2025; 119(2): 35–55
https://doi.org/10.58221/mosp.v119i2.31045
Moderna Språk

Research Article

Syntactic network characteristics of English e-commerce live-streaming discourse

Lin Wang*
School of International Culture and Communication, Communication University of Zhejiang

Hongyan Shi
School of Foreign Languages, Zhejiang Normal University

Received September, 2024; accepted May, 2025;
published online October, 2025

Abstract: This study explores the syntactic network characteristics of English e-commerce live-streaming discourse by employing a syntactic dependency treebank and complex network analysis. The main findings are: (1) The syntactic network of English e-commerce live-streaming discourse exhibits small-world and scale-free properties, which are hallmark traits of complex networks. (2) The central nodes of the network are *be*, *I*, and *the*, with *be* serving as the most central node, while *I* and *the* act as local central nodes. (3) The central node *be* demonstrates both strong centrifugal and centripetal forces. Its centrifugal force is most frequently associated with subject relations and adjective complements, while its centripetal force is characterized by auxiliary and clausal complements. These findings indicate that the syntactic structure of English e-commerce live-streaming discourse is highly robust. This robustness underscores the discourse's functional purpose: to convey information clearly while engaging users through personalization and specificity. Furthermore, the study highlights the critical role of *be* in attributive and descriptive constructions. Overall, this research provides insights into the syntactic organization of e-commerce discourse and demonstrates the effectiveness of complex network analysis in linguistic studies.

Keywords: syntactic complex network, English e-commerce live-streaming discourse, characteristics, central nodes, binding force

1 Introduction

E-commerce live-streaming discourse refers to the discourse acts that occur in the context of live broadcasting on e-commerce platforms, which diverges in its communicative objectives compared to conventional scenarios. Compared with traditional e-commerce, live-streaming e-commerce is a novel business genre (Shi & Dou, 2023) as well as a special type of mediated communication (Liu & Cheng, 2025), which is characterized by authenticity, visibility, real-time interaction, and entertainment (Ma & Mei, 2018). As a key medium for product marketing, live-streaming e-commerce discourse plays a crucial role in influencing consumer behavior. English live-streaming e-commerce has gained significant prominence on international e-commerce platforms, where the quality and characteristics of live-streaming discourse directly impact sales performance.

***Corresponding author:** Lin Wang, E-mail: 512497740@qq.com

A search using the phrase "live streaming" on the Web of Science platform yielded 7,594 items, with the earliest studies dating back to 2008. The majority of studies focus on computer science, accounting for 71.92%. Linguistic research, however, accounted for less than 1%, with only 19 papers. In the linguistic approach, most studies are qualitative in nature and are primarily published in journals such as *Pragmatics* and *Journal of Pragmatics.* These studies predominantly focus on the role of hosts as intermediaries, emphasizing how their discourse significantly influences viewers' behavior. For example, Shi and Dou (2023) conducted an analysis of 100 live-streaming videos from Taobao Live, exploring how successful broadcasters strategically manage relationships with viewers through specific discourse patterns. In contrast, quantitative studies in this area are relatively scarce. Among the few examples, Yang and Wang (2022) investigated the gender effects in Chinese e-commerce live-streaming discourse, focusing on the relationship between buyers and sellers.

While these studies have contributed to the field, several research gaps remain. First, most studies have focused on Chinese live-streaming discourse, with relatively little attention paid to the counterpart English e-commerce live-streaming discourse, which is essential in the context of cross-border e-commerce. Second, understanding the linguistic characteristics of e-commerce live-streaming discourse is a fundamental step to improve its quality and effectiveness. Third, exploring the global and local characteristics of e-commerce live-streaming discourse requires the application of quantitative methods grounded in linguistic theory.

Quantitative linguistics has proven effective in exploring language properties in areas such as general linguistics and second language acquisition (e.g., Liu, 2008; Ouyang & Jiang, 2018; Jiang et al., 2019; Hao et al., 2024). Furthermore, since language can be theoretically modelled as a complex network (Hudson, 2007; Xu, 2009; Liu, 2022), the systematic investigation of its global characteristics necessarily relies on complex network analysis methodologies and computational techniques. Specifically, constructing a syntactic network composed of nodes and edges is an effective approach to identifying linguistic features through the analysis of network parameters. Dependency Grammar, which models language as a network of nodes (words) and edges (syntactic dependencies), is particularly well-suited for the complex network method (Hudson, 1984; Liu, 2017). This approach, combined with network science, provides a robust framework for analysing the structural properties of language, including its global and local characteristics.

Global characteristics refer to the macro-level properties of the complex network, primarily involving two key features: the small-world property and the scale-free property (Liu, 2017). Local characteristics, on the other hand, emphasize the localized patterns of connections and dependencies within the network, rather than its overall structure. These characteristics reveal how individual elements contribute to the syntactic organization of the discourse. Key aspects of local characteristics include node degree and local centrality. The degree of a node reflects its binding forces, which correspond to its syntactic functions. According to the probability valency pattern (PVP) theory, this binding force aligns with the valency of a word, offering a theoretical basis for explaining the syntactic features of central nodes.

In this study, we construct a syntactic complex network for English e-commerce live-streaming discourse based on a dependency treebank to investigate both global and local syntactic characteristics. Specifically, the study addresses the following research questions:

(1) What are the global syntactic characteristics of the English e-commerce live-streaming discourse network?
(2) What are the local syntactic characteristics of the English e-commerce live-streaming

discourse network?
(3) What are the main central nodes?
(4) What are the syntactic functions of the most central node?

2 Research Methods

2.1 The dependency treebank

The data of this study originate from the English live broadcasts on international e-commerce platforms like Amazon Live and YouTube Shopping. The hosts are professional e-commerce broadcasters from English-speaking countries and enjoy a high reputation. The sales domains cover clothing and furniture supplies. A total of more than 200 minutes of live e-commerce videos were recorded for this study. After the recording, the videos were transcribed into text and manually checked and corrected. Each word is lemmatized, for example, *be* is the lemmatized form of *am*, is, *are*, *was*, *were*, *being* and *been*.

This study constructs a dependency treebank of English e-commerce live-streaming discourse (excluding punctuation). A dependency treebank is a syntactic corpus annotated by the Dependency Grammar. According to Word Grammar, a type of dependency grammar, the basic element is the word, and the syntactic structure of a sentence consists of nothing but the typed dependencies among individual words. The dependency relation is the unbalanced syntactic relation between two words, the head and the dependent (Hudson, 2010). In the dependency treebank, the heads, the dependents and their dependency relations are all annotated. The methods of building the dependency treebank can refer to Liu (2022). Figure 1 shows the dependency structure of Example 1.

Example 1 I really hate onions

Figure 1

Dependency structure of Example 1

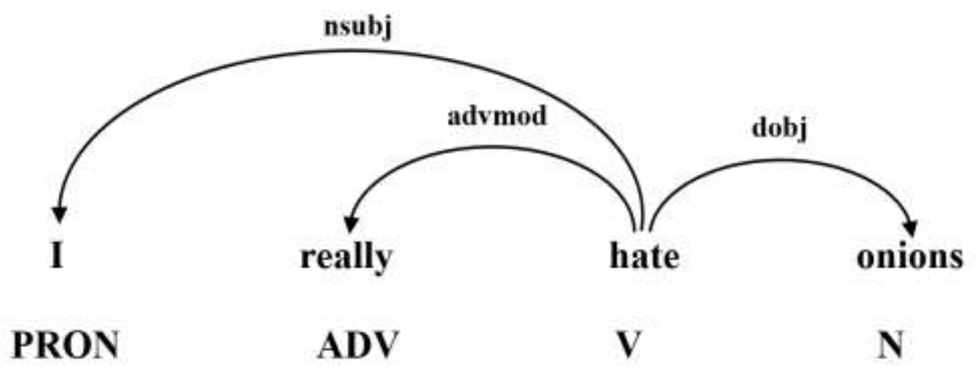


Figure 1 presents the part of speech of each word and the syntactic relations. The dependency type labels the directed arc that links the head and the dependent. For example, the syntactic relation between *hate* and *onions* is the direct object relation and *onions* depends on *hate*.

Stanford Parser[1] (Version 4.2.0) is utilized to perform syntactic dependency analysis on the corpora. Developed by the Natural Language Processing research team at Stanford University in the United States, the Stanford Parser can achieve a relatively high annotation accuracy, surpassing other similar tools, which was validated by Liu and Zhang (2021) who comparatively examined the accuracy of dependency structures annotated by Mate Parser, Stanford Parser, and Malt Parser syntax respectively, and found that Stanford Parser outperformed the other two syntactic analysing tools, both in the English literary and non-literary texts. The greatest average of literary accuracy text was 88.74% and non-literary text accuracy 88.94%. With the aid of Stanford Parser (Version 4.2.0), dependents, heads/governors, the word order, part-of-speech and dependency relations are easily available, and after the automatic processing, the dependency treebank was manually checked and verified.

2.2 The dependency network and analysis indicators

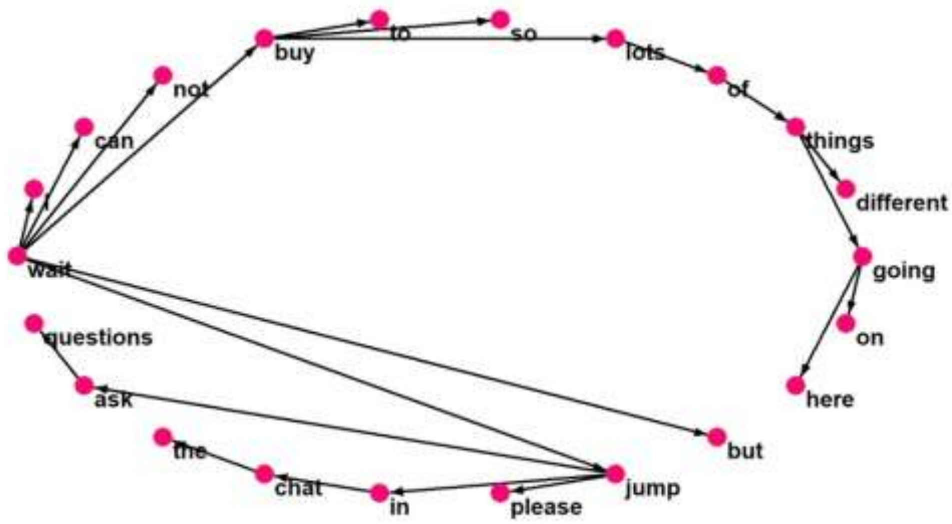
Complex networks are categorized into directed and undirected networks. In directed networks, edges between nodes have a specific direction, while in undirected networks, edges lack direction (Liu, 2022). Language complex networks model language as a complex system (Hudson, 2007; Liu, 2022). These networks can be broadly categorized into static language networks and dynamic language networks. Static language networks are primarily used to analyse the overall structure of a language as a static resource, such as its vocabulary or grammatical rules. In contrast, dynamic language networks focus on various relationships in language usage, capturing the dynamic and evolving nature of linguistic complexity. Syntactic dependency networks and co-occurrence networks have been the focus of extensive study. The dependency networks reveal deeper structural relationships, such as syntactic dependencies between words, while co-occurrence networks capture surface-level relationships between linguistic elements (Liu, 2017). In order to further explore the dependency network characteristics of English e-commerce live-streaming discourse, this study constructs a directed dependency network based on the dependency treebank. Figure 2 shows the directed dependency network of Example 2.

> Example 2
> I cannot wait to buy so lots of different things going on here, but please jump in the chat, ask questions.

1 This can be downloaded at the website https://nlp.stanford.edu/software/lex-parser.shtml#Download

Figure 2

Schematic diagram of the directed dependency network of Example 2



The basic constituent units of a complex network are nodes and edges/arcs In undirected networks, edges are the relationship between the lexical items, while edges convert to arcs in the directed networks. Nodes represent language units, and arcs are the directed relationships between them. In the syntactic complex network, nodes are words and arcs indicate the dependency relationships between words. For example, in Figure 2, *wait* is a node, and there are six arcs between *wait* and other words. Each arc represents one syntactic relation, such as the arc between *I* and *wait* referring to the subject relation. The total number of arcs around a given node is the degree of that node. The arrow indicates the direction of the arc and is related to the combining ability of the word. For example, the degree of the node *wait* is 6, and the degree of the node *in* is 2. The node degree reflects the connectivity of a given node and is the most basic and easily understandable network indicator. Node degree in directed network can be divided into out-degree and in-degree. The total number of arcs from a given node to other nodes is called the out-degree of that node, and the total number of arcs from other nodes to that node is called the in-degree of that node (Liu, 2017). Specifically, in example 2, the in-degree of the node *things* is 2, and the out-degree is 1. According to the Dependency Grammar, the in-degree indicates the total number of dependency relationships governed by the node word, and the out-degree indicates the total number of dependency relationships in which the node word is headed by other words.

Subsequently, *Pajek* (Version 6.01), a complex network analysis and visualization tool, is employed to analyse the dependency network. The tool can calculate centrality and degree distribution. Particularly, *Pajek* is primarily utilized to analyse the characteristics of the network and its nodes. What's more, *R* (R core tema, 2024) is employed to provide power-law fitting of the degree distribution.

The main analysis indicators of complex networks include the small-world property, the scale-free property, node degree, degree distribution, correlation, and centrality/centralization (Liu, 2017). Complex networks are widely used to analyse the structure and dynamics of various systems, including language. If a language complex network exhibits both the small-world and scale-free properties, it demonstrates the common characteristics of a complex network. These properties provide critical insights into network efficiency, robustness, and functionality, making them essential for understanding the underlying structure of complex systems.

To determine whether a network exhibits these properties, specific criteria must be met. A small-world network is defined by two key conditions: a high clustering coefficient (C) and a short average path length (d). The clustering coefficient measures how closely connected a node's neighbours are. For a network to be classified as small-world, its clustering coefficient (C) must be significantly higher than that of a random network (C_{rand}) with the same number of nodes and edges. The average path length represents the shortest path length averaged over all possible pairs of vertices in the network. In small-world networks, the average path length (d) should be approximately equal to that of a random network (d_{rand}), ensuring global efficiency in information or signal transmission. In summary, a small-world network is characterized by high local clustering ($C >> C_{rand}$) and short global distances ($d \approx d_{rand}$). A scale-free network, on the other hand, is characterized by a node degree distribution that follows or approximates a power-law distribution. In such networks, most nodes are sparsely connected, while a small number of nodes, referred to as hubs, have a disproportionately high number of connections. This leads to significant fluctuations in node degree values, making it difficult to define a characteristic degree, such as an average degree or eigendegree (Hudson, 2007; Liu, 2017). In essence, a scale-free network is defined by its power-law degree distribution and the presence of hubs, which contribute to its highly heterogeneous structure.

To further explore the internal structure of a language complex network and examine its local structural features, analysing the central nodes is essential. Central nodes play a crucial role in the emergence of small-world and scale-free properties, offering valuable insights into the organization and functionality of language networks (Chen & Liu, 2011). Studying these nodes can help uncover how linguistic structures are organized and how key elements contribute to the network's overall efficiency and robustness.

3 Results and Discussion

This section mainly answers four questions in the introduction part. First, the global and local characteristics of the dependency network of English e-commerce live-streaming discourse are analysed; then it focuses on the central nodes of this syntactic complex network.

3.1 Global characteristics

Global characteristics of the complex network refers to the small-world property and the scale-free property. By understanding the global characteristics, live-streamers can optimize the overall organization of their discourse to ensure clarity, efficiency, and persuasive power. For instance, small-world structures help create a seamless flow of information, making it easier for consumers to process and respond to the message. Scale-free structures emphasize the strategic use of high-frequency words to reinforce key ideas and keep consumers engaged.

The small-world property is widespread in various complex networks in the real world, and different types of language complex networks have also been found to have this property. For complex networks, the small-world property is very important, and it indicates that these networks have achieved a high degree of global and local connectivity with a smaller number of edges (Ferrer-i-Cancho & Solé, 2001; Ferrer-i-Cancho et al., 2004; Liu, 2022). Specifically, a linguistic complex network presents the small-world property, indicating that there is a short distance between the nodes of the network, and the nodes and their neighboring nodes can form a relatively dense local sub-network (Liu, 2017).

To determine whether a network has the small-world property, it is necessary to estimate the average path length and clustering coefficient of the corresponding random network. The random network refers to a network with the same number of nodes and the same number of edges, but the nodes are randomly connected. Table 1 lists the parameters of the dependency network of English e-commerce live-streaming discourse (ELD) and the corresponding random network. The parameters and their meaning are shown in Appendix I.

Table 1

Parameters of the dependency network of English e-commerce live-streaming discourse and the corresponding random network

N	E	D	k	d	d_{rand}	C	C_{rand}	γ	R^2
3,220	29,461	8	9.15	3.038	3.012	0.216	0.068	2.206	0.725

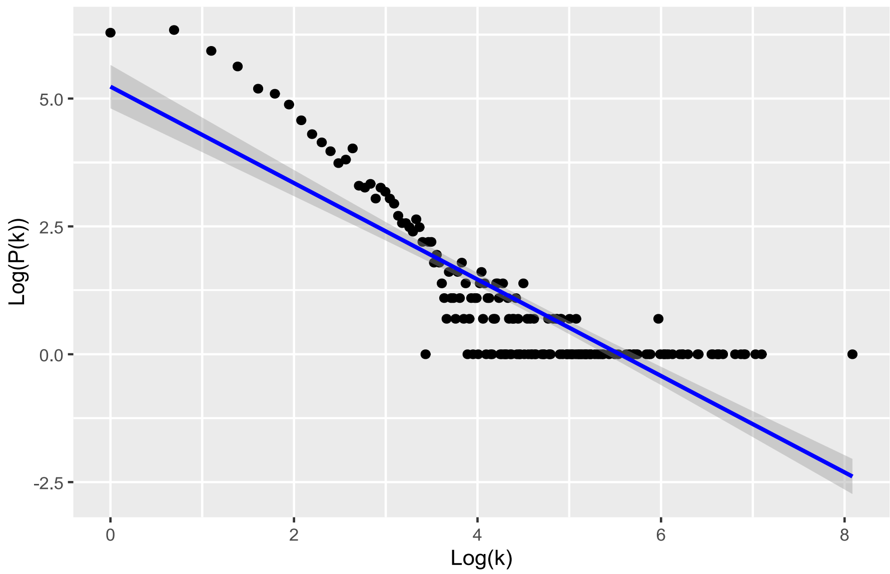
In the dependency network, the number of nodes refers to the number of node words, totalling 3,220. The number of edges in this syntactic network corresponds to the number of dependency relations between nodes, amounting to 29,461 edges. The network diameter, defined as the maximum distance between nodes, is 8. The average degree, representing the average number of connections per node, is 9.15. The average path length, which measures the average distance between any two nodes (3.038), exhibits a high degree of proximity with that of the corresponding random network (3.012). The clustering coefficient (C_i) measures the likelihood of an edge existing between a node's neighbouring nodes. It is calculated as the ratio of the actual number of edges among the neighbouring nodes to the maximum possible number of edges. For this network, the clustering coefficient is 0.216, much higher than that of the corresponding random network (0.068). Additionally, the degree distribution power-law exponent of this syntactic network is 2.206.

A network is considered a small-world network if it exhibits a greater clustering coefficient and shorter average path length. More formally, a small-world graph is defined as a network in which the average path length (d) is approximately equal to that of a random network (d_{rand}), while the clustering coefficient (C) is significantly larger than that of a random network ($C >> C_{rand}$) (Watts & Strogatz, 1998; Liu, 2008). It is found in Table 1 that the average path length of the undirected syntactic network of this treebank is approximately equal to that of the corresponding random network. This structure facilitates efficient global connectivity because the hubs that are a few nodes with a very high degree act as bridges that connect different parts of the network, reducing the average path length between nodes. Liu (2008) explained that syntactic networks often exhibit reduced average distances due to a language's natural tendency toward efficient global connectivity. Furthermore, the clustering coefficient of this network is significantly higher than that of the random network, indicating a strong tendency for local clustering. This discrepancy can be attributed to the inherent structural properties of the syntactic network, which reflect strong local clustering due to the grammatical and semantic relationships between words. Additionally, preprocessing steps such as lemmatization may have influenced the network's structure by reducing the number of nodes and concentrating connections, potentially amplifying the clustering coefficient. In contrast, the random network maintains the same number of nodes, with edges distributed randomly, resulting in a lower clustering coefficient. In conclusion, the syntactic complex network of English e-commerce live-streaming discourse demonstrates the small-world property, characterized by efficient global connectivity and high local clustering.

The scale-free property is also widespread in various complex networks. Solé et al. (2010) once pointed out that both the scale-free property and the small-world property are statistically based on universal language laws. The scale-free property refers to a characteristic of certain networks where the degree distribution (the number of connections each node has) follows a power-law distribution (Barabási & Albert, 1999). This means that most nodes have very few connections in a scale-free network, while a small number of nodes, called hubs, have a disproportionately large number of connections. Figure 3 is the degree distribution scatter plot of the syntactic complex network of English e-commerce live-streaming discourse.

Figure 3

Degree distribution scatter plot of the syntactic complex network of English e-commerce live-streaming discourse



From Table 1 and Figure 3, the power-law determination coefficient <R^2> of the degree distribution is 0.725. This finding is close to coefficient of determination for the power law of best fit to P(k) (0.743) and R^2 (0.79) in the study of Liu & Cong (2013) and Liu (2008) based on the novel and news respectively. This indicates that the degree distribution of the syntactic complex network of English e-commerce live-streaming discourse shows a relatively good fit to the power-law distribution. In other words, this syntactic complex network exhibits the scale-free property, as indicated by the degree distribution's reasonably good fit to the power-law distribution. The result matches with the findings based on different genres, such as the news (Liu, 2008; Chen & Liu, 2011; Cong & Liu, 2014), novel (Liu & Cong, 2013) and foreign language (L2) writing (Jiang et al., 2019). It also echoes with the hypothesis previously proposed by several scholars regarding universality of this network pattern in human language (Ferrer-i-Cancho, 2005; Masucci & Rodgers, 2006). What's more, the coefficient γ obtained in this study, 2.206, falls within the range of 2 to 3 reported in prior study (Barabási & Albert, 1999). They put forward that the network is a scale-free network if the constant is between 2 and 3. The combination of these two coefficients corroborates the scale-free nature of the syntactic network. The power-law degree distribution suggests that only a few nodes in the network have significantly larger degrees than others. These nodes function as hub nodes, playing a central role in maintaining the network's structure and connectivity (Barrat et al., 2008). This structural characteristic is not only a mathematical property but also closely related to the way language systems operate. The characteristic of scale-free networks has been extensively documented across a diverse array of linguistic complex networks (Ferrer-i-Cancho & Solé, 2001; Ferrer-i-Cancho et al., 2004; Liu, 2022). This pattern is particularly evident when examining networks derived from a variety of textual genres, such as novel (Liu & Cong, 2013) and news (Liu, 2022). Researchers have explored how these properties manifest in different genres to gain insights into the dynamics of language use and evolution. For instance, Liu (2022) constructed four types of undirected Chinese complex networks, including the semantic network, the syntactic dependency network, the word co-occurrence network and Chinese character co-occurrence network. All these networks exhibit the scale-free property, highlighting the universal presence of this feature in linguistic complex networks.

The scale-free property of the complex network reflects the principle of least effort in language, where language systems tend to balance efficiency for those who produce language and those who comprehend it. The principle of least effort, as proposed in linguistic studies, suggests that language evolves to minimize effort for both speakers and listeners. Each subsystem of the language tends to allow a few units with high connectivity or frequency to act as hubs. These highly connected units are used in a wide range of contexts, reducing the cognitive and physical burden on language producers. In contrast, most language units are used in relatively fixed contexts, simplifying the comprehension process for listeners. This balance between a few highly connected units and many specialized units ensures that language remains both efficient to produce and easy to comprehend, highlighting the adaptive nature of linguistic systems.

3.2 Local characteristics

In the previous section, we focused on the global characteristics of the syntactic complex network of English live-streaming e-commerce discourse, specifically its small-world and scale-free properties. In the following section, we delve deeper into the local characteristics of the network, which refer to the micro-level properties that describe the structural relationships between individual nodes (words) and their immediate surroundings within the network.

The central node serves as the primary entry point for examining the local characteristics of the network, as it plays a crucial role in shaping the small-world and scale-free properties of the network (Chen & Liu, 2011). To identify the most probable central nodes in this syntactic complex network, we utilized the British National Corpus (BNC) to compare high-frequency words. The BNC, originally created by Oxford University Press in the 1980s and early 1990s, contains 100 million words of text spanning a wide range of genres, including casual discussions, fiction, magazines, newspapers, and academic texts.

First, we calculated the top 20 high-frequency words in both the syntactic complex network and the BNC corpus, as shown in Table 2. The word frequency statistics for the BNC are based on the results provided by Adam Kilgarriff (available at http://www.kilgarriff.co.uk/bnc-readme.html). Frequency is a foundational metric in quantitative linguistics (Liu & Liang, 1986; Liu, 2017). As shown in Table 2, the top three high-frequency words in the syntactic complex network are *be*, *I*, and *the*. Their respective rankings in the BNC are 2, 12, and 1. Based on the comparison, these three words are identified as the central nodes of the syntactic network. It is similar to the finding of the previous research that shows the hubs are inclined to be function words (such as articles and prepositions) (Solé et al., 2010). Three words are central nodes of the syntactic network but have different status as their influence to the overall structure is quite different (Chen & Liu, 2016).

Table 2

The top 20 high-frequency words in our treebank and BNC

ELD Rank	Frequency	Word	BNC Rank	Frequency	Word
1	1,979	be	1	6,187,267	the
2	522	the	2	4,239,632	be
3	468	of	3	3,093,444	of
4	447	have	4	2,687,863	and
5	423	in	5	2,186,369	a
6	378	a	6	1,924,315	in
7	345	to	7	1,620,850	to
8	331	for	8	1,375,636	have
9	324	like	9	1,090,186	it
10	283	and	10	1,039,323	to
11	260	with	11	887,877	for
12	256	that	12	884,599	I
13	254	on	13	760,399	that
14	236	I	14	695,498	you
15	235	get	15	681,255	he
16	222	you	16	680,739	on
17	220	it	17	675,027	with
18	197	do	18	559,596	do
19	183	gon	19	534,162	at
20	182	love	20	517,171	by

Table 3

The top 20 word frequency ranks and degree ranks in the syntactic complex network of English e-commerce live-streaming discourse

Rank (Frequency)	Frequency	Word	Rank (Degree)	Degree	Word
1	1,979	be	1	2,283	be
2	522	the	2	1,130	I
3	468	of	3	1,061	the
4	447	have	4	1,039	you
5	423	in	5	978	it
6	378	a	6	782	and
7	345	to	7	742	a
8	331	for	8	651	that
9	324	like	9	606	to
10	283	and	10	552	this
11	260	with	11	485	so
12	256	that	12	479	of
13	254	on	13	466	have
14	236	I	14	443	do
15	235	get	15	423	in
16	222	you	16	394	we
17	220	it	17	359	like
18	197	do	18	287	not
19	183	gon	19	269	for
20	182	love	20	267	get

Table 3 presents the top twenty words ranked by frequency and degree in the syntactic complex network of English e-commerce live-streaming discourse. As shown in Table 3, the ranking of words by frequency and by degree is not entirely consistent. It echoes with the statement of Liu (2008) that the degree of a word is not equivalent to its frequency, although both are closely related. From Table 3, we can also see that functional or grammatical words play an important role in degree and frequency. Word frequency refers to the number of times a word appears in the discourse, while degree represents the number of direct connections a node has to other nodes in the network. For instance, *be*, *the*, and *of* are the three most frequent words, whereas the top three words by degree are *be*, *I*, and *the*. Notably, *so* appears 485 times in the corpus but does not rank among the top 20 by degree, highlighting the difference between frequency and degree as measures of centrality.

The fact that *I* and *you* are the second and fourth most frequent word in our treebank but ranks 12th and 14th in BNC respectively. The frequent use of *I* in e-commerce live-streaming suggests that hosts often rely on first-person narratives to share their personal experiences, opinions, or recommendations about products. This approach helps build credibility and fosters a sense of authenticity. The prominence of *you* reflects the direct engagement with the audience in live-streaming discourse. Hosts frequently address viewers personally to create a conversational tone and make the audience feel involved.

The degree of a node reflects its dominance or influence within the network. Nodes with high degrees are considered central; however, since degree only accounts for direct connections and ignores indirect ones, it is referred to as “local centrality.” In this study, *be* and *the* are identified as the nodes with the highest degrees, making them the local central nodes of the network. This finding is consistent with the results reported by Jiang et al. (2019). Similarly, it can be observed that these three nodes consistently maintain the leading positions. In their study, the degrees of *is*, *I*, and *the* are 120, 105, and 67, respectively, ranking first, second, and fifth. In our study, the degrees of *be, I,* and *the* are 2,283, 1,130, and 1,061, respectively, ranking first, second, and fourteenth. Such consistency in trends across different studies highlights the robustness of the observed phenomenon and underscores the need for further exploration into the underlying mechanisms driving these relationships.

To further analyze the network structure, we used the network analysis software *Pajek* (de Nooy et al., 2005), available at [http://mrvar.fdv.uni-lj.si/pajek/], to calculate the network parameters for *be, I,* and *the,* as well as the global network parameters after removing these three nodes. Initially, the networks were undirected but were converted into directed networks to better represent the specific relationships of the selected nodes. Table 4 displays the network parameters of *be, I,* and *the*, while Table 5 compares the global network parameters before and after removing these nodes. To ensure comparability and minimize statistical errors, all data were standardized. Values above the average were encoded as 1, while those below the average were encoded as 0. This standardization facilitates a clearer comparison of network properties and highlights the structural impact of the selected nodes.

Table 4

Network parameters of the nodes be, I, and the

Network	*be*	*I*	*the*
Node degree	2,283	1,130	1,061
Standardized node degree	1	1	1
In-degree	1,554	14	0
Standardized in-degree	1	1	0
Out-degree	729	1,116	1,061
Standardized out-degree	1	1	1
Closeness	0.56771	0.669141	0.500473
Standardized closeness	1	1	1
Internal closeness	0.38632	0.055794	0
Standardized internal closeness	1	1	0
External closeness	0.24297	0.965665	0.495756
Standardized external closeness	1	1	1
Betweenness	0.18096	0.18148	0.01500
Standardized betweenness	1	1	1
Betweenness centrality	0.64387	0.050082	0

Table 5

Comparison of network data after removing three nodes

	Number of nodes	Average degree	Average path length	Number of isolated nodes	Diameter	Density
Complete	3,220	9.1342	3.67478	0	9	0.002837
Remove *be*	3,212	8.4461	3.89109	244	11	0.001407
Remove *I*	3,219	8.7860	3.62861	132	9	0.001548
Remove *the*	3,219	8.8074	3.65141	331	9	0.001520

Centralization refers to the global centrality of the entire network, while centrality pertains to the importance of a specific node within the network. Local centrality is a metric in network analysis used to evaluate the importance of a node, with a specific focus on its position and role within its immediate network environment. It assesses a node's significance among its directly connected neighbors, which is often determined by the number of connections (i.e., degree) the node has. Nodes with high local centrality, known as local hub nodes, exert significant influence within their localized area of the network. These nodes hold key positions among their neighbors and can have a substantial impact on the structure and function of the local network.

The primary measures of local centrality include Degree Centrality, Betweenness Centrality, Closeness Centrality, and others. Degree centrality is the degree of the node, and specifically evaluates a node's importance based on its direct connections. In contrast, both closeness centrality and betweenness centrality take the entire network into account, thereby reflecting a node's global significance within the network (Liu, 2017).

From Table 4, the important network parameters values, such as node degree, closeness, and betweenness, for the high-frequency node words *be*, *I*, and *the* are listed. Node degree can be divided into in-degree and out-degree. In-degree refers to the number of nodes directly connected to a given node, while out-degree measures the number of nodes it directly connects to. The core idea of closeness is that a node occupies a central position in the network if it has short distances to other nodes. For directed networks, closeness can be further divided into internal closeness and external closeness. Betweenness, on the other hand, measures the extent to which a node acts as a "mediator" between other nodes in the network. A node with high betweenness can control the flow of information and thus plays a central role in the network.

The data were standardized to facilitate comparisons. Values above the average were assigned 1, while those below the average were assigned 0. As shown in Table 4, all standardized parameters—including node degree, in-degree, out-degree, closeness, internal closeness, external closeness, and betweenness—for the node words *be*, *I*, and *the* are equal to 1, representing the highest values among all nodes. This indicates that these three nodes function as local central nodes within the network, playing crucial roles in maintaining its structural integrity and connectivity.

Then, which node is the most probable central node of the network? Table 4 lists the changes in network data after removing the three nodes, providing insights into their structural impact on the overall network.

According to Chen et al. (2015), one method for network analysis involves assessing the degree of a node, which refers to the number of distinct types of connections a node can establish. In other words, a node with a higher degree—indicating its ability to connect to a greater number of other nodes and maintain more extensive connections—is more likely to be considered a "hub" or hold a pivotal position within the network's architecture. In a directed network, the centrifugal force of a node corresponds to its out-degree, while the centripetal force corresponds to its in-degree (Liu, 2017). By combining the data from Tables 4 and 5, it is evident that *be* exhibits a high degree, high in-degree, and high out-degree, making it a local central node. Consequently, *be* is more likely to serve as the global center of the network compared to *I* and *the*. Additionally, as shown in Table 3, there is no one-to-one correspondence between word frequency and centrality, indicating that word frequency alone cannot determine the centrality. However, *be* ranks first in both frequency and degree, leading to the initial conclusion that *be* is the most central node in the network.

Closeness centrality and betweenness centrality provide further evidence of the global importance of *be* within the network (Liu, 2017). The closeness centrality of *be* is intermediate among the three nodes, but its external closeness centrality is the lowest, indicating that *be* maintains short distances to neighboring nodes and occupies a central position in the network. Furthermore, the betweenness centrality of *be* is significantly higher than that of *I* and *the*, confirming its role as the most central node in the global network. The betweenness centrality of *be* is higher than that of *I*, and its frequency is also greater. This corresponds with Zipf's Principle of Least Effort (1949), which posits that language users tend to favor high-frequency words to minimize cognitive and expressive effort during communication. According to this principle, frequently used words like *be* are more likely to occupy central positions in communication networks, as they facilitate efficient information exchange. The high betweenness centrality of *be* further underscores its structural importance in connecting other nodes within the network. *A*dditionally, the higher betweenness centrality of *be* highlights its critical role as a bridge, connecting different parts of the network. This dual importance, in both frequency and network structure, aligns with its frequent use in natural language.

After removing *be*, several network characteristics undergo significant changes: The network density drops to approximately half of its original value; the average path length and diameter both increase; the number of isolated nodes rises sharply, with 244 nodes becoming disconnected from the network.

These changes demonstrate that *be* functions as a central hub, connecting many nodes and shortening distances between them. The increase in the network diameter after removing *be* suggests that it lies on many shortest paths, further explaining its high betweenness centrality. As both a local and global central node, *be* plays an indispensable role in maintaining the network's structure and connectivity.

When *the* is removed, the average degree, density, and average path length of the network all decrease, while the number of isolated nodes increases significantly. The degree of *the* is much higher than the average degree of the original network, so removing it naturally reduces the average degree. However, *the* exhibits unique characteristics: its in-degree, internal closeness centrality, and betweenness centrality are all zero. A betweenness centrality of zero indicates that *the* is not a global central node. An in-degree of zero suggests that *the* cannot dominate other nodes and can only function as a dependent. This is consistent with its role as a definite article. After removing *the*, the network density decreases to about 50% of its original value, and 331 nodes become isolated, highlighting *the*'s close connections with other nodes in the network.

Similarly, removing *I* results in decreases in the average degree, density, and average path length, along with an increase in the number of isolated nodes. However, the network diameter remains unchanged, showing a similar trend to the removal of *the*. The degree of *I* is higher than the average degree of the original network, so its removal naturally reduces the average degree. Nodes that rely on *I* to enter the syntactic network become isolated, although the number of isolated nodes is smaller compared to *the* or *be*. This is likely because *I* can function as both a head and a dependent, allowing some nodes to remain connected even after its removal. Unlike *the*, which only enters the network unidirectionally, *I* maintains bidirectional connections, similar to *be*. While *I* has the highest closeness centrality among the three nodes, its betweenness centrality is negligible, indicating that it is not the most important node in the global network. The impact of *I* on the network's average path length and density is comparable to that of *the*. After removing *I*, the network density is reduced to approximately half of its original value, suggesting that *I* also plays a significant role in connecting nodes within the network. In summary, while *be, I,* and *the* are all central nodes, their importance differs. *Be* is the most central node in the entire syntactic network, serving as both a local and global hub. In contrast, *the* and *I* are primarily local central nodes with more limited global influence.

Current complex network parameters primarily reveal the global characteristics of networks, providing insights into the global features of linguistic systems. Regardless of the type of linguistic network, nodes represent elements of the linguistic system, and the relationships between these elements, degree of the node, reflect their binding forces at the respective linguistic structural level. While global network parameters reveal the structural importance of *be*, a more detailed analysis of its syntactic role requires examining its binding forces within the framework of PVP theory.

According to the theory, the binding force of words is divided into centripetal force and centrifugal force. The centripetal force refers to the ability of a word (category) to be headed by other words (categories), while the centrifugal force refers to the ability of a word (category) to dominate other words (categories). Syntactic valency is closely related to dependency relationships. Liu and Feng (2007) introduced probability into valency research and proposed the probability valency pattern theory, which enhanced the application value of the PVP theory.

This theory emphasizes that valency patterns should be described both qualitatively, in terms of dependency relationships (centrifugal or centripetal), and quantitatively, through their probability distributions.

In the syntactic complex network of this English e-commerce live-streaming discourse, the verb *be* (including all lemmatized forms) is the most central node of the network, and its syntactic binding force is the strongest. Specifically, *be* participates in a total of 2,864 dependency relationships, of which 1,523 are heads, accounting for 53.18%, and 1,341 are dependents, representing 46.82%. This balanced distribution of centripetal and centrifugal forces indicates that *be* plays a dual role in the network, functioning both as a head and a dependent. These findings highlight the unique position of *be* as a central hub that connects various parts of the network, ensuring structural coherence.

This balanced distribution, combined with its high frequency and centrality, establishes *be* as the most critical node in the syntactic network of English e-commerce live-streaming discourse. Table 6 provides a detailed distribution of its binding force distribution.

Table 6

The Distribution of Binding forces of node be

Centrifugal Force				Centripetal Force			
Rank	**Dependency relations**	**Frequency**	**Percentage**	**Rank**	**Dependency relations**	**Frequency**	**Percentage**
1	nsubj	439	28.10%	1	aux	537	41.24%
2	acomp	251	16.07%	2	ccomp	189	14.52%
3	xcomp	159	10.18%	3	advcl	140	10.75%
4	advmod	146	9.35%	4	conj	133	10.22%
5	aux	123	7.87%	5	auxpass	102	7.83%
6	cc	81	5.19%	6	rcmod	63	4.84%
7	prep	78	4.99%	7	xcomp	58	4.45%
8	mark	68	4.35%	8	dep	39	3.00%
9	conj	51	3.27%	9	parataxis	18	1.38%
10	advcl	48	3.07%	10	advmod	6	0.46%
Total	1444		92.45%	Total		1285	98.69%

As shown in Table 6, sorted by frequency distribution, the top 10 dependency relationships of the centrifugal force of the node *be* account for 92.45%, while the top 5 relationships make up 71.57%. The most frequent centrifugal force is the subject relationship (nsubj). The top 10 dependency relationships of the centripetal force of *be* constitute 98.69%, with the top 5 relationships accounting for 84.56%. The most frequent centripetal force is the auxiliary relationship (aux).

Specifically, the centrifugal force of *be* includes subject relationship, adjective complement (acomp), open clausal complement (xcomp), adverb modifier (advmod), auxiliary (aux), coordination (cc), prepositional modifier (prep), marker (mark), conjunct (conj), and adverbial clause modifier (advcl). The centripetal force of *be* includes auxiliary, clausal complement (ccomp), adverbial clause modifier, conjunct, passive auxiliary (auxpass), relative clause modifier (rcmod), atypical dependency relationship (dep), parataxis (parataxis), and adverb modifier.

In terms of centrifugal force, the most frequent relationship is the subject relationship, where *be* serves as the head word. Among its subjects, there are 1082 pronouns (77.2%) and 270 nouns (19.2%). Among pronoun subjects, *I* appears most frequently (43.4%), followed by *they*, *you*, *we*, and *he*. Notably, the third-person singular pronoun *he* appears only six times

(2.2%). This indicates that English e-commerce live-streaming hosts adopt a personalized and experiential approach when introducing and promoting products. The high frequency of first-person pronouns, particularly *I*, suggests that hosts rely on personal experiences and opinions to establish credibility and build trust with their audience. The low frequency of third-person singular pronouns like *he* further highlights the focus on direct interaction and personal connection rather than impersonal descriptions. This approach is a defining characteristic of e-commerce live-streaming discourse, where the goal is to make promotions feel genuine and trustworthy. For example:

Example 2: I was pretty impressed with the quality. I am more impressed with the fact that all I have to do is turn(ing) it upside down, and it definitely has a good grind.

Additionally, two other dependency relations, adjective complement and adverb modifier, account for 25.42%, emphasizing the central role of descriptive and attributive constructions in the analyzed texts. This underscores the importance of be as a linguistic tool for linking subjects with descriptive elements, which is especially critical in e-commerce texts. The primary goal is to describe products, services, or processes in a clear, concise, and persuasive manner, distinguishing e-commerce texts from other genres. For instance, academic prose often features highly informational content, frequent nouns, prepositions, and longer words (Biber, 1988). In contrast, e-commerce communication focuses on engaging and influencing potential customers through evaluative and promotional language.

In terms of centripetal force, *be* serves as the dependent in various dependency relationships. The second most frequent centripetal relationship is the clausal complement (ccomp), where a dependent clause functions as an object of the verb or adjective[2]. For example:

Example 3:
That does not **mean** it **is** useful to me.

The high frequency of the clausal complement involving *be* highlights its role in embedding subordinate clauses into larger sentence structures, enabling texts to express nuanced and complex meanings. In this treebank, *be* frequently appears as the verb of the subordinate clause, with its top three high-frequency heads being *think* (29), *be* (28), and *know* (21). For example:

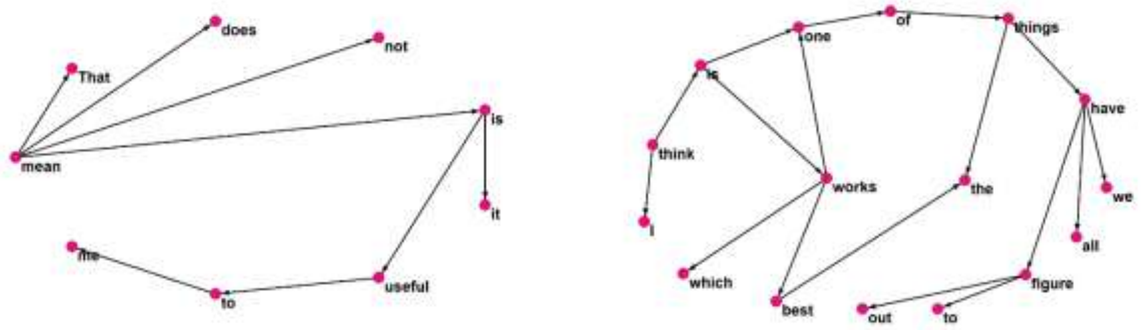
Example 4: I **think** one of the things we all have to figure out **is** which one works the best.

These high-frequency heads suggest that e-commerce live-streaming discourse often features subjective judgments, explanations, or factual assertions, which are defining characteristics of this genre. The primary goal is to persuade or inform the audience about products or services.

2 Marie-Catherine de Marneffe and Christopher D. Manning. (2008). “Stanford typed dependencies manual”. Stanford University, Technical report. https://downloads.cs.stanford.edu/nlp/software/dependencies_manual.pdf.

Figure 3

Directed syntactic dependency network of Example 3 and Example 4



4 Conclusion

The syntactic network of English e-commerce live-streaming discourse exhibits both small-world and scale-free properties, aligning with the typical characteristics of complex networks. Within this network, the central nodes are *be*, *I*, and *the*, though their roles differ. *Be* emerges as the most central node, while *I* and *the* function as local central nodes. The prominence of these nodes reflects the linguistic features of e-commerce discourse: descriptive, user-focused, and specific. This pattern underscores the functional purpose of such discourse—to convey information clearly while engaging users through personalization and specificity.

The study also highlights the robustness of the syntactic structure in English e-commerce live-streaming discourse. Even with the removal of the most central node, *be*, the majority of the network's connectivity remains intact. Regarding the centrality of *be*, its out-degree and in-degree are nearly balanced. From the perspective of probability valency pattern theory and dependency grammar, this balance indicates that *be* exerts both a strong centrifugal force as a head in dependency relations and a strong centripetal force as a dependent. The centrifugal force of *be* mainly encompasses roles such as subject relations and adjective complements, emphasizing its function in attributive and descriptive constructions—key elements in genres where descriptions and attributes are central, such as advertising and product descriptions.

Conversely, the centripetal force of *be* mainly involves syntactic roles such as auxiliary verb relations and clause complements, highlighting its importance in complex grammatical constructions like progressive tenses, passive voice, and subordinate clauses. These dual forces position *be* as both a *descriptive anchor* and a *grammatical enabler* in e-commerce live-streaming discourse. Its binding forces support the primary goals of e-commerce texts: to describe products clearly and explain processes effectively. As such, *be* serves as a central and indispensable verb in the linguistic structure of e-commerce content.

The research methods applied in this study not only provide a macroscopic analysis of the global characteristics of the language network but also enable a microscopic examination of individual nodes within the network. This dual approach offers significant methodological support and technical tools for investigating linguistic structural features. In the next phase of research, we will conduct a comparative analysis of the syntactic network characteristics across other genres, as well as Chinese e-commerce live-streaming discourse and Chinese-English mixed e-commerce live-streaming discourse.

Acknowledgement

Funding: This article is supported by Zhejiang Provincial Philosophy and Social Sciences Planning Project (24JCXK05YB).

References

Barabási, A., & Albert, R. (1999). Emergence of scaling in random networks. *Science 286*(5439), 509–512. https://doi.org/10.1126/science.286.5439.509

Barrat, A., Barthelemy, M., & Vespignani, A. (2008). *Dynamical processes in complex networks*. Cambridge University Press.

Biber, D., (1988). *Variation across speech and writing*. Cambridge University Press.

Chen, X., & Liu, H. (2011). Central nodes of the Chinese syntactic networks. *Chinese Science Bulletin 56*(10), 735–740.

Chen, X., Liu, H., & Kim, G. (2015, August 24–26). *Proceedings of the Third International Conference on Dependency Linguistics*, pp. 74–81, Uppsala, Sweden.

Chen, X., & Liu., H. (2016). Function nodes in the Chinese syntactic networks. In A. Mehler, A. Lücking, S. Banisch, P. Blanchard, & B. Job (Eds.), *Towards a theoretical framework for analyzing complex linguistic networks. Series on understanding complex systems*, Springer.

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

De Nooy, W., Mrvar, A., & Batagelj, V. (2005). *Exploratory social network analysis with Pajek*. Cambridge University Press.

Ferrer-i-Cancho, R., & Solé, R. (2001). The small world of human language. *The Royal Society of London. Series B: Biological Sciences 268*(1482), 2261-2265. https://doi.org/10.1098/rspb.2001.1800

Ferrer-i-Cancho, R., Solé R., & Köhler, R. (2004). Patterns in syntactic dependency networks. *Physical Review E 69*(1), 51915. https://doi.org/10.1103/PhysRevE.69.051915

Ferrer-i-Cancho, R. (2005). The structure of syntactic dependency networks: Insights from recent advances in network theory. In G. Altmann, V. Levickij & V. Perebyinis (Eds.), *The problems of quantitative linguistics* (pp. 60–75), Ruta, Chernivtsi.

Hao, Y., Wang, X., Bin, S., Yang, Q., & Liu, H. (2024). How syntactic complexity indices predict Chinese L2 writing quality: An analysis of unified dependency syntactically–annotated corpus. *Assessing Writing 61*, 100847. https://doi.org/10.1016/j.asw.2024.100875

Hudson, R. (1984). *Word grammar*. Blackwell.

Hudson, R. (2007). *Language networks: The new word grammar*. Oxford University Press.

Hudson, R. (2010). *An introduction to word grammar*. Cambridge University Press.

Jiang, J., Yu, W., & Liu, H. (2019). Does scale-free syntactic network emerge in second language learning? *Frontier in Psychology 10*, 925. https://doi.org/10.3389/fpsyg.2019.00925

Liu, D., & Zhang, Z. (2021). An investigation into the reliability of automatic dependency parsing tools in Chinese-English cross language studies. *Foreign Language Research* (6), 9–16.

Liu, Y., & Liang, N. (1986). Basic engineering for Chinese processing modern Chinese words frequency count. *Journal of Chinese Information Processing* (1), 17–25.

Liu, H. (2008). The complexity of Chinese dependency syntactic networks. *Physica A 387*(12), 3048–3058. https://doi.org/10.1016/j.physa.2008.01.069

Liu, H. (2017). *An introduction to quantitative linguistics*. The Commercial Press.
Liu, H. (2022). *Dependency relation and language networks*. China Science Publishing & Media Ltd.
Liu, Q., & Cheng, W. (2025). “I’m telling you”: The use of interactional metadiscourse in Chinese live streaming commerce. *Journal of Pragmatics 237*, 14–29. https://doi.org/10.1016/j.pragma.2025.01.001
Liu, H., & Cong, J. (2013). Language clustering with word co-occurrence networks based on parallel texts. *Chinese Science Bulletin 58*(10), 1139–1144. https://doi.org/10.1007/s1143 4-013-5711-8
Liu, H., & Feng, Z. (2007). Probabilistic valency pattern theory for natural language processing. *Linguistic Sciences* (5): 32–41.
Ma, H., & Mei, H. (2018). Empirical research on the decision-making influence factors in consumer purchase behavior of webcasting platform. In *Proceedings of the Twelfth International Conference on Management Science and Engineering Management* (pp. 1017–1028). Springer, Cham.
Masucci, A. P., & Rodgers, G. J. (2006). Network properties of written human language. *Physical Review E*74, 026102. https://doi.org/10.1103/PhysRevE.74.026102
Ouyang, J., & Jiang, J. (2018). Can the probability distribution of dependency distance measure language proficiency of second language learners? *Journal of Quantitative Linguistics 25*(4), 295–313. https://doi.org/10.1080/09296174.2017.1373991
Shi, X., & Dou, H. (2023). How broadcasters enhance rapport with viewers in live streaming commerce: A genre–based discourse analysis. *Pragmatics 33*(4), 592–617. https://doi.org/10.1075/prag.22009.shi
Solé, R., 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
Watts, D., & Strogatz, S. (1998). Collective dynamics of ‘small-world’ networks. *Nature* (393), 440–442. https://doi.org/10.1038/30918
Xu, S. (2009). *Exploration in language research*. The Commercial Press.
Yang, N., & Wang, Z. (2022). Addressing as a gender-preferential way for suggestive selling in Chinese e-commerce live streaming discourse: A corpus-based approach. *Journal of Pragmatics* (197), 43–54. https://doi.org/10.1016/j.pragma.2022.05.014
Zipf, G. K. (1949). *Human behaviour and the principle of least effort: An introduction to human ecology*. Addison-Wesley Press.

Appendix Ⅰ Parameters and their meaning

Parameter	Meaning
N	Nodes
E	Edges
D	Network Diameter
d	Average Path Length
d_{rand}	Average Path Length of the Random Network
C	Clustering Coefficient
C_{rand}	Clustering Coefficient of the Random Network
γ	Power-law Exponent of the Degree Distribution
R^2	Power-law Determination Coefficient of the Degree Distribution
Node degree	Node degree is the number of edges directly connected to a node.
In-degree	In a directed network, the in-degree is the number of edges that other nodes point to a node.
Out-degree	In a directed network, out-degree is the number of edges from a node that point to other nodes.
Closeness	Closeness measures the inverse of the average distance of a node to all other nodes in the network.
Internal closeness	Internal closeness is the reciprocal of the average distance of a node to all other nodes in the same community.
External closeness	External closeness is the reciprocal of the average distance of a node to all other nodes in a different community (or network).
Betweenness	Betweenness measures the proportion of all shortest paths in the network that a node has.
Betweenness centrality	Betweenness centrality is the same as betweenness centrality, and refers to the proportion of a node's shortest path in a network.

Appendix Ⅱ Abbreviation of dependency relations and their meaning mentioned in this study

Abbreviation	Full form
acomp	adjective complement
advcl	adverbial clause modifier
advmod	adverb modifier
aux	auxiliary
auxpass	passive auxiliary
cc	coordination
ccomp	clausal complement
conj	conjunct
cop	copula
dep	dependent
mark	marker
nsubj	nominal subject
parataxis	parataxis
prep	prepositional modifier
rcmod	relative clause modifier
xcomp	open clausal complement