers to the ability of a word to dominate other words. Since the study is based on word co-occurrence networks, the parameter "degree" refers to the ability of words to be adjacent to other words in sentences. Since about 50% of the syntactic dependencies are found between linearly adjacent words in Chinese sentences [63], word co-occurrence networks in this study could reflect, to some extent, the syntactic patterns of Chinese.

Fig 2 shows the results of fitting the degree distributions to the power law.

According to Table 2, the absolute value of the exponential γ of the four power law distributions is increasing, which indicates that the enhancement of the power law of the degree distribution. In other words, few high-degree nodes are seeing increases, or the degrees of most of the nodes are experiencing reductions.

Since the degree distributions fit to the power law, observed networks are all scale-free networks. The scale-free feature of networks reveals that only a very small number of nodes/words have a strong ability to combine with other nodes/words, and most of the nodes/words have a weak capacity. This phenomenon can be explained by the "principle of least effort" [11]. The ability of a word to combine with other words is in proportion to its flexibility of use. Only a limited number of words have high flexibility, a quality which allows them to be used in a rich variety of contexts; their high frequency of use facilitates communication for both speakers/writers and listeners/readers. To be more specific, a small number of contextually flexible words can reduce the encoding effort for speakers/writers, and the remaining words that are usually used in a non-ambiguous context can reduce the decoding effort for listeners/readers. Therefore, the two parts work together and achieve a dynamic balance.

Table 2. Diachronic change of network parameters.

Parameters	Network 1	Network 2	Network 3	Network 4
N	1,527	3,034	2,227	2,526
$\langle k \rangle$	6.6051	4.2149	5.1612	4.7403
$\langle k_{rnd} \rangle$	6.6064	4.2373	5.1863	4.7593
P	0.0022	0.0007	0.0012	0.0020
P_{rnd}	0.0043	0.0014	0.0023	0.0019
$\langle l \rangle$	3.6553	4.2421	4.3153	4.7460
$\langle l_{rnd} \rangle$	4.0943	5.6825	4.8679	5.1763
D	9	13	14	16
D_{rnd}	7	13	9	10
$\langle C \rangle$	0.1542	0.0716	0.0681	0.0563
$\langle C_{rnd} \rangle$	0.0061	0.0016	0.0028	0.0013
NC	0.246	0.163	0.107	0.199
NC_{rnd}	0.0093	0.0109	0.0131	0.0088
Q-modularity measure	0.3481	0.4846	0.4345	0.4832
Q_{rnd}	0.3935	0.5152	0.4522	0.4768
S (Small-worldness measure)	35.9567	57.6239	27.5611	28.2322
γ	-1.173	-1.270	-1.353	-1.449
R_1^2	0.775	0.811	0.851	0.824
β	-0.418	-0.396	-0.185	-0.355
R_2^2	0.835	0.729	0.566	0.661
δ	0.956	0.454	0.304	0.292
δ_{rnd}	5.797	3.966	5.652	4.229
α	-0.609	-0.576	-0.375	-0.543
α_{rnd}	-2.543	-2.262	-2.5	-2.317
R_3^2	0.803	0.688	0.564	0.571
$R_3^2_{rnd}$	0.9986	0.9951	0.9996	0.9996

<https://doi.org/10.1371/journal.pone.0192545.t002>

By observing the change of the parameter γ , we can see that the power law distribution of the Chinese word co-occurrence networks is becoming stronger over time. It means that the binding ability of a small number of words is increasing, and more and more words have a relatively weak capacity. The scale-free network is also called an authoritarian network, and from this point of view, the Chinese network has become more authoritarian over time.

Evolution of the average degree of nearest neighbors

The average degree of nearest neighbors, $\bar{K}_{nn}(k)$, reflects the influence of one node's degree on its neighbors' degrees. It represents the correlation between the degree of a node and the degrees of its neighbors. Complex language networks tend to show a clear degree of correlation. Dynamic language networks often exhibit negative correlations, which means that language units with strong (weak) binding abilities tend to be combined with nodes with weak (strong) binding capabilities. Fig 3 shows the linear fitting of the $\bar{K}_{nn}(k)$ of the four Chinese networks.

In Fig 3, the X-axis represents the degrees and the Y-axis represents the average degrees of nodes which are neighbors of nodes with certain degree. It can be seen from Fig 3 that there has been a negative correlation between these two values for all four diachronic Chinese networks. Therefore, a high-degree node (a word with a strong combination capacity) tends to link to low-degree nodes (a word with a weak combination ability). However, the change is

Fig 2. Power-law fitting (dual-log) of degree distributions of four diachronic Chinese networks.

<https://doi.org/10.1371/journal.pone.0192545.g002>

not one-directional. The absolute value of the slope decreased from periods 1 to 3, while it increased slightly in period 4.

Evolution of the small-world feature

According to the model of Watts and Strogatz [54], small-world networks can be defined by two complex network parameters—namely, the clustering coefficient $\langle C \rangle$ and the average path length ($\langle l \rangle$). The evolution of the average path length (or characteristic path length) is shown in Fig 4. According to the S value in Table 2, these four networks are all small-world networks. However, from Fig 4 we can see that although $\langle C \rangle$ is continuously getting smaller, $\langle l \rangle$ is increasing over time. This means that the Chinese word co-occurrence network is expanding although it is still a small-world network.

Meara has suggested that clustering seems to be a fundamental attribute of words in a text and may be related to the repetition of some words in the text [50]. It is also believed that the structure of a complex network affects the functionality of a system [64]. Studies have shown that small-world networks can provide efficient communication between nodes [65, 66].

According to Fig 4, the average path length showed an increasing trend, while the clustering coefficient showed a decreasing tendency. A MANOVA test (Model as independent variable,

Fig 3. Power-law fitting (dual-log) of average degree of nearest neighbors of four diachronic Chinese networks.

<https://doi.org/10.1371/journal.pone.0192545.g003>

Time as concomitant variable, and $\langle C \rangle$, $\langle l \rangle$ as dependent variable) shows that there is a significant difference between two models with $P \approx 0.035 < 0.05$. This means that although the Chinese word co-occurrence network maintains as a small-world network, it has become increasingly larger globally, but the local connection has become increasing weaker.

Evolution of the degree-dependent clustering coefficients

Hierarchy is the core organizational principle of complex networks and it is also the best observation spot for exploring the phenomenon of networks [15]. Ravasz, Somera, Mongru, Oltvai and Barabási proposed using the degree-dependent clustering coefficient as a verification index for the network's hierarchy feature, which is also a commonly accepted method in most language network studies [16]. Hierarchical networks are characterized by the tendency of low-degree nodes to form tightly connected subnets, with the high-degree nodes then linking these subnets into a connected system.

Fig 5 shows the power-law fitting results for the degree-dependent clustering coefficients of the four Chinese networks. The results show that all four networks exhibit hierarchical structures, which means that the words with weak combination capacity tend to form small groups first, and the words with strong combination capacity connect them into a whole.

These four distributions of the degree-dependent clustering coefficients are very similar to those of the Fig 3 in the [57], all of which belong to the hierarchical network. In these four

Fig 4. Evolution of the clustering coefficient ($\langle C \rangle$) and the average path length ($\langle L \rangle$).

<https://doi.org/10.1371/journal.pone.0192545.g004>

diachronic language networks, $\overline{C}(k)$ decreases as parameter k increases. This illustrates the existence of hierarchical structures, which means that low-degree nodes are more likely to form small groups with a high density, and these small groups are weakly connected. Meanwhile, the high-degree nodes (usually representing functional words) connect these small groups into a whole. From the diachronic point of view, the degree of fit shows a decreasing trend.

In order to further investigate the difference between the degree-dependent clustering coefficients of nodes with different degrees, we calculate the means of the top 10, the middle 10 and the lowest 10 degree-dependent clustering coefficients for the four networks. The results are shown in Fig 6. It can be seen that there is no significant change for the average degree-dependent clustering coefficient of the top 10 high-degree nodes; it maintains the same level. There is an obvious decrease from period 1 to period 2 for the average degree-dependent clustering coefficient of the middle 10 high-degree nodes; however, it becomes more stable for the remaining periods. There is an observed decreasing tendency for the average degree-dependent clustering coefficient of the last 10 high-degree nodes. The evolution of the hierarchy is closely related to the distribution of node degrees and degree dependent parameters, which we will discuss more in detail later.

Fig 5. Power-law fitting of the degree-dependent clustering coefficients of four diachronic Chinese networks.

<https://doi.org/10.1371/journal.pone.0192545.g005>

Modularity is a concept that is closely related to hierarchicality. Barabási, Ravasz and Oltvai argued that hierarchical organization and modularity are the most important organizational principles for many complex networks, and it is often assumed that the module structure and the hierarchical structure are usually found together in a complex network [57]. Many real world networks are hierarchical, such as www networks and biological networks, which generally have both scale-free and community structures [16]. Such communities are also called modules. Sporns, Chialvo, Kaiser and Hilgetag argued that modular networks are small-world networks [67]. On the one hand, the high clustering of nodes in the same module can ensure efficiency in completing specific functions, such as visual dynamic tracking in biological networks; on the other hand, the small average path length in a macro sense can ensure that the system can complete some more general functions, such as working memory.

In fact, the degree of modularity refers to the difference between a network with a community structure and a random network, since a random network lacks a community structure. The greater the difference, the more significant the community is. Newman proposed the following modularity calculation method: because the modularity of a random network is less than 0.3, a network with a modularity greater than 0.3 has a clear community structure [58]. As can be seen from the Table 2, Q-modularity values of four networks are all greater than 0.3, so there is a clear module structure in all of them. Furthermore, the value of Q has also evolved. Overall, the Q is increasing over time.

Fig 6. The means of the top 10, the middle 10, and the final 10 degree-dependent clustering coefficients.

<https://doi.org/10.1371/journal.pone.0192545.g006>

Evolution of mean utterance length and mean word length

The change of the Chinese word co-occurrence network is accompanied by the change of the mean utterance length (MUL) and mean word length (MWL).

Although MUL is not a network parameter, it is generally regarded as one in language network research. The reason for this is mainly that the construction of the network is closely related with the UL, no matter whether the network is based on dependency syntax or word co-occurrence relationships. Cross-sentence relationships are not allowed. In this paper, an utterance is defined as a “chunk,”—that is, a continuous word string. It is similar to a clause in grammar.

Fig 7 illustrates the four MUL distributions. The X-axis represents the length of the utterances, and the Y-axis represents the number of utterances sharing the same length.

It can be seen from Fig 7 that the distribution does not change significantly from period 1 to period 2; however, it becomes broader in period 3, and it continues the same broadening

Fig 7. MUL distributions (TP1 corresponds to Network 1 text).

<https://doi.org/10.1371/journal.pone.0192545.g007>

tendency in period 4. In addition, the distribution is obviously right-skewed, and the maximum number of utterances lessens.

In addition to the increasing of the MUL, the MWL of texts also increases. This means that the MUL of Chinese shows an increasing trend no matter whether it is measured based on words or Chinese characters. In Table 3, the WL (static), measured based on word types, and the WL (dynamic), measured based on tokens, are both increasing over time.

Discussion

Language is a complex adaptive system, and the more ancient a language is, the more complex it may be [14], but how does a language become complex? In this paper, Chinese is used as a means for investigating this question from the perspective of hierarchical network systems. We built four diachronic word co-occurrence networks based on texts of ancient Chinese, middle ancient times Chinese, modern times Chinese, and modern Chinese, and then analyzed the topological changes of these networks over time.

The questions that we address first are whether Chinese word co-occurrence networks show any complex characteristics and whether network parameters have changed over time. According to our observations, all four diachronic Chinese word co-occurrence networks show similar complex network characteristics, such as scale-free and small-world features. However, they are different in terms of fine details. More specifically, the power law fitted

Table 3. The diachronic change of the MUL of Chinese (measured based on words) and the MWL of Chinese (measured based on characters).

	TP 1	TP 2	TP 3	TP 4
MUL	4.07	3.82	4.74	5.55
MWL(static)	1.17	1.26	1.35	1.79
MWL(dynamic)	1.48	1.66	1.71	2.01

<https://doi.org/10.1371/journal.pone.0192545.t003>

distributions of words' degrees have become extreme. This means that the high-degree words (they are usually also the central words, such as functional words) are developing stronger combination capacity compared to other words, while most other words have a weak ability to combine with other words. Since the average clustering coefficient $\langle C \rangle$ of networks has declined, it can be inferred that the connections between low-degree words are also diminishing. The small-world feature of networks ensures the efficiency of communication. Judging from observed data, although the four diachronic networks all show the small-world feature, the evolution of the average clustering coefficients and the average path lengths, two parameters related to the small-world feature, shows that the language network is actually getting larger. This means that the language system is becoming more complex in the lexical network sense. Moreover, the enhancement of the scale-free feature means that people are, consciously or unconsciously, adhering to the "Principal of Least Effort" more and more, so that they can avoid the increasing difficulty of communicating along with the growing of the language network [68].

The average degree of nearest neighbors is also an important network. Previous studies have confirmed that dynamic language networks show significant negative correlations with this feature, which means that low-degree words are more likely to be connected with high-degree central words. According to our findings, this negative correlation decreased from periods 1 to 3, meaning that high-degree words were beginning to connect more often with medium-degree words. Low-degree words were doing so as well. However, this trend rebounded from periods 3 to 4, although the negative correlation was still weaker than in periods 1 and 2. We find that the changing tendency of the negative correlation is consistent with the change of the network centrality (see *NC* in Table 2) and the betweenness (see *SI*). In fact, the average degree of nearest neighbors is closely related to the centrality of nodes. For example, Ferrer-i-Cancho and Solé found out that the average degree of nearest neighbors decreases after removing central nodes from the language networks [30].

Regarding the hierarchy features of language networks (i.e., the degree-dependent clustering coefficient), the four word co-occurrence networks all show similar distributions. Clustering coefficients of nodes decreased in average with the increase of the node degree, which means that the Chinese word co-occurrence network has consistently hierarchical characteristics. Low-degree words with weak combination capacities have tended to form tightly connected sub-lexical groups, and then the high-degree words, which are usually functional words, have linked these sub-lexical groups into a whole. However, over time, the degree-dependent clustering coefficient has decreased. The reason for this is that the degree-dependent clustering coefficients of low-degree words are decreasing continuously, while those of high-degree words and medium-degree words do not change significantly. It is generally believed that network hierarchy is accompanied by its modularity. The results of this study show that the modularity of all four diachronic networks is greater than 0.3, indicating that they all have clear modular structures. However, from period 1 to periods 2, 3 and 4, the modularity of the Chinese word co-occurrence network generally increased, which shows the deepening trend of the communitization of Chinese vocabulary networks.

In synthesizing the above analyses, we can generally say that the number of the sub-lexical groups of Chinese word co-occurrence network is increasing over time. The high-degree words in the network remain the central nodes of the entire network. They are crucial and they maintain the small-world and scale-free features of the network, which ensure the efficiency of language communication. Meanwhile, the modularity of low-degree words is increasing, which indicates that the subgroups are becoming specialized and their number is increasing. Because of this specialization, they are more inclined to connect with medium-degree words with which they share some similarities, and then finally connect into a whole

through the high-degree words. Interestingly, Borodkin, Kenett, Faust and Mashal [21] considered the average clustering coefficient as a parameter of local communitization, modularity as the parameter of communitization for mediators, and the average path length as the parameter of macro-scale communitization in networks. He found out lexical networks of the second language displayed greater local connectivity and less modular community structure than the network in the native language. This shows that if we see the language evolution from the point of view of its complexity, the increasing modularity of networks may actually mean higher language complexity. The study we present verifies this point with authentic diachronic data.

During its evolution, the Chinese word co-occurrence network experienced some changes while keeping its basic characteristics, including its small-world and scale-free qualities, the negative correlation of the average degree of nearest neighbors, and most importantly, the hierarchical feature. In general, the evolution of the Chinese word co-occurrence network shows the characteristics of stability and self-adaptation. As a complex adaptive system, the evolution of the Chinese word co-occurrence network, while having its own internal evolution as a dynamic symbol system, is inseparable from external social and cultural factors, which will be discussed in more detail below.

The relationship between the linear and hierarchical structures is also a question examined in this paper. Since the networks are based on word co-occurrence relationships, their sub-structures are closely related to the chunks (defined in section 2) and the distribution of the length of utterances (i.e., MUL). MUL can also be used as a rough measurement of syntactic complexity. The greater the MUL, the more complex the syntax. Over time, MUL increases, and in modern Chinese (network 4), it shows significant growth. In the networks, the parameter average path length is very closely related to the MUL because utterances are actually linear words sequences. Therefore, the length of utterances affects the length of the paths between two words. Jiang and Liu examined the relationship between sentence length and average dependency distance in Chinese and found that there is a positive correlation between these two parameters—that is, the longer the sentence, the greater the average dependency distance [69]. However, grammatical factors and cognitive factors such as memory limit the growth of the average dependent distance, making it as short as possible. This phenomenon is referred to as “minimizing dependency distance.” Although the word co-occurrence relationship we observed in this study is different from dependent syntax, they are indeed very similar. Approximately half of the dependency relationships are found between linearly adjacent words. In some ways, the average dependency distance and the average path length are similar [62]. According to our observations, the average path length also grows with the increase of MUL, but much more slowly. This may due to the influence of various human cognitive factors (e.g., short-term memory) on the evolution of language, especially in terms of grammar. For example, studies have shown that no more than seven blocks of information can be temporarily stored in the brain at any given time [70]. This may also mean that although Chinese has a hierarchical feature, it is developing more linear serialization and the word order of Chinese is also becoming more flexible. In addition, the reason why the average path length has maintained relatively steady growth may also be due to the emergence of some word collocation patterns [71, 72], which are very similar to the motivation for dependency minimization [69].

In fact, the increase in MUL is not only expressed in the number of words, but also by the average length of words, which is also growing (see Table 3). From the perspective of language information load, a higher MLU or word length usually indicates a higher level of language proficiency [73, 74]. Moreover, the growth is closely related to the need for more nuanced language expressions. As society develops, new things and new concepts appear. Language communication and expression become more refined in order to meet the need to communicate

information. However, efficiency is also a fundamental principle of language communication. The longer the utterances and the words, the greater the cognitive complexity. For balancing these two needs, some syntactic changes are inevitable. The emergence of new word collocation patterns is likely to be the result of this adaptation.

By analyzing four diachronic Chinese word co-occurrence networks, we speculate the evolution mechanism was as follows: Initially, low-degree words, which are usually low-frequency words, formed into dense subgroups. Some high-degree words or hubs, usually function words, then linked these subgroups into a whole for ensuring the connection of the entire language network. Such a network is a very efficient organizational system. It ensures both the adequacy of the description and the efficiency of expression. As human society developed, people needed to find new words and syntactic structures to fit their changing world. Word and utterance length began to grow; the sub-lexical groups also became more specialized. This led to the increase of the width and length of the average path length. The small world became bigger. At the same time, the hierarchy feature of the network underwent a corresponding change. Due to the specialization of the sub-lexical groups, the connections between these low-degree words became less. Therefore, the clustering coefficient got smaller. The clustering coefficient of high-degree words remained similar, while that of the medium-degree words was reduced from period 1 to 2, and then remained at the same level. All these indicate that the modularity of the language network continues to develop. For dealing with the dilemma of increasing language network, people started to make some syntactic adjustments—that is, new word collocation patterns and more flexible word orders. That may be also the main reason for the observed phenomenon that the language network persisted as a small-world and scale-free network. It was to ensure the efficiency and to reduce the efforts of language communication. However, speculation regarding new word collocation patterns and more flexible word order requires further empirical investigation.

Conclusion

This study constructed four diachronic word co-occurrence networks based on texts of ancient Chinese, middle ancient times Chinese, modern times Chinese, and modern Chinese, and investigated the evolution of Chinese as a hierarchical system. We find the following:

1. Hierarchy (based on the definition of network topology) exists in four diachronic Chinese networks, indicating that Chinese has always been a hierarchical complex adaptive system.
2. If the Chinese hierarchical system is divided into low (local groups), middle (middle layer clusters or small communities) and high (global structure) levels, then the evolution of Chinese (syntactic) network can be described as a) the connections of words at the micro-language level are continually weakening; b) the number of meso-level word communities had a significant growth in middle ancient times (network 2), and then maintained at a relatively high level; and c) the global network in the macro level is expanding. This means that more and more low-degree words tend to be connected to the medium-connection words and then the high-connection words will link these groups into a highly efficient small-world language network.
3. As the number of sub-words-communities as well as modularity increases, the written Chinese word co-occurrence network becomes larger, which is consistent with the increase of the length of linear sequences, as measured by mean word length and mean utterance length. The emergence of new word collocation patterns and more flexible word orders may be important reasons for the language system to maintain its hierarchical structure.

4. We summarize the diachronic language developing mechanism of written Chinese word co-occurrence network based on observed changes. With the development of human society, new things and new concepts emerged. There was a growing need for refined language expressions. The words of specific fields began to form either large or small sub-lexical groups. Words, clauses, and sentences also began to lengthen. For balancing the increase of language units and the efficiency of language communications, new word collocation patterns started to appear and word orders became more flexible. The language system could then offer more refined language expressions while at the same time achieving efficient language communication.

This paper investigates the evolution of written Chinese lexical networks spanning more than 2,000 years. Based on the co-occurrence of words in sentences, we have generated maps of the lexical structure of the texts. As a result, after probing into the changes of indicator values of the networks that are closely related to the hierarchy feature, we have found some diachronic patterns in the language network. This research brings new insights to understanding the emergence, adaptability, and stability of human language system.

Supporting information

S1 File. S1 Supporting information.
(DOCX)

Author Contributions

Conceptualization: Heng Chen, Haitao Liu.

Data curation: Heng Chen.

Formal analysis: Heng Chen, Haitao Liu.

Funding acquisition: Haitao Liu.

Investigation: Heng Chen.

Methodology: Heng Chen, Haitao Liu.

Project administration: Haitao Liu.

Resources: Heng Chen.

Software: Heng Chen.

Supervision: Haitao Liu.

Visualization: Heng Chen.

Writing – original draft: Heng Chen, Haitao Liu.

Writing – review & editing: Heng Chen, Xinying Chen, Haitao Liu.

References

1. Nowak M. A., Plotkin J. B., & Jansen V. A. (2000). The evolution of syntactic communication. *Nature*, 404(6777), 495–498.
2. Solé R. V., Corominas-Murtra B., Valverde S., & Steels L. (2010). Language networks: Their structure, function, and evolution. *Complexity*, 15(6), 20–26. <https://doi.org/10.1002/cplx.20305>
3. Pinker S. (2000). *The Language Instinct* (Rev. Ed.). New York: Harper Collins.
4. Kretzschmar W. A. (2009). *The Linguistics of Speech*. Cambridge University Press, Cambridge.

5. Larsen-Freeman D., & Cameron L. (2008). Research methodology on language development from a complex systems perspective. *The Modern Language Journal*, 92(2), 200–213. <https://doi.org/10.1111/j.1540-4781.2008.00714.x>
6. Massip-Bonet À, & Bastardas-Boada A. (Eds.). (2013). *Complexity Perspectives on Language, Communication and Society*. Springer Berlin Heidelberg, Berlin, Heidelberg.
7. Ellis N. C., Römer U., & O'Donnell M. B. (2016). Usage-based Approaches to Language Acquisition and Processing: Cognitive and Corpus Investigations of Construction Grammar. *Language Learning*, Vol. 66(S1).
8. Chomsky N. (1957). *Syntactic Structures*, The Hague/Paris: Mouton.
9. Miyagawa S., Berwick R., & Okanoya K. (2013). The emergence of hierarchical structure in human language. *Frontiers in Psychology*, 4, 71. <https://doi.org/10.3389/fpsyg.2013.00071> PMID: 23431042
10. Cong J., & Liu H. (2014). Approaching human language with complex networks. *Physics of Life Reviews*, 11(4), 598–618. <https://doi.org/10.1016/j.plrev.2014.04.004> PMID: 24794524
11. Zipf G. K. (1949). Human behavior and the principle of least effort. *American Journal of Sociology*, 110(110), 306–306.
12. McWhorter J. (2003). *The power of Babel: A natural history of language*. London: Arrow.
13. McWhorter J. (2011). *Linguistic simplicity and complexity: Why do languages undress?* (Vol. 1). Boston: De Gruyter Mouton.
14. Ellis N. C., & Larsen-Freeman D. (2006). Language emergence: implications for applied linguistics—introduction to the special issue. *Applied Linguistics*, 27(4), 558–589.
15. Clauset A., M, C., & Newman M. E. (2008). Hierarchical structure and the prediction of missing links in networks. *Nature*, 453(7191), 98–101. <https://doi.org/10.1038/nature06830> PMID: 18451861
16. Ravasz E., Somera A. L., Mongru D. A., Oltvai Z. N., & Barabási A. L. (2002). Hierarchical organization of modularity in metabolic networks. *science*, 297(5586), 1551–1555.
17. Simon H. A. (1996). *The Science of the Artificial* (Third edition). MIT Press, London.
18. Halliday M. (1985). *An Introduction to Functional Grammar*. London: Edward Arnold.
19. Niazi M. A., & Hussain A. (2013). *Cognitive Agent-based Computing-I: A Unified Framework for Modeling Complex Adaptive Systems using Agent-based & Complex Network-based Methods*. Springer Netherlands, Dordrecht.
20. Beckner C., Blythe R., Bybee J., Christiansen M. H., Croft W., Ellis N. C. et al. (2009). Language is a complex adaptive system: Position paper. *Language learning*, 59(s1), 1–26.
21. Borodkin K., Kenett Y., Faust M., & Mashal N. (2016). When pumpkin is closer to onion than to squash: The structure of the second language lexicon. *Cognition*, 156, 60–70. <https://doi.org/10.1016/j.cognition.2016.07.014> PMID: 27513870
22. Karuza E. A., Thompsonschi S. L., & Bassett D. S. (2016). Local patterns to global architectures: influences of network topology on human learning. *Trends in Cognitive Sciences*, 20(8), 629–640.
23. Amancio D. R., Oliveira O. N. Jr., & Costa L. D. F. (2013). Unveiling the relationship between complex networks metrics and word senses. *Epl*, 98(1), 18002.
24. Ferrer-i-Cancho R., Solé R. V., & Köhler R. (2004). Patterns in syntactic dependency networks. *Physical Review E*, 69(5), 051915. <https://doi.org/10.1103/PhysRevE.69.051915> PMID: 15244855
25. Mehler A., Lücking A., Banisch S., Blanchard P., & Job B. (Eds.). (2016). *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks*. Springer Berlin Heidelberg, Berlin Heidelberg.
26. Newman, M. (2010). *Networks: An Introduction*. Oxford Scholarship Online.
27. Gros C. (2008). *Complex and Adaptive Dynamical Systems: A Primer*. Springer, Berlin Heidelberg.
28. Li P., Farkas I., & MacWhinney B. (2004). Early lexical development in a self-organizing neural network. *Neural Networks*, 17(8), 1345–1362. <https://doi.org/10.1016/j.neunet.2004.07.004> PMID: 15555870
29. Ke J., & Yao Y. (2008). Analysing language development from a network approach. *Journal of Quantitative Linguistics*, 15(1), 70–99.
30. Ferrer-i-Cancho R. & Solé R. V. (2001). The small world of human language. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1482), 2261–2265. <https://doi.org/10.1098/rspb.2001.1800> PMID: 11674874
31. Borge-Holthoefer J., & Arenas A. (2010). Semantic networks: Structure and dynamics. *Entropy*, 12(5), 1264–1302. <https://doi.org/10.3390/e12051264>
32. Abramov O., & Mehler A. (2011). Automatic language classification by means of syntactic dependency networks. *Journal of Quantitative Linguistics*, 18(4), 291–336.

33. Amancio D. R. (2015a). A complex network approach to stylometry. *PloS One*, 10(8), e0136076. <https://doi.org/10.1371/journal.pone.0136076> PMID: 26313921
34. Amancio D. R. (2015b). Authorship recognition via fluctuation analysis of network topology and word intermittency. *Journal of Statistical Mechanics: Theory and Experiment*, 2015(3), P03005.
35. Barceló-Coblijn L., Corominas-Murtra B., & Gomila A. (2012). Syntactic trees and small-world networks: Syntactic development as a dynamical process. *Adaptive Behavior*, 20(6), 427–442. <https://doi.org/10.1177/1059712312455439>
36. Gegov E., Gegov A., Gobet F., Atherton M., Freudenthal D., & Pine J. (2012). *Cognitive modelling of language acquisition with complex networks*. In Floares A. (Ed.). *Computational Intelligence*. Nova Science, New York.
37. Lahiri, S. (2013). Complexity of word collocation networks: a preliminary structural analysis. *arXiv pre-print arXiv:1310.5111*.
38. Liu H. & Cong J. (2013) Language clustering with word co-occurrence networks based on parallel texts. *Chinese Science Bulletin*, 58(10): 1139–1144.
39. Akimushkin C., Amancio D. R., & Oliveira O. N. Jr. (2017). Text authorship identified using the dynamics of word co-occurrence networks. *Plos One*, 12(1), e0170527. <https://doi.org/10.1371/journal.pone.0170527> PMID: 28125703
40. Liu H., & Cong J. (2014). Empirical characterization of modern Chinese as a multi-level system from the complex network approach. *Journal of Chinese Linguistics*, 42(1), 1–38.
41. DeFrancis J. (1984). *The Chinese Language: Fact and Fantasy*. University of Hawaii Press, Honolulu.
42. Ge L. (2001). *Out of the Margins: The Rise of Chinese Vernacular Fiction*. University of Hawaii Press, Honolulu.
43. Boltz W. G. (1994). *The Origin and Early Development of the Chinese Writing System*. (Vol. 78), American Oriental Society, New Haven.
44. Pulleyblank E. G. (1995) Outline of Classical Chinese Grammar. UBC Press, Vancouver.
45. Mair V. H. (Ed.) (2001). *The Columbia History of Chinese Literature*. Columbia University Press, New York.
46. Chen H., Liang J., & Liu H. (2015). How does word length evolve in written Chinese?. *Plos One*, 10(9), e0138567.
47. Xu S (2015). *The History of Chinese Dialect*. Peking university press, Beijing.
48. Pagel M., Atkinson Q. D., & Meade A. (2007). Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature*, 449(7163), 717–720.
49. Chao Y. R. (1946). The logical structure of Chinese words. *Language*, 22(1):4–13.
50. Meara P. (2016). Texts as vocabulary networks. In Chodkiewicz H., Steinbrich P., & Krzemińska-Adamk M. (Eds.). *Working with Text and Around Text in Foreign Language Environments*. pp. 193–204. Cham: Springer International Publishing.
51. Mehri A., Darooneh A. H., & Shariati A. (2012). The complex networks approach for authorship attribution of books. *Physica A Statistical Mechanics & Its Applications*, 391(7), 2429–2437.
52. Amancio D. R., Oliveira O. N. Jr., & Costa L. D. F. (2012). Structure-semantics interplay in complex networks and its effects on the predictability of similarity in texts. *Physica A Statistical Mechanics & Its Applications*, 391(18), 4406–4419.
53. Ferrer-i-Cancho R., Capocci A., & Caldarelli G. (2007). Spectral methods cluster words of the same class in a syntactic dependency network. *International Journal of Bifurcation & Chaos*, 17(07), 2453–2463.
54. Watts D. J., & Strogatz S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684), 440–442. <https://doi.org/10.1038/30918>
55. Erdős P., & Rényi A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5 (1), 17–60.
56. Humphries M. D., & Gurney K. (2008). Network 'small-world-ness': A quantitative method for determining canonical network equivalence. *PloS One*, 3(4), e0002051. <https://doi.org/10.1371/journal.pone.0002051> PMID: 18446219
57. Barabási A. L., Ravasz E., & Oltvai Z. (2003). Hierarchical organization of modularity in complex networks. In *Statistical mechanics of complex networks*. Springer Berlin Heidelberg. pp. 46–65
58. Newman M. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 103(103), 8577–8582.
59. Barrat A., Barthélemy M., & Vespignani A. (2008). *Dynamical Processes on Complex Networks*. Cambridge University Press, New York.

60. Herbst T., & Götz-Votteler K. (Eds.). (2007). *Valency: Theoretical, Descriptive and Cognitive Issues*. Mouton de Gruyter, Berlin · New York.
61. Liu H. & Feng Z. (2007). Ziran yuyan chuli de gailü peijia moshi lilun (Probabilistic valency pattern theory for natural language processing). *Yuyan Kexue (Language Science)*, 6(3):32–41.
62. Liu H. (2008). The complexity of Chinese syntactic dependency networks. *Physica A Statistical Mechanics & Its Applications*, 387(12), 3048–3058.
63. Liu H., Zhao Y., & Li W. (2009). Chinese syntactic and typological properties based on dependency syntactic treebanks. *Poznan Studies in Contemporary Linguistics*, 45(4), 509–523.
64. Boccaletti S., Latora V., Moreno Y., Chavez M., & Hwang D. U. (2006). Complex networks: structure and dynamics. *Physics Reports*, 424(4), 175–308.
65. Latora V., & Marchiori M. (2001). Efficient behavior of small-world networks. *Physical Review Letters*, 87(19), 198701. <https://doi.org/10.1103/PhysRevLett.87.198701> PMID: 11690461
66. Bullmore E., & Sporns O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3), 186–198.
67. Sporns O., Chialvo D. R., Kaiser M., & Hilgetag C. C. (2004). Organization, development and function of complex brain networks. *Trends in cognitive sciences*, 8(9), 418–425.
68. Ferrer-i-Cancho R. & Solé R. V. (2003). Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences of the United States of America*, 100(3), 788–91.
69. Jiang J., & Liu H. (2015). The effects of sentence length on dependency distance, dependency direction and the implications-based on a parallel English—Chinese dependency treebank. *Language Sciences*, 50, 93–104. <https://doi.org/10.1016/j.langsci.2015.04.002>
70. Miller G. A. (1994). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 101(2), 343–352. <https://doi.org/10.1037//0033-295X.101.2.343>
71. Chen H. & Liang J. (2017). Chinese word length motif and its evolution. In: Liu Haitao & Liang Junying (Eds.). *Motifs in Language and Text*. Berlin/Boston: De Gruyter Mouton. pp. 37–64.
72. Chen H., & Liu H. (2018). Quantifying evolution of short and long-range correlations in Chinese narrative texts across 2000 years. *Complexity*, vol. 2018, Article ID 9362468. Forthcoming.
73. Rice M. L., Smolik F., Perpich D., Thompson T., Rytting N., & Blossom M. (2010). Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. *Journal of Speech, Language, and Hearing Research: JSLHR*, 53(2), 333–349.
74. Piantadosi S. T., Tily H., & Gibson E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, 108(9), 3526.