

Article

A Social Network Analysis on the Danmaku of English-Learning Programs

Man-Ni Chu ¹, Xin Huang ¹, Jia-Lien Hsu ² and Hai-Lun Tu ³, *

¹ Graduate Institute of Cross-Cultural Studies, Fu Jen Catholic University, New Taipei City 242062, Taiwan; mannichu@gmail.com (M.-N.C.); hx2504541481@gmail.com (X.H.)

² Department of Computer Science and Information Engineering, Fu Jen Catholic University, New Taipei City 242062, Taiwan; alien@csie.fju.edu.tw

³ Department of Library and Information Science, Fu Jen Catholic University, New Taipei City 242062, Taiwan

* Correspondence: 148245@mail.fju.edu.tw

Abstract: This study utilizes the danmaku on the Bilibili platform as the research subject to examine how their characteristics vary according to the nature or focus of English teaching videos. By employing social network analysis, the study reveals distinctive features in danmaku. For videos categorized under linguistic knowledge (phonetics, vocabulary, and grammar), the danmaku comments predominantly center around topics such as phonetics, vocabulary, and grammar. Conversely, in videos categorized under language skills (listening, speaking, reading and writing), the danmaku comments primarily reflect a vocabulary review for three of the four skills, with only the listening skill showing slight deviations. This underscores the centrality of vocabulary in skill-oriented videos. The findings highlight the unique role of danmaku in distinguishing between knowledge and skills within the context of English teaching videos.

Keywords: social network analysis; danmaku; English knowledge and skills

1. Introduction



Academic Editor: Andrea Prati

Received: 22 December 2024

Revised: 4 February 2025

Accepted: 8 February 2025

Published: 13 February 2025

Citation: Chu, M.-N.; Huang, X.; Hsu, J.-L.; Tu, H.-L. A Social Network Analysis on the Danmaku of English-Learning Programs. *Appl. Sci.* **2025**, *15*, 1948. <https://doi.org/10.3390/app15041948>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Numerous studies have explored online interactive models in education, ranging from MOOCs with pre-recorded videos to live synchronous classes. Among these, danmaku, a unique form that combines the advantages of pre-recorded content and real-time interaction, represents a third category. Danmaku, when applied to online learning, enables students to actively engage with course content through peer participation in danmaku interactions. This engagement is influenced by the intensity and design of the curriculum, creating a learning atmosphere that fosters immersion, particularly in professional contexts. Such phenomena underscore the contributions of danmaku to online learning environments.

Implemented on Bilibili, one of China's major video-sharing platforms, danmaku features a live commenting system that allows users to engage with peers asynchronously. This interaction fosters a sense of community and enhances user engagement in learning (e.g., [1] for English; [2] for Mathematics). The sense of synchronicity generated by danmaku not only enriches the content over time but also gives learners the impression of studying collaboratively, thereby boosting motivation and enthusiasm for learning.

Compared to MOOCs, danmaku offers a unique advantage in enhancing participant engagement. As highlighted by [3], a study involving 4466 participants across 10 highly rated MOOCs emphasized the importance of peer interaction in fostering engagement. Unlike structured peer interactions in MOOCs, danmaku is entirely learner-generated, making it an exemplary form of authentic peer interaction in online learning. Jiang et al.

(2022) [1] compared the learning experiences provided by MOOCs and Bilibili, concluding that Bilibili offers a superior environment for fostering engagement due to its interactive features. Specifically, danmaku demonstrated significantly higher effectiveness in stimulating learning interest compared to MOOC platforms. While there was no notable difference in grammar acquisition, danmaku was more effective in enhancing vocabulary acquisition, linguistic intuition, and conversational fluency. Furthermore, Zhang et al. (2023) [4] highlighted the use of a series of L2 vlogs on Bilibili for Spanish learning, emphasizing the critical role of interaction-oriented learning.

In a related study, Yang (2020) [5] investigated the influence of danmaku videos on learners' social interaction and their role in increasing motivation and engagement. The interactive nature of danmaku strengthens the sense of connection among learners, positively impacting participation, comprehension, and learning outcomes. The cumulative nature of danmaku enables the presentation of diverse perspectives, as learners from different backgrounds share insights, enriching the viewing experience and encouraging critical thinking.

Zeng et al. (2024) [6] integrated danmaku into educational data analysis using the TextMind software for psycholinguistic analysis (https://www.researchgate.net/publication/285653495_Developing_Simplified_Chinese_Psychological_Linguistic_Analysis_Dictionary_for_Microblog, accessed on 7 February 2025). Their study examined 58,143 danmaku comments in an online course on the fundamentals of digital electronics. The results demonstrated how danmaku fosters engagement, offering personalized recommendations to students and practical guidance for improving participation in online education platforms. However, contrary findings were reported by [7], who analyzed the use of danmaku in TED-Ed science videos. They found that merely increasing the volume of comments failed to facilitate deep learning. Similarly, Li et al. (2022) [8] observed that danmaku did not meet learners' expectations when the interaction between learners and teachers remained one-sided, with no feedback from the instructors. In Table 1, we show a summary of prior studies.

Table 1. The summary of prior studies on danmaku.

Authors and Year	Domain and Platform	Investigation	Main Technology
Yang (2020) [5]	General, Bilibili	The influence of danmaku videos on learners' social interaction and their role in increasing motivation and engagement	The two-pronged model combining a semiotic resource perspective and a social practice perspective
Peng & Wang (2021) [7]	Science, TED-Ed science videos	The impact of danmaku in learning: the correlation between the number of danmaku comments which subjects leave and the test of comprehension	Spearman's rank correlation analysis
Li et al. (2022) [8]	General, Video lectures	Identifying the interaction mode that danmaku meets	Semi-structured interviews
Jiang et al. (2022) [1]	English, MOOCs and Bilibili	Comparing the learning experiences provided by MOOCs and Bilibili	Interviews and statistics
Zeng et al. (2024) [6]	General, Bilibili	To understand students' learning patterns and present corresponding intervention strategies for different types of students	TextMind

Previous studies on danmaku have primarily focused on aspects such as the timing of messages, textual content, and emotional expressions, with an emphasis on data mining and text analysis. Most research has concentrated on danmaku in different types of open courses, often using individual videos as the primary research subjects, with limited

exploration of systematic learning collections. Moreover, studies specifically targeting systematic English teaching through danmaku remain scarce. To address this gap, the present study focuses on a systematic collection of English learning materials, using learners' danmaku data as the primary research subject.

In this study, we propose a social network analysis (SNA) approach to visualize the danmaku for better understanding and revealing the interaction. This study diverges from [6] by employing SNA to highlight the relevance of danmaku to various topics in English language instruction. Using Python-based web scraping to collect danmaku data, we aim to examine whether the interactions are related to specific subcategories of English instruction, such as phonetics, vocabulary, and grammar in professional contexts, as well as listening, speaking, reading, and writing in skills-based contexts. This approach seeks to determine whether danmaku is aligned with the two overarching concepts of English teaching—linguistic knowledge and language skills. Moreover, this research intends to explore whether danmaku facilitates a better learning environment through cumulative interactions or, conversely, whether an overload of information leads to distractions.

We hypothesize that while the danmaku mechanism can enhance the interactive experience, excessive engagement might hinder effective cognitive processing, impeding the deep learning emphasized in educational frameworks. In support of this, Li et al. (2022) [8] noted that 40.9% of the interactions in English videos were related to supplemental knowledge and answering queries, highlighting the potential for meaningful engagement. This study seeks to uncover whether similar patterns emerge in our analysis and whether danmaku can indeed provide a conducive environment for deep learning.

The other parts of this paper are composed as follows. In the Related Work Section, we present the use of SNA alongside related technologies. The Methods Section details our research methodology. In Section 4, we showcase the results of our study, followed by a discussion and conclusion in Section 5. The Section 5 summarizes key findings and provides insights for future research directions.

2. Related Work

In this section, we present the related work of SNA and the technologies in our research.

2.1. Social Network Analysis

Social network analysis (SNA) has become an essential area of research, particularly in computer science and social sciences. SNA is defined as the study of social structures through the use of networks and graph theory. Serving as a powerful framework for understanding complex interactions within various fields, SNA examines how relationships between individuals influence the behaviors and outcomes within a network. SNA has evolved from sociological roots to a multidisciplinary approach that integrates insights from computer science, economics, and organizational studies [9]. Some methodologies in SNA are proposed, such as "Centrality Measures" [10] and "Community Detection" [10,11]. SNA can be utilized in various domains, including co-authorship networks, social media analytics and epidemiology [12]. Recent studies indicate that SNA is experiencing rapid growth, particularly with the advent of big data analytics and machine learning techniques. However, challenges persist, such as data privacy concerns and the need for more robust analytical frameworks to assess user influence effectively [13]. The integration of machine learning with SNA is seen as a promising direction for future research [13].

The evaluation metrics of a social network graph are centrality, degree, betweenness, closeness, eigenvector centrality, diameter/radius, average geodesic distance, average degree, reciprocity, density, and global clustering coefficient [14]. In our research, we utilize eigenvector centrality for our evaluation.

Eigenvector centrality [15] evaluates a node's importance based on the importance of its neighbors, in contrast to degree centrality, which only considers the number of direct connections. As a result, eigenvector centrality provides a more comprehensive assessment of node significance in a network [16], incorporating the influence of well-connected nodes with high centrality [17].

2.2. Pre-Trained Language Model

In recent years, pre-trained language models have become a pivotal technology in the field of natural language processing (NLP) [18]. RoBERTa (robustly optimized BERT pre-training approach) builds upon BERT (bidirectional encoder representations from transformers) [19] with several significant improvements, including the use of larger training datasets, extended training duration, larger batch sizes, and longer input sequences. Additionally, it removes the next sentence prediction (NSP) task and adopts a dynamic masking strategy for the masked language model (MLM) task [20].

The **Chinese-RoBERTa-WWM-Ext-Large** is a pre-training model of Chinese BERT for its advanced understanding of Chinese language tasks. This improves its ability to capture contextual meaning, making it highly effective for tasks like text classification, sentiment analysis, and question answering in Chinese [20].

2.3. Clustering Algorithms

K-means, first introduced by [21] is a partitional clustering method developed for classifying and analyzing multivariate observational data. The algorithm partitions the data into k clusters by minimizing the average squared distance between points within the same cluster. Its main advantages are simplicity and speed [22]. K-means is a partitional clustering technique within cluster analysis, an unsupervised exploratory method that is generally classified into two categories: hierarchical and partitional clustering. Hierarchical clustering constructs a tree-like structure by iteratively merging or splitting clusters, ultimately forming a complete hierarchical structure. In contrast, partitional clustering methods, such as K-means, divide the data into a predefined number of clusters, with each data point assigned to exactly one cluster, without any hierarchical relationships [23].

The K-means algorithm is one of the most common, unsupervised methods for its simplicity, efficiency, and scalability in clustering tasks. The K-means algorithm performs well when the number of clusters is predefined, and the dataset is structured, making it ideal for segmenting data into distinct groups.

2.4. TF-IDF

TF-IDF is a classic method for measuring term importance, combining term frequency (TF), which reflects the significance of a term within a document, and inverse document frequency (IDF), which gauges its distribution across the entire corpus. Rare terms are assigned higher weights due to their greater discriminative value [24]. Initially proposed by [25] in the field of information retrieval, this concept highlights the importance of both term frequency and specificity for effective retrieval. TF-IDF, combining TF and IDF, has since become a fundamental approach in information retrieval [26]. In *Introduction to Modern Information Retrieval* [24], cosine similarity is used to compute the similarity between a query and a document by measuring the cosine angle between their respective vectors. This study employs cosine similarity to calculate the similarity between danmaku vectors.

2.5. The Levenshtein Distance-Based Method

The Levenshtein distance-based method uses Levenshtein's algorithm [27], which measures the minimum number of edit operations (insertion, deletion, substitution) required to transform one string into another, to calculate text similarity. In this study,

an undirected, unweighted edge is created between two danmaku nodes if the Levenshtein distance between their texts is 1. Originally developed for error correction in binary data [27], the Levenshtein algorithm has been widely applied in fields such as computational linguistics [28] and bioinformatics [29].

3. Methods

In this section, we present our research methods, including data collection, and social network analysis procedure.

3.1. Data Collection

This study adopted a systematic computational methodology to analyze danmaku data and construct a social network representing user interactions. The workflow, shown in Figure 1, comprises five key stages: web scraping, preprocessing, embedding, clustering, and network construction.

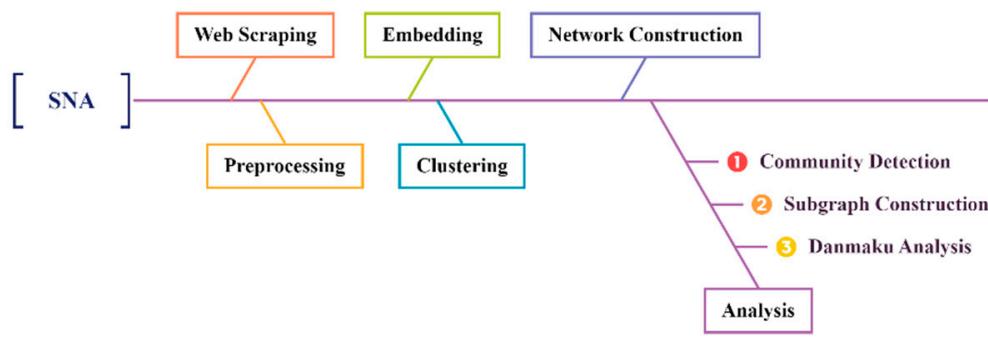


Figure 1. SNA workflow.

3.1.1. Web Scraping

Initially, web scraping was employed to extract data from the Bilibili platform, focusing on relevant videos and their associated danmaku comments. This process ensured the comprehensive collection of the raw textual data necessary for subsequent analytical tasks.

This study was conducted on Bilibili, a comprehensive video-sharing platform, established in 2009. To identify relevant content, searches were performed using keywords such as “English pronunciation”, “English vocabulary”, “English grammar”, “English listening”, “English speaking”, “English reading”, and “English writing”. The primary selection criterion was the quantity of danmaku comments (real-time comments displayed on videos). Secondary factors, including playback counts, coin donations, likes, and shares, were also considered. From the seven identified categories, the three most popular English learning collections were selected for analysis. Each collection comprises a varying number of videos.

Eventually, data for this study were collected between October and November 2023, encompassing 21 English learning collections on Bilibili. These collections comprised a total of 2057 individual videos and generated 1,721,873 danmaku comments. The danmaku data included both the text content (danmaku comments) and the sender’s unique user ID (UID), representing interactions from 331,263 participants.

3.1.2. Preprocessing

We develop a preprocessing pipeline to standardize and prepare the textual data for analysis, involving three primary steps: decoding, text normalization, and deduplication. Initially, HTML-encoded entities in the danmaku text (e.g., <, &) were decoded using the `html.unescape` function to restore the original user input. Subsequently, text normalization was performed, which included converting Chinese text to simplified characters, trans-

forming English text to lowercase, and converting full-width characters to their half-width equivalents. Lastly, to address the issue of excessive repetition in online content, sequences of more than three consecutive identical characters were truncated to a maximum of three. This step preserved the semantic integrity of expressions such as “好好學習” (study diligently) and internet slang such as “666” (indicating admiration).

3.1.3. Embedding

Following preprocessing, sentence embeddings were generated using the Chinese-Roberta-WWM-Ext-Large model, a pre-trained transformer-based model optimized for capturing nuanced semantic relationships in Chinese text, serving as a foundation for subsequent social network analysis.

3.1.4. Clustering

Sentence embeddings served as the input for the K-means clustering algorithm, which categorized the danmaku comments into distinct clusters. The number of clusters k in this study is determined using an empirical rule. As shown in Equation (1), k is the floor of the square root of n , with n representing the data size.

$$k = \lfloor \sqrt{n} \rfloor \quad (1)$$

3.1.5. Network Construction

The final step before analysis involved constructing a social network model for each collection based on two fundamental concepts: “User behavior” and “textual similarity between danmaku comments”. “User behavior” is related to a danmaku submission, characterized by the content and frequency of danmaku comments, while textual features are analyzed using the K-means clustering algorithm to explore relational patterns in messages.

The network consists of three types of nodes—users, danmaku, and clusters. To facilitate visualization in Gephi [30], node-related information is stored in CSV files, which include the node content and shape (polygon). The node content includes user IDs, danmaku comments, and the cluster assignment of each comment. To distinguish danmaku comments that consist of a single number, cluster nodes are labeled as “# + cluster number”. For instance, #0 and #1 represent the first and second clusters identified by the K-means algorithm, respectively. The “polygon” attribute is used to differentiate node types in Gephi: user nodes are represented as circles (polygon = 1), danmaku nodes as squares (polygon = 4), and cluster nodes as pentagons (polygon = 5).

The network contains two types of undirected edges: user–danmaku edges (U-D) and danmaku–cluster edges (D-C). U-D represents the relationship between users and the danmaku comments they submit, forming a many-to-many relationship where one user can submit multiple danmaku comments, and a single danmaku comment can be submitted by multiple users. The weight of the edge reflects the number of times a user submits the same content, with a minimum weight of 1. For example, if user “U1” submits danmaku comment “D1” three times, the edge weight between U1 and D1 would be 3. This could occur if the user is watching a series and submits the same comment, “D1”, at different times. D-C represents the relationship between danmaku comments and the clusters to which they are assigned, as determined by the K-means algorithm described in the previous section. These edges are unweighted and form a many-to-one relationship, where each danmaku comment belongs to a single cluster, but each cluster may contain multiple danmaku comments. For instance, if danmaku comments “D1” and “D2” are both assigned to cluster 1, each will be connected to node “C1” by an unweighted, undirected edge.

To ensure the reproducibility of the results during the execution of the Python process, a random seed of 42 was set.

Based on these concepts, this study constructs a social network model as Figure 2 with three types of nodes and two types of edges to investigate interaction characteristics between users in different clusters. It analyzes the behavioral patterns of user interactions through danmaku and identifies the key topics that drive user participation in danmaku interactions.

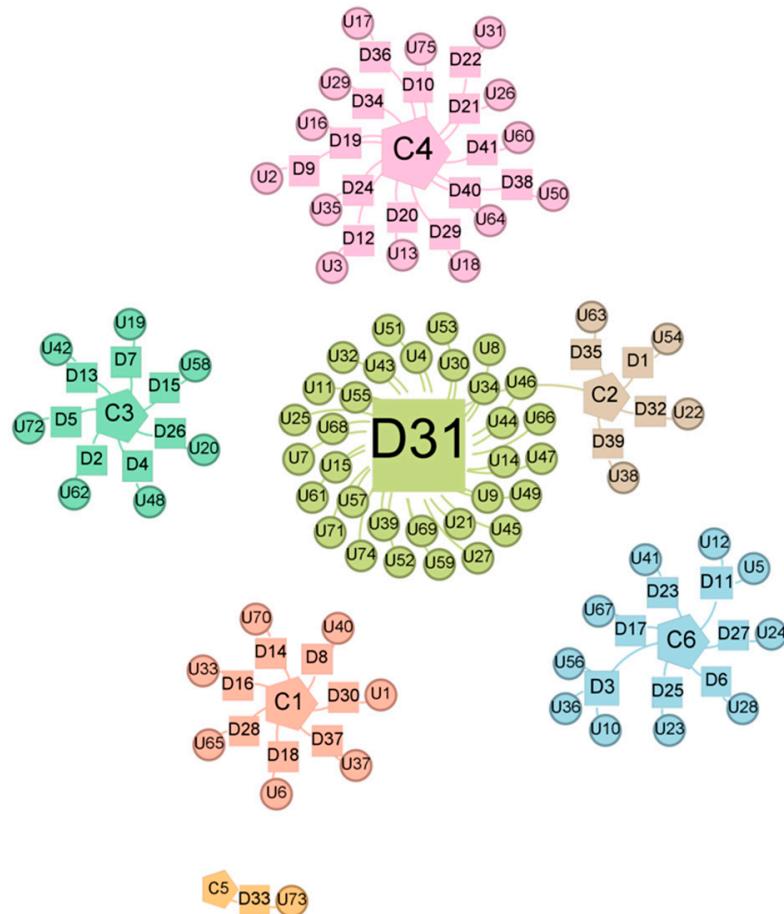


Figure 2. Gephi visualization of social networks: a case study. The detail of danmaku nodes is presented in Appendix A. Different colors in the figure represent different communities.

3.2. Social Network Analysis Procedure

In this phase, the study focuses on analyzing danmaku nodes within the top three subgraphs of each collection, leveraging insights gained from community detection methods.

3.2.1. Community Detection

After constructing the network, community detection was performed. Community detection aims to uncover naturally occurring groups or clusters within a network without prior knowledge of the number or size of these groups [17]. A common approach involves maximizing the modularity score, which evaluates the quality of a particular division of the network into communities [17]. This study follows a similar approach. According to research by [31], the Leiden algorithm outperforms the heuristic Louvain algorithm [32] in terms of both speed and the quality of community connectivity. Therefore, this study applies the Leiden algorithm to analyze 21 social networks, aiming to obtain the modularity and community structure of an integrated network comprising three types of nodes.

Modularity measures the extent of assortative mixing, where nodes with similar attributes tend to form connections within the same community. A higher modularity value, approaching 1, indicates a strong presence of intra-community connections relative to a randomized network, thereby reflecting significant structural properties of the network [17].

The complexity of social networks means that factors such as the diversity and frequency of user-submitted comments can significantly impact the interpretation of interactions. High-centrality danmaku comments, for example, can create distinct communities with the user nodes from which they originate, while separating from other nodes within the same cluster. This separation enhances the understanding of danmaku interactions. Thus, the greater the number of communities identified by a community detection algorithm, the more diverse the underlying topics, reflecting a wider range of user behaviors and textual features. In contrast, a smaller number of communities indicate a more concentrated set of topics.

3.2.2. Subgraph Construction

For the analysis of community nodes, the top three communities from each network are selected based on node count, yielding a total of 63 communities. The next step is to construct a bipartite network containing only user and danmaku nodes. The user and danmaku nodes from the original communities are first identified, along with the edges connecting them, while excluding cluster nodes and their associated edges. In this bipartite network, edges are retained between user nodes and danmaku nodes that belong to the same community, with edge weights representing the frequency of users posting the corresponding danmaku comments. Given the relatively small scale of danmaku within communities, the analysis focuses on character-level relationships.

Spelling and grammatical errors, commonly associated with internet language [33], are prevalent in danmaku comments as a form of online communication. For instance, spelling variations such as “禮貌” (transliteration: “Li Mou”, translation: “manners”) and “禮帽” (transliteration: “Li Mou”, the spelling error case of “manners”) frequently occur. To address this, both cosine similarity and Levenshtein distance are employed to establish direct connections between danmaku nodes, supplementing potential omissions in clustering results generated by pre-trained models.

The cosine similarity method leverages TF-IDF (term frequency-inverse document frequency) weighting to represent the danmaku comments, converting text into vectors and calculating cosine similarity between the danmaku comments. If the similarity exceeds 0.5, an undirected, unweighted edge is created between the corresponding nodes.

Additionally, for certain collections where danmaku comments are predominantly in English, a Levenshtein distance-based method is applied. In this approach, an edge is created between nodes if the Levenshtein distance between their content equals 1. Consequently, edges between danmaku nodes are constructed when either of the two conditions is satisfied: cosine similarity greater than 0.5 or Levenshtein distance equal to 1.

To identify the top ten danmaku nodes in each community network, weighted eigenvector centrality is utilized.

For each community, a visualization ready for Gephi was created by selecting the top 10 danmaku nodes based on weighted eigenvector centrality. The node data was filtered and sorted by centrality values to identify these key nodes. A graph was then constructed to include all nodes directly connected to these top nodes by a single edge, forming a target node set. Both edge and node files were filtered to retain only the data relevant to the target nodes, with edge weights preserved from the original community to represent interaction strength. The processed data was exported for visualization in Gephi, providing an intuitive representation of interaction patterns.

To enhance the readability of the figures, we utilized labels to represent the contents of danmaku nodes and provided a detailed description of each label along with its translation in the tables of the appendix. In addition, we draw user nodes in green and danmaku nodes in red below.

4. Results

We categorized English learning programs into two main groups: linguistic knowledge and language skills. The linguistic knowledge category includes phonetics, vocabulary, and grammar, while the language skills category, following established classifications, encompasses listening, speaking, reading, and writing. Each of these seven categories was analyzed using the top three videos from Bilibili, resulting in a total of 21 videos that are denoted as P1~P3 (phonetics), V1~V3 (vocabulary), G1~G3 (grammar), L1~L3 (listening), S1~S3 (speaking), R1~R3 (reading) and W1~W3 (writing). For each video, we constructed edges based on interactions between users and the danmaku comments they submitted. From this data, we identified the three largest communities within each video (e.g., V1_1~V1_3 in V1) and calculated the centrality of nodes within these communities.

Our analysis revealed distinctive patterns in the linguistic knowledge category. For instance, danmaku comments centrality for phonetics videos (e.g., P1) prominently featured content-specific terms like “一個是捲到齒齦後，一個是捲到硬齶” (One curls toward after the alveolar ridge, while the other curls toward the hard palate.), see Figure 3 (All nodes exemplified in the text will be highlighted in **bold** in the corresponding table.). In vocabulary videos (e.g., V2, V3), central danmaku comments often combined English and Chinese explanations, such as “overlook忽視” and “beneath在下方”, see Figure 4. Similarly, grammar videos (e.g., G2, G3) highlighted a series of syntactic discussions of subjunctive mood (“虛擬語氣”) or the functions of word class, such as “狀語修飾動詞”(adverbials modify verbs), see Figure 5. These findings illustrate a clear focus on the respective content within the linguistic knowledge’s category, which we grouped into a distinct class.

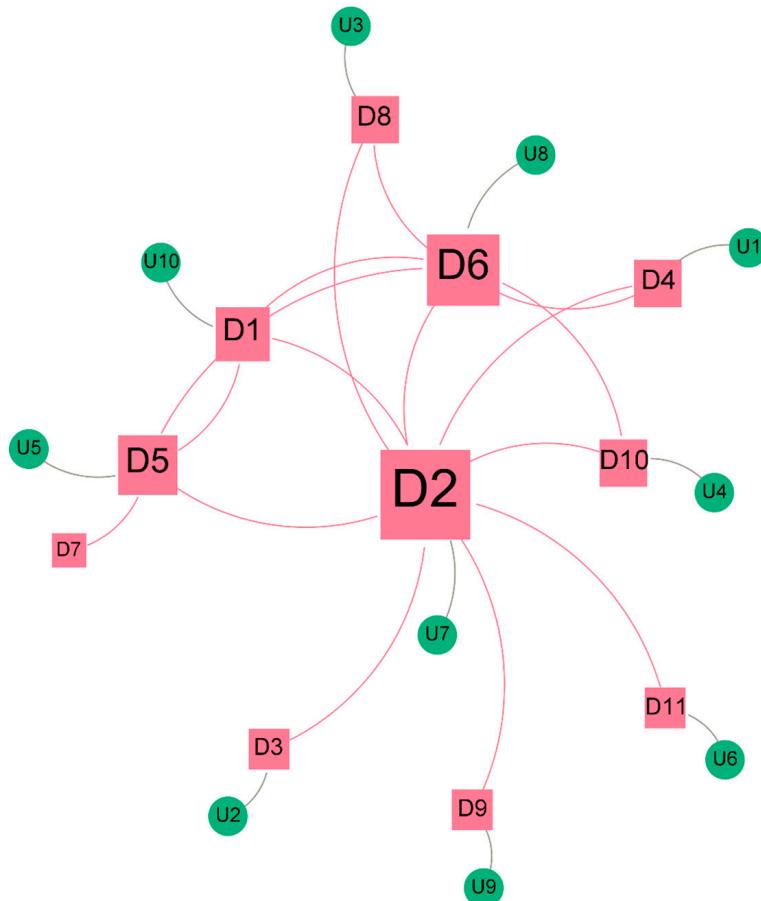


Figure 3. Visualization of social network in P1. It shows that danmaku comments centrality for phonetics videos prominently featured content-specific terms. The detail of danmaku nodes is presented in Appendix B.

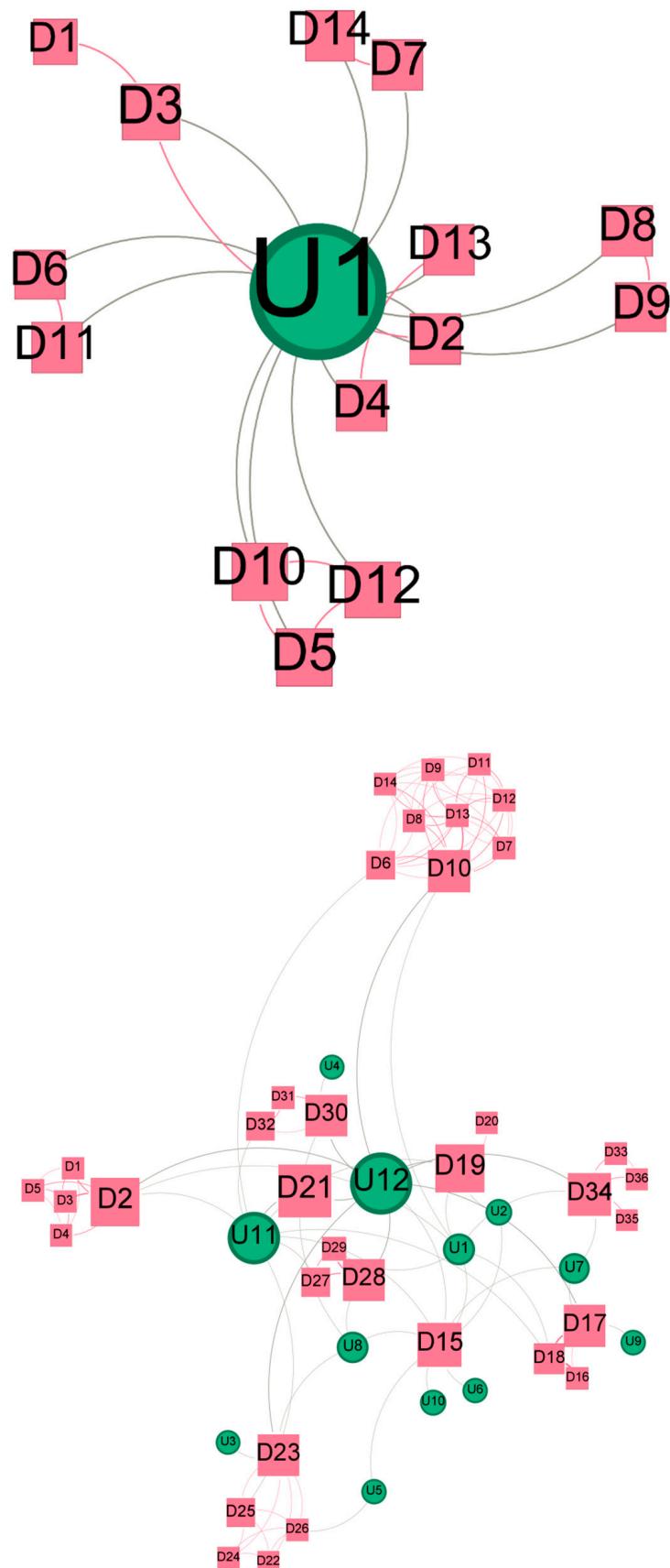


Figure 4. Visualization of social network in V2 (upper panel) and V3 (lower panel). In vocabulary videos, central danmaku comments often combine English and Chinese explanations. The detail of danmaku nodes is presented in Appendixes C and D.

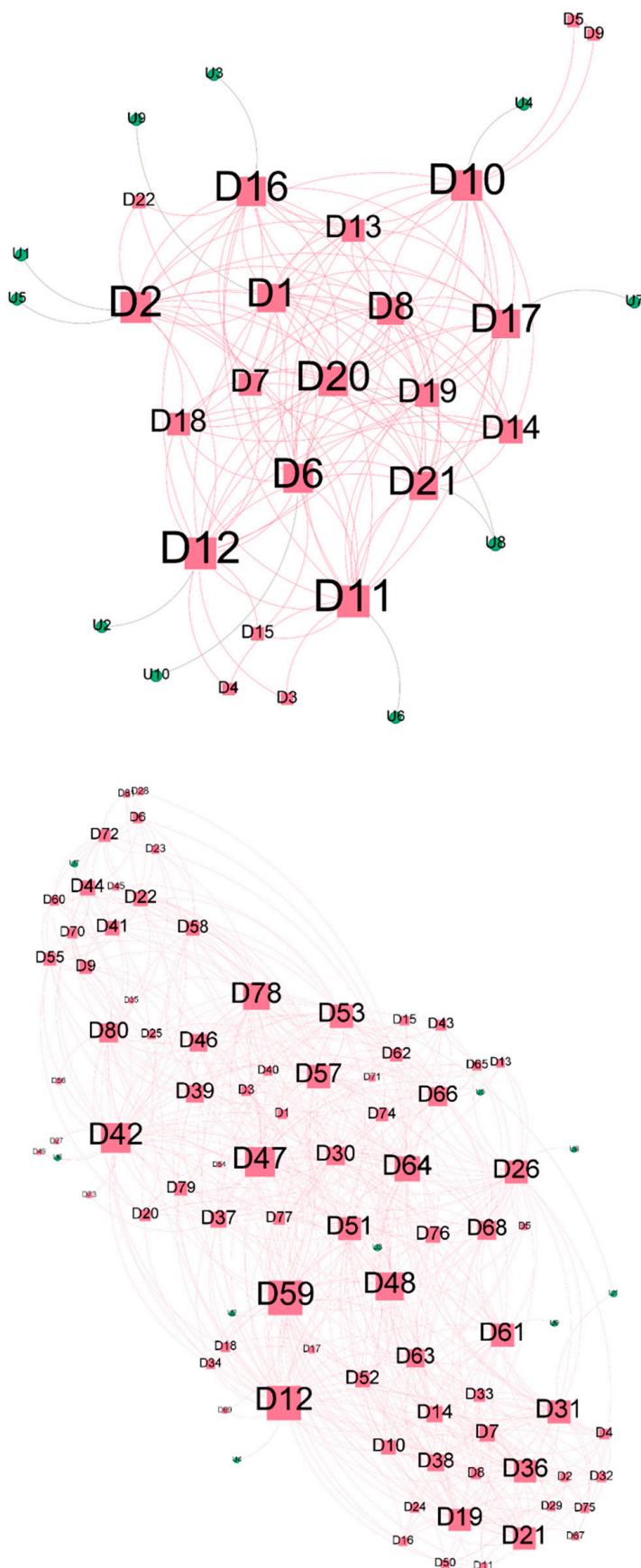


Figure 5. Visualization of social network in G2 (upper panel) and G3 (lower panel). Grammar videos highlighted a series of syntactic discussions of “subjunctive mood” or the functions of word class. The detail of danmaku nodes is presented in Appendixes E and F.

For the language skills category, the analysis reveals nuanced differences in the content and nature of danmaku comments. In L1, although the lack of context makes it difficult to infer precise content, comments from the third community, such that "...沒聽出來..." (...couldn't hear...) clearly indicates their failure in listening comprehension, see Figure 6. This aligns with previous findings that learners seek to share similar experiences of misunderstanding. In S1 and S3, the danmaku comments reflect the integration of newly learned vocabulary during the learning process. Examples include "quantum, 量子" and "mosquito n.蚊子" with consistent annotation of the part of speech. Similar trends are observed in R3 (e.g., "radical極端的") and W3 (e.g., "cultivate, foster培養"), where vocabulary translation is emphasized, see Appendixes H–K. However, these comments show little indication of specific skill-focused training, as the primary focus appears to be on vocabulary explanation rather than the underlying skill itself.

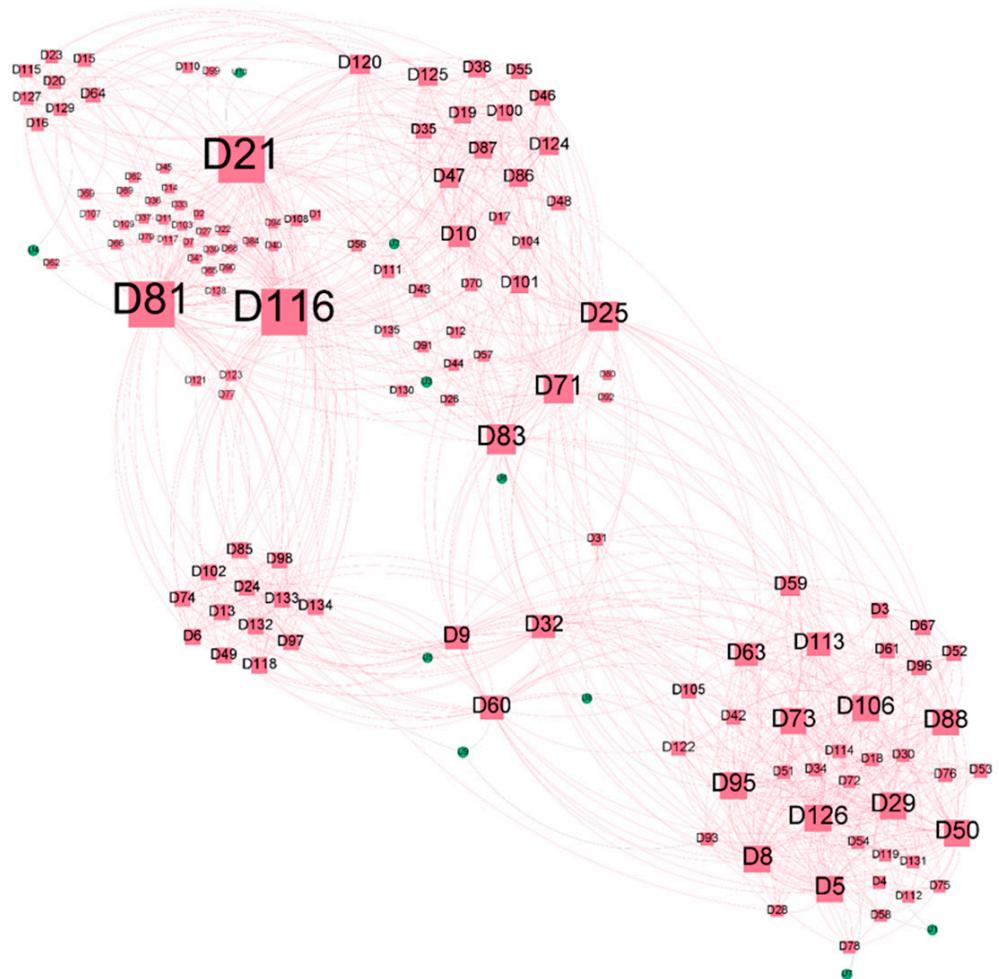


Figure 6. Visualization of social network in L1. It shows that learners seek to share similar experiences of misunderstanding. The detail of danmaku nodes is presented in Appendix F.

Notably, in L2 and across S1 and S3, longer sentences frequently appear in the danmaku comments, suggesting transcription of phrases or sentences introduced during instruction, see Figure 7. This is distinct from the broader trends in the skills category. Among the four skills, listening and speaking stand out as having unique danmaku patterns compared to the others, highlighting their distinctiveness in the learning process.

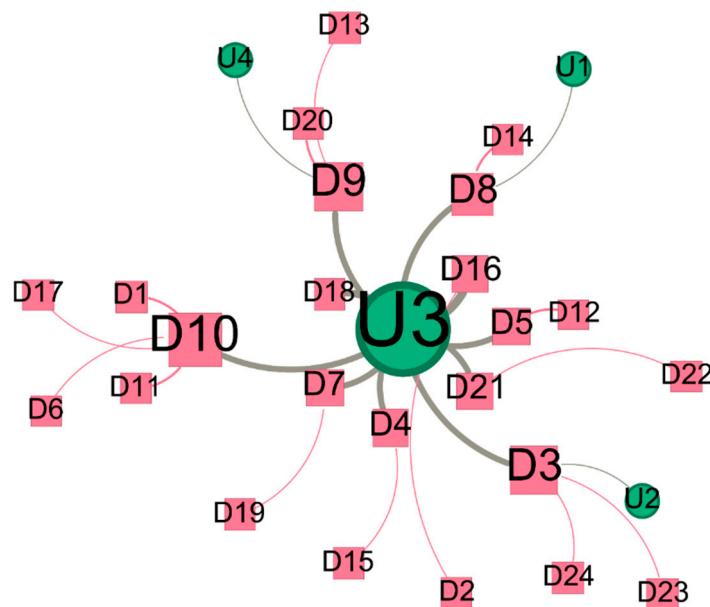


Figure 7. Visualization of social network in L2. Longer sentences frequently appear in the danmaku comments. The detail of danmaku nodes is presented in Appendix J.

5. Discussion and Conclusions

In language learning environments, viewers often use danmaku to discuss linguistic features of the target language and to address comprehension challenges. For instance, Zhang and Cassany (2019) [34] identified three primary categories of danmaku related to Spanish: (1) content focused on Spanish (61%), (2) learning Spanish as a foreign language, and (3) Spanish–Chinese translation. Building on this, our study further explores the role of danmaku in English learning, categorizing its content into two main areas: knowledge and skills.

In the linguistic knowledge category, danmaku comments reflect various aspects of English learning, including phonetics, vocabulary, and grammar. In the language skills category, however, a unique pattern emerges: only *listening* displays distinct danmaku comments, primarily involving learners discussing their own mishearings. In contrast, the other three skills—*speaking*, *reading*, and *writing*—primarily feature vocabulary-focused comments, with danmaku used as a tool for recording and memorizing words. This suggests that vocabulary serves as the foundation for language learning; without a solid vocabulary base, developing other skills is as precarious as building a castle in the air. However, the findings also reveal that danmaku, as employed by English learners, predominantly serves a single purpose—vocabulary-focused learning—with little variation in its application to other aspects of language learning.

According to [35], a study of Chinese undergraduate students in the United States during 2011 and 2012 revealed three key skill profiles: (1) speaking was consistently weaker than the other three skills ($S < L, R, W$); (2) speaking and writing were weaker than listening and reading ($SW < LR$). These findings suggest that reading is relatively easier for Chinese English learners compared to the other skills, while speaking poses the greatest challenge. Building on these insights, it is worth investigating whether the weaker speaking proficiency among danmaku users influences their behavior—specifically, their tendency to transcribe longer sentences in danmaku as a means of reviewing and reinforcing content. While this hypothesis aligns with the observed usage patterns, further research is needed to confirm whether the reliance on transcription is indeed linked to difficulties in mastering speaking.

Jiang et al. (2022) [1] highlighted that Bilibili, due to its interactive danmaku features, provides a more engaging learning environment compared to MOOCs. Building on this, the present study employs an SNA approach to visualize danmaku interactions, offering visual evidence of its effectiveness in facilitating vocabulary acquisition. The findings demonstrate that interactive learning through danmaku fosters a more vibrant and engaging learning atmosphere.

Furthermore, Zeng et al. (2024) [6] discussed how danmaku enhances student participation. Our study complements their work by providing a visualized understanding, thereby operationalizing Zeng et al. (2024) [6]'s assertion that danmaku can improve engagement on online education platforms.

In [36], the researchers employed a variety of methods—including social network analysis, surveys, longitudinal designs, and data visualization—to explore the impact of peer interactions on second language (L2) acquisition during study-abroad experiences. Their findings reveal a strong correlation between the diversity of students' social networks, the frequency of interactions, and improvements in language proficiency, particularly during the initial phase of their study-abroad period. Building on this framework, our study represents a novel approach by employing social network analysis to classify different English instructional videos based on danmaku comments. The findings demonstrate that in the linguistic knowledge category, the danmaku comments closely align with the specific content of the videos, such as phonetics, vocabulary, and grammar. In the skills category, however, distinct patterns emerge only for listening and speaking, where the danmaku exhibits unique characteristics compared to the other skills. This highlights the potential of danmaku as a tool for categorizing instructional content and identifying learner engagement patterns across various aspects of English language learning.

For educators, the feedback provided through danmaku encourages them to not only consider content that students can independently learn online but also design classroom activities that promote discussion and interaction via danmaku. This approach enhances students' ability for autonomous learning while naturally shifting the traditional teacher-centered teaching model to a more student-centered approach, aligning with contemporary trends in education.

For platform designers, optimizing danmaku functionality could involve leveraging network analysis to highlight significant danmaku content and keywords based on collections, videos, or video timelines. For example, social network graphs or leaderboards displayed alongside the video could provide real-time insights into learners' discussion hotspots at specific video progress points. This would benefit content creators by identifying potential learning feedback and challenging concepts, improving teaching materials. It would also help learners identify key learning points and resolve their doubts efficiently.

Moreover, if platforms retain students' danmaku as a form of learning output and feedback, these could serve as valuable data for platform designers when developing diverse English language teaching modules. This collaborative effort between platform designers and educators would facilitate a deeper understanding of students' online learning behaviors and support the continuous refinement of teaching strategies.

Although we provide a visualization of danmaku, the visualization may not be easy to interpret when the graph is extremely dense. When designing an interactive GUI, some filtering and zoom-in/zoom-out functions are required. In addition, more SNA functions, such as centrality and modularity, could be integrated into the next version of our approach.

As discussed, danmaku in online learning fosters an atmosphere of active participation by engaging students with interactive content. This study provides visual evidence demonstrating that danmaku creates a more engaging learning environment compared to

MOOCs. The practical recommendations for educators and platform designers form a key contribution of this research.

However, the integration of an interactive GUI, such as filtering and zoom-in/zoom-out functions, could enhance the immersive quality of SNA visualizations. This would make the visual representations more dynamic and user-friendly. Consequently, the recommendations for educators and platform designers in both teaching and practice would become more actionable and impactful.

Author Contributions: Conceptualization, M.-N.C. and X.H.; methodology, X.H., J.-L.H. and H.-L.T.; software, X.H., J.-L.H. and H.-L.T.; validation, M.-N.C. and X.H.; formal analysis, M.-N.C. and X.H.; investigation, X.H.; resources, X.H.; data curation, X.H.; writing—original draft preparation, M.-N.C. and X.H.; writing—review and editing, J.-L.H. and H.-L.T.; visualization, X.H.; supervision, M.-N.C. and H.-L.T.; project administration, M.-N.C.; funding acquisition, M.-N.C. and H.-L.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Fu Jen Catholic University, Taiwan grant number A0113010.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. The detailed information of Figure 2.

Video	Label	English Translation	Original Content
W3	D1	Just bought it, output template	剛買出模板
W3	D2	Exam at 3 PM, start studying at 10 AM	下午3點考上午10點看
W3	D3	1	1
W3	D4	Just finished the teaching qualification exam yesterday, one hour left before the test, cramming for the CET-6 essay now	昨天剛考完教資，還有一個小時考試，來突擊六級作文
W3	D5	Exam in the afternoon, studying now	下午考試現在看
W3	D6	1000 people	1000人
W3	D7	Passed CET-4, now continuing with CET-6, September 18	四級過啦，繼續來看6級 9.18
W3	D8	Passed CET-6, thank you	六級過了謝謝
W3	D9	Passed CET-6, thank you very much, came back to fulfill my wish	六級過了過了，非常感謝，來還願了
W3	D10	Came back to fulfill my wish, finally passed, scored 139 on the essay	來還願了，終於過了，寫作拿了139
W3	D11	2	2
W3	D12	Beep! Check-in!	滴！打卡！
W3	D13	Taking the CET-6 exam in the afternoon, good luck to me	下午裸考六級，祝自己好運
W3	D14	Confident pass for CET-6	六級穩過
W3	D15	Exam tomorrow	明天考試啦

Table A1. Cont.

Video	Label	English Translation	Original Content
W3	D16	Pass CET-6, please!	六級過過過 拜託了
W3	D17	Back to fulfill my wish	回來還願
W3	D18	Confident pass	穩過
W3	D19	Back to fulfill my wish, passed CET-6, happy, thank you uploader	回來還願, 六級過了, 開心, 感謝博主
W3	D20	Passed CET-6, came back to fulfill my wish, and gave the uploader a coin	六級過啦, 前來還願給up主投幣
W3	D21	Wish to pass CET-6!	許願六級過!
W3	D22	Passed CET-6 in December 2022! Came back to fulfill my wish! Thank you so much!	2022年12月考的六級過了! 來還願! 謝謝瑞斯拜!
W3	D23	22 12 6	22 12 6
W3	D24	Came back to fulfill my wish, thank you uploader	來還願, 感謝up主
W3	D25	Here I am	來啦
W3	D26	If I pass, I'll recharge for you	過了我就給瑞充電
W3	D27	Template	mu ban
W3	D28	CET-6 in the afternoon	下午六級
W3	D29	Passed CET-6, came back to fulfill my wish!	六級過了, 來還願了!
W3	D30	One-shot pass for CET-6, please	六級一把過求求了
W3	D31	Starting check-in	開始打卡
W3	D32	Check-in completed for 1 day~	已完成打卡1天~
W3	D33
W3	D34	I'm back again! It's time for CET-6! Cramming at the last minute!	我又來了! 該考六級了! 我來臨時抱佛腳了!
W3	D35	Looking for a study partner	cpdd
W3	D36	Passed CET-6, came back to fulfill my wish, thank you so much	六級已過, 前來還願, 感謝阿瑞
W3	D37	One-shot pass in the afternoon	下午一遍過
W3	D38	This is a template to help you write when you're stuck	這是模板, 讓你找不到話的時候寫
W3	D39	600+	600+
W3	D40	Came back to fulfill my wish, passed CET-4	來還願了, 四級過了
W3	D41	Passed! My essay score improved by over 30 points, thank you, teacher!	過啦過啦, 作文提高了三十多分, 謝謝老師!

Appendix B

Table A2. The detailed information of Figure 3.

Community	Label	English Translation	Original Content
P1_3	D1	One is “a~”, and the other is “ai” (the character for “sorrow”).	一個是a~, 一個是悲哀的哀
P1_3	D2	One rolls back to the gums, the other rolls to the hard palate.	一個是捲到齒齦後, 一個是捲到硬齶
P1_3	D3	I got it! Try adding a small “ri” sound to the previous one.	我會啦, 你在前一個的基礎上加一個小的“日”的音試試

Table A2. Cont.

Community	Label	English Translation	Original Content
P1_3	D4	Are the two pronunciations different in that one is longer and the other shorter?	是不是倆者一個是發音較長, 一個較短?
P1_3	D5	One is a longer “u” sound, and the other is a light “oh” sound.	一個是長一點的u的音, 一個是哦輕聲的音
P1_3	D6	Is one “um” and the other “oh”?	是不是一個唔, 一個歐
P1_3	D7	The first one is a light “wo”, and the second is a light “wu”.	前一個是wo輕聲, 後一個是wu輕聲
P1_3	D8	One is “emmm”, and the other is a closed-mouth “hmm”.	一個是emmm, 一個是不開口的嗯
P1_3	D9	It feels like one is “ch” and the other is “chu”, moving into a full pinyin sound.	感覺像是一個是ch, 一個是chu, 往後拼音
P1_3	D10	One ends with tightly closed lips, and the other leaves the tongue tip on the upper gums.	一個雙脣閉緊結束, 一個舌尖停留上齒齦
P1_3	D11	One does not add a trill (voiced), and the other does.	一個不加顫音(濁音), 一個加

Appendix C

Table A3. The detailed information of upper panel in Figure 4.

Community	Label	English Translation	Original Content
V2_2	D1	precious means “valuable or cherished”.	precious 寶貴的
V2_2	D2	spacious means “having a lot of space”; broad means “wide or extensive”.	spacious 寬敞的; broad 廣闊的
V2_2	D3	spacious means “wide or open”; broad means “extensive”; precious means “valuable”.	spacious, broad 廣闊的; precious 寶貴的
V2_2	D4	undermine means “to weaken or destroy”.	undermine 破壞
V2_2	D5	overlook means “to ignore or fail to notice”.	overlook 忽視
V2_2	D6	special means “unique or distinct”; species means “a group of organisms with shared traits”.	special 特殊的; species 物種
V2_2	D7	proper means “suitable or appropriate”; property means “assets or possessions”; asset means “a valuable resource or property”.	proper 適合的; property 財產; asset 資產(不動產)
V2_2	D8	summit means “the highest point or peak”.	summit 頂點
V2_2	D9	summit means “a high point”; peak means “the top or apex”.	summit 高峯; peak 頂點
V2_2	D10	neglect means “to ignore”; overlook means “to fail to notice or look over”.	neglect 忽略; overlook 忽視, 俯瞰
V2_2	D11	species means “a group of organisms with similar traits”.	species 物種
V2_2	D12	neglect means “to ignore or pay no attention to”; overlook means “to fail to notice”.	neglect 忽視 overlook
V2_2	D13	undermine means “to weaken or damage”.	undermine 損壞
V2_2	D14	proper means “appropriate or suitable”; property means “assets or possessions”.	proper 合適的; property 資產

Appendix D

Table A4. The detailed information of lower panel in Figure 4.

Community	Label	English Translation	Original Content
V3_1	D1	annual every year	annual 每年
V3_1	D2	annual every year	annual 每年的
V3_1	D3	annual every year, once a year	annual 每年的一年一次的
V3_1	D4	annual every year, one year	annual 每年的一年的
V3_1	D5	annual every year, once a year	annual 每年的的一年一次的
V3_1	D6	beneath under	beneath 在...下方
V3_1	D7	beneath under, lower than, inferior to, under cover of, beneath one's dignity, below, underneath	beneath 在...下方 低於 次於 在...掩蓋下 有失...的身份 在下方 在底下
V3_1	D8	beneath under, lower than, inferior to, under cover of, beneath one's dignity, below, underneath	beneath 在...的下方 低於 次於 在...的掩蓋下 有失...的身份 在下方 在底下
V3_1	D9	beneath under	beneath 在...下方
V3_1	D10	beneath under	beneath 在下方
V3_1	D11	beneath under, socially inferior	beneath 在下方、地位低於
V3_1	D12	beneath under, socially inferior	beneath 在下方 地位低於
V3_1	D13	beneath under, socially inferior, lower than	beneath 在下方 地位低於 次於
V3_1	D14	beneath under, lower than	beneath 在下方 地位次於
V3_1	D15	camel	camel 駱駝
V3_1	D16	county	county 縣
V3_1	D17	county district	county 郡
V3_1	D18	county district, county	county 郡縣
V3_1	D19	gene	gene 基因
V3_1	D20	gene	geng 基因
V3_1	D21	gram	gram 克
V3_1	D22	instinct intuition	instinct. 本能, 直覺
V3_1	D23	instinct	instinct 本能
V3_1	D24	instinct intuition	instinct 本能直覺
V3_1	D25	instinct intuition, nature	instinct 本能 直覺 本性
V3_1	D26	instinct intuition, inherent nature	instinct 直覺 本能生性
V3_1	D27	legal law	legal 法律
V3_1	D28	legal lawful	legal 法律的
V3_1	D29	legal lawful, legitimate	legal 法律的 合法的
V3_1	D30	legislation law	legislation 法律
V3_1	D31	legislation laws, regulations	legislation 法律法規
V3_1	D32	legislation laws, regulations, legislation	legislation 法律法規立法
V3_1	D33	resident resident doctor, inhabitant, settler	resident 住院的 居民 定居者
V3_1	D34	resident resident	resident 居民
V3_1	D35	resident resident, residence	resident 居民 residence居住
V3_1	D36	resident resident, resident doctor, settled	resident 居民住院醫生定居的

Appendix E

Table A5. The detailed information of upper panel in Figure 5.

Community	Label	English Translation	Original Content
G2_2	D1	Can this be understood as a subjunctive mood?	可以理解為虛擬語氣嗎
G2_2	D2	Isn't this the subjunctive mood?	這不是虛擬語氣嗎
G2_2	D3	Here the teacher is saying that in this case, Zhang San is complaining whether this sentence can be expressed in this tense, and then the inability to do so leads to the reason why the subjunctive mood is used. What was previously discussed is the subjunctive mood for various tenses, and here it talks about the past tense, which is one type of subjunctive mood.	這裏老師是在說,在他說的這個案件裏,張三抱怨這句話能不能用這個時態表示,然後不能進而引出為什麼要用虛擬語氣,之前說的是各種時態時候的虛擬語氣,這裏說的是過去時,也就是虛擬語氣的分類之一
G2_2	D4	It is contrary to past facts, so it is the subjunctive mood.	與過去的事實相反,所以是虛擬語氣
G2_2	D5	Like a shortened clause.	和縮寫版從句一樣
G2_2	D6	The subjunctive mood is just a way of speaking casually.	虛擬語氣就是口嗨的意思
G2_2	D7	This is the impossible completion subjunctive mood.	這裏就用不可能完成的虛擬語氣
G2_2	D8	This is the conditional mood in the subjunctive mood.	是虛擬語氣中條件語氣
G2_2	D9	I thought they were the same, both are clauses.	我以為都是一樣的,都是從句
G2_2	D10	Just like the subjunctive mood.	和虛擬語氣一樣
G2_2	D11	The subjunctive mood is always fake.	虛擬語氣都是假的
G2_2	D12	Subjunctive mood?	虛擬語氣?
G2_2	D13	Note: The subjunctive mood is included in conditional adverbial clauses, but not all conditional adverbial clauses are in the subjunctive mood.	注意:虛擬語氣被包含於條件副詞從句,條件副詞從句不都是虛擬語氣
G2_2	D14	The one in front is not the subjunctive mood.	前面的這個不是虛擬語氣
G2_2	D15	The note means that "will" and "shall" can also be used as auxiliary verbs for the subjunctive mood.	註釋的意思是will和shall也可做虛擬語氣的助動詞
G2_2	D16	This is not the subjunctive mood!	這裏不是虛擬語氣!
G2_2	D17	It might be the subjunctive mood.	有可能是虛擬語氣
G2_2	D18	Isn't this still the subjunctive mood?	這不還是虛擬語氣嗎?
G2_2	D19	Is this the subjunctive mood?	這個是不是虛擬語氣?
G2_2	D20	This is not the subjunctive mood.	不是虛擬語氣
G2_2	D21	This is still the subjunctive mood.	這裏還是虛擬語氣啊
G2_2	D22	This assumption is not a subjunctive mood that is contrary to reality!	這裏的假設不是與現實相反的虛擬語氣!

Appendix F

Table A6. The detailed information of lower panel in Figure 5.

Community	Label	English Translation	Original Content
G3_1	D1	Adverbial modifier of verb, brother is a noun	狀語修飾動詞, brother是名詞
G3_1	D2	Are you still modifying a noun when you are already a top scorer? It is an adverbial, adverbs cannot modify nouns	都是狀元了你還能修飾名詞嗎它是副詞詞性,副詞不能修飾名詞呀
G3_1	D3	Adverbial modifies “verb”, not “predicate verb”, here meeting is a verb	狀語是修飾“動詞”的, 而不是“謂語動詞”,這裏meeting是動詞.
G3_1	D4	Time adverbs can modify nouns	時間副詞可以修飾名詞
G3_1	D5	Adverbial modifies a verb, here under the tree modifies the noun the boy. Attributive is adjective-based	狀語修飾的是動詞, 這裏under the tree是對名詞the boy的修飾. 定語是形容詞性的
G3_1	D6	It is a location adverbial clause	是地點狀語從句把
G3_1	D7	Nouns can modify nouns!	名詞可以修飾名詞!
G3_1	D8	Novel is a noun, relative clause modifies noun	novel名詞, 定語從句修飾名詞
G3_1	D9	Isn't this an adverbial?	這不是狀語了麼
G3_1	D10	Attributive modifies noun, where does this modification of subject come from?	定語修飾名詞 哪來的修飾主語
G3_1	D11	Attributive is used to modify limits	定語用於修飾限定
G3_1	D12	Attributive modifies nouns, adverbial modifies verbs?	定語修飾名詞, 狀語修飾動詞?
G3_1	D13	Preposition modifies adjective verb	介詞修飾 形容詞 動詞
G3_1	D14	Attributive: modifies nouns and pronouns	定語:修飾名詞和代詞的
G3_1	D15	Adverb is to modify a verb, adverb, word in a sentence	副詞就是修飾動詞 副詞 句子的詞
G3_1	D16	Attributive is not for subject modification, it modifies nouns	定語可不是專門修飾主語的哈 修飾名詞的
G3_1	D17	To modify a verb, it must be an adverb; if it's an adjective, it can only modify a noun	修飾動詞要是副詞, in order是形容詞的話就只能修飾名詞啦
G3_1	D18	Incorrect, adverbial does not modify noun so it cannot modify indefinite pronouns	記錯了, 狀語不修飾名詞所以不能修飾不定代詞
G3_1	D19	Attributive modifies noun, what are you listening to?	定語是修飾名詞的 聽啥呢
G3_1	D20	Classmate, adverbial doesn't modify noun	同學, 狀語不是修飾名詞的
G3_1	D21	Attributive modifies noun?	定語修飾名詞是嘛?
G3_1	D22	I mixed up with adverbial	我跟狀語混了
G3_1	D23	Can this not be used as a location adverbial to modify the predicate?	這個就不能做地點狀語修飾謂語嗎
G3_1	D24	The subsequent attributive modifies the noun baby	後面的定語修飾的名詞baby
G3_1	D25	The adverbial clause modifies the predicate	修飾主句謂語的是狀語從句
G3_1	D26	Because attributive modifies nouns, adverbial modifies verbs, describes those things	因為定語是修飾名詞, 狀語是修飾動詞形容那些的
G3_1	D27	How can “how” be an adverbial, modifying “did”?	how做狀語啊 修飾did啊
G3_1	D28	Location and time are generally adverbials	地點和時間一般都是做狀語

Table A6. Cont.

Community	Label	English Translation	Original Content
G3_1	D29	This is how I understand it, attributive modifies noun, as pre-attributive before the noun and post-attributive after the noun, so those closely following the noun are most likely attributives	我是這樣理解的, 定語修飾名詞, 在名詞前面為前置定語, 在名詞後面為後置定語, 所以緊跟名詞的概率為定語
G3_1	D30	It modifies the verb	修飾動詞的吧
G3_1	D31	Attributive modifies nouns	定語修飾名詞
G3_1	D32	Nouns can act as attributives to modify another noun	名詞可以作定語, 修飾另一個名詞
G3_1	D33	Attributive actually refers to a word modifying nouns	定語其實就是修飾名詞的詞
G3_1	D34	Pronouns are nouns, adverbial can modify nouns, so why can't adverbial modify indefinite pronouns?	代詞屬名詞, 狀語可修飾名詞, 那為什麼狀語不能修飾不定代詞?
G3_1	D35	Not a predicate verb	不是謂語的動詞
G3_1	D36	Attributive modifies nouns	定語是修飾名詞的
G3_1	D37	Adverbial doesn't modify verbs, why can it modify the subject?	狀語不是修飾v嗎, 為什麼可以修飾主語[n]
G3_1	D38	Attributive modifies the preceding noun	定語修飾前面的名詞
G3_1	D39	Isn't adverbial clause modifying verbs?	狀語從句不是修飾與動詞的嗎
G3_1	D40	Adverbial also modifies verbs	狀語同樣也是修飾動詞
G3_1	D41	It's adverbial	是狀語吧
G3_1	D42	Modifies adverbial?	修飾狀語?
G3_1	D43	Adverbial modifies verbs, adjectives, adverbs, or the entire sentence	狀語修飾動詞, 形容詞, 副詞或整個句子
G3_1	D44	Look at adverbial	狀語去看
G3_1	D45	Adverbial's part of speech is adverb, and it doesn't describe nouns	狀語的詞性是副詞, 並且狀語不形容名詞
G3_1	D46	Adverbial clause can't only modify verbs	狀語從句不是隻能修飾動詞的嗎
G3_1	D47	Adverbial modifies verbs here, so it's not adverbial	狀語修飾動詞吧 這裏修飾名詞所以不是狀語
G3_1	D48	Because adverbial doesn't modify nouns	因為狀語不修飾名詞
G3_1	D49	Adverbial modifies sentence or verb	狀語修飾句子或動詞
G3_1	D50	The preceding one is attributive, it modifies the noun, helping is a verb	前面那個, 定語是修飾名詞的, 幫助是動詞
G3_1	D51	Adverbial modifies action, attributive modifies noun	狀語修飾動作啊, 定語修飾名詞啊
G3_1	D52	It can't be adverbial because it modifies a noun, adverbial can't modify nouns	不能作狀語是因為它修飾的是名詞, 狀語不能修飾名詞
G3_1	D53	Adverbial modifies verbs	狀語修飾動詞
G3_1	D54	Because it's a prepositional phrase, it's an adverb's part of speech and cannot modify nouns, so it's adverbial	因為那個是個介詞短語所以是副詞的詞性不能修飾名詞所以是狀語
G3_1	D55	Isn't adverbial?	副詞不是狀語嗎
G3_1	D56	Adverbial usually modifies the predicate, here it should modify began	狀語一般修飾謂語, 在這裏我覺得應該是修飾began.

Table A6. Cont.

Community	Label	English Translation	Original Content
G3_1	D57	Adverbial modifies verbs	狀語修飾動詞啊
G3_1	D58	Adverbial—location adverbial—modifies verb before it	狀語-地點狀語-提到動詞前面修飾動詞
G3_1	D59	Adverbial modifies verb, attributive modifies noun	狀語是修飾動詞的, 定語是修飾名詞的
G3_1	D60	Complement is not adverbial	賓補不是狀語
G3_1	D61	Adverbial doesn't modify nouns, nouns can be modified by attributives	狀語不用來修飾名詞, 名詞可以被定語修飾
G3_1	D62	It modifies verbs	就是修飾動詞
G3_1	D63	Complement can only modify the object, the object modified is definite, attributive modifies nouns which are general	賓補只能修飾賓語 修飾的賓語是確定的 定語是修飾名詞的 修飾的名詞是籠統的.
G3_1	D64	Attributive modifies nouns, adverbial modifies verbs, time, location, etc.	定語修飾名詞, 狀語修飾動詞時間地點啥的
G3_1	D65	Isn't modifier an adjective?	不是修飾詞纔是形容詞嗎
G3_1	D66	Adverbial can't modify nouns, it only modifies verbs, adjectives, adverbs, and the entire sentence, mainly modifying verbs	狀語不能修飾名詞, 他只修飾動詞 形容詞 副詞和整個句子, 實際主要是修飾動詞
G3_1	D67	Attributive modifies nouns, bored is an adjective	定語修飾名詞 bored是形容詞
G3_1	D68	Attributive modifies nouns, adverbial modifies verbs, adjectives, and adverbs	定語修飾名詞, 狀語修飾動詞形容詞副詞
G3_1	D69	Meaning doesn't change if it's attributive; otherwise, it's adverbial in modifying the noun or verb	意思不變則是定語, 反之是狀語 在修飾名詞前面的動詞
G3_1	D70	Sentence acts as adverbial, adverbial clause	句子做狀語, 狀語從句,
G3_1	D71	Sentence acts as (time) adverbial, modifies verb stopped	句子做(時間)狀語, 修飾動詞stopped
G3_1	D72	Adverbial isn't "de", it's "di", and when it's a location adverbial, "di" is omitted	狀語不是"的", 是"地", 且做地點狀語, "地"就省略了喔
G3_1	D73	What does adverbial modify in subject-predicate-complement structure?	狀語在主系表中修飾什麼?
G3_1	D74	Time adverbial modifies the entire sentence, attributive modifies a word	時間狀語修飾整個句子, 定語修飾一個詞
G3_1	D75	It can't be attributive because it modifies a noun, so it doesn't fit	應該不能作定語吧, 定語修飾名詞, 也就是得修飾a book, 意思也不對啊
G3_1	D76	Attributive modifies noun, here under the tree modifies reading verb, so it's adverbial	定語是修飾名詞的, 在樹下修飾的是reading動詞, 所以是狀語
G3_1	D77	Emphasis sentence can only modify subject, object; adverbial can't modify verbs	強調句只能修飾主語賓語狀語它不能修飾動詞
G3_1	D78	Adverbial doesn't modify verbs	狀語不是修飾動詞
G3_1	D79	1/3 is not a modifier?	1/3不是修飾詞嗎?
G3_1	D80	Adverbial isn't for modifying predicate?	狀語不是修飾謂語嗎
G3_1	D81	Time and location adverbials can also emphasize	時間和地點狀語也能強調吧.

Appendix G

Table A7. The detailed information in Figure 6.

Community	Label	English Translation	Original Content
L1_2	D1	to live couldn't be heard	to live 沒聽出來
L1_2	D2	determine couldn't be heard +1	determine 沒聽出來+1
L1_2	D3	and was heard as the	and 聽成 the
L1_2	D4	heard at the as that	把 at the 聽成了 that
L1_2	D5	a was heard as the	a 聽成 the
L1_2	D6	could hear but couldn't write	能聽出來但寫不出來
L1_2	D7	but couldn't hear show	倒是 show 沒聽出來
L1_2	D8	a was heard as the	a 都聽成 the
L1_2	D9	couldn't hear the	聽不出來 the啊
L1_2	D10	couldn't hear to	沒聽出來 to
L1_2	D11	opens couldn't hear the s	opens 沒聽出來 s
L1_2	D12	when couldn't be heard	when 沒聽出來
L1_2	D13	couldn't hear anything, crying	啥都沒聽出來, 哭了
L1_2	D14	mother couldn't be heard	mother 居然沒聽出來
L1_2	D15	that was couldn't be heard at all, wow	that was 完全沒聽出來, 哇靠
L1_2	D16	I have learned couldn't be heard at all	i have learned 完全沒聽出來
L1_2	D17	the first sentence couldn't be heard	第一句沒聽出來
L1_2	D18	so I heard the	所以我聽成 the
L1_2	D19	couldn't hear in	沒聽出來 in
L1_2	D20	or couldn't be heard at all	or 完全沒聽出來
L1_2	D21	I couldn't hear	i 沒聽出來
L1_2	D22	applauded couldn't be heard	applauded 沒聽出來
L1_2	D23	this sentence couldn't be heard at all	這句完全沒聽出來
L1_2	D24	I couldn't hear anything	我都沒有聽出來
L1_2	D25	couldn't hear the	沒聽出來 the
L1_2	D26	under couldn't be heard	under 沒聽出來
L1_2	D27	couldn't hear it was deceased	沒聽出來是去世
L1_2	D28	their was heard as the...	their 聽成 the..
L1_2	D29	the was heard as a	the 又聽成了a
L1_2	D30	the was heard as then	the 聽成then
L1_2	D31	this the just couldn't be heard	這個 the 就聽不出來
L1_2	D32	the couldn't be heard again	the 又沒聽出來
L1_2	D33	I give up, couldn't hear it	我認輸了, 沒聽出來
L1_2	D34	laughing, the space aboard was heard as the baseboard...	笑死, the space aboard 聽成 the baseboard.....
L1_2	D35	couldn't hear out	沒聽出來out
L1_2	D36	late couldn't be heard...	late 沒聽出來...
L1_2	D37	process couldn't be heard	process 沒聽出來
L1_2	D38	I have couldn't be heard	i have 沒聽出來

Table A7. Cont.

Community	Label	English Translation	Original Content
L1_2	D39	lofty couldn't be heard	lofty 沒聽出來
L1_2	D40	only the place name couldn't be heard	只有地名沒聽出來
L1_2	D41	sustain couldn't be heard	sustain 沒聽出來,
L1_2	D42	wrote it as the	寫成 the 了
L1_2	D43	it is couldn't be heard	it is 沒聽出來
L1_2	D44	aboard couldn't be heard	aboard 沒聽出來
L1_2	D45	at least couldn't be heard	at least 沒聽出來
L1_2	D46	with couldn't be heard	with 沒聽出來
L1_2	D47	it couldn't be heard	it 沒聽出來
L1_2	D48	day couldn't be heard	day 沒聽出來
L1_2	D49	couldn't hear Thor	沒把錘哥聽出來
L1_2	D50	heard a as the	把 a 聽成 the
L1_2	D51	heard the day	聽成 the day
L1_2	D52	to was heard as the again	to 又聽成 the 了
L1_2	D53	I also heard on the	我也聽成 on the
L1_2	D54	them was heard as the	them 聽成了 the
L1_2	D55	with couldn't be heard	with 沒聽出來
L1_2	D56	dumped it couldn't be heard	dumped it 沒聽出來
L1_2	D57	hit couldn't be heard	hit 沒聽出來
L1_2	D58	always hear a as the	總是把a聽成 the
L1_2	D59	heard a as the, but couldn't hear the again, incredible	a 聽成了the. 後面的 the 又沒聽出來, 服了自己
L1_2	D60	I really couldn't hear the	我真聽不出來 the
L1_2	D61	to was heard as the	to 聽成 the
L1_2	D62	couldn't hear wrote	沒聽出來 wrote
L1_2	D63	I wrote the	我寫了 the
L1_2	D64	this and couldn't be heard at all	這個 and 完全沒聽出來
L1_2	D65	dies couldn't be heard...	dies 沒聽出來...
L1_2	D66	isn't couldn't be heard	isnt 沒聽出來
L1_2	D67	always hear to as the	總把 to 聽成 the
L1_2	D68	i guess couldn't be heard	i guess 沒聽出來
L1_2	D69	this fear couldn't be heard	這個 fear 沒聽出來
L1_2	D70	re couldn't be heard	re 沒聽出來
L1_2	D71	the couldn't be heard	the 沒聽出來
L1_2	D72	heard as the one	聽成了 the one
L1_2	D73	the was heard as a	the 聽成 a
L1_2	D74	oh my, couldn't hear it	我的天 聽不出來
L1_2	D75	us a was heard as the	us a 聽成了 the
L1_2	D76	heard on the	聽成 on the
L1_2	D77	couldn't hear their	竟然沒聽出來 their
L1_2	D78	a is always heard as the, the is always heard as a	a永遠聽成 the, the 永遠聽成 a

Table A7. Cont.

Community	Label	English Translation	Original Content
L1_2	D79	secret couldn't be heard	secret 沒聽出來
L1_2	D80	the amount of couldn't be heard	the amount of 沒聽出來
L1_2	D81	couldn't hear i	沒聽出來 i
L1_2	D82	ever couldn't be heard, heard as let (◎·一·◎)	ever 沒聽出來, 聽成了 let (◎·一·◎)
L1_2	D83	couldn't hear a and missed the	沒聽出來a 漏了 the
L1_2	D84	usually really couldn't be heard	usually 真的沒聽出來
L1_2	D85	couldn't be heard either	也沒聽出來
L1_2	D86	and couldn't be heard	and 沒聽出來
L1_2	D87	did not hear "is"	is沒聽出來
L1_2	D88	heard "a" as "the"	a 聽成了 the
L1_2	D89	could not hear "ever", very difficult	沒聽出來 ever 好難
L1_2	D90	could not hear "honest"	honest 沒聽出來
L1_2	D91	could not hear "would"	would 沒聽出來
L1_2	D92	could not hear "pass the prime"	pass the prime 沒聽出來
L1_2	D93	heard the first word as "the"	第一個詞聽成了 the
L1_2	D94	could not hear "how about"	how about 沒聽出來
L1_2	D95	heard "a" as "the"	a 聽成the
L1_2	D96	every time I hear "the" as "a" and "a" as "the"!	前面的! 我每次都是 the 聽成 a a 聽成 the!
L1_2	D97	heard it here	這裏聽出來了
L1_2	D98	really could not hear "i"	真沒聽出來i
L1_2	D99	could not hear "order", missed "that" at the end	order 後面 that 沒聽出來
L1_2	D100	could not hear "there's"	there's 沒聽出來
L1_2	D101	could not hear "and the"	and the 沒聽出來
L1_2	D102	could not hear the part at the end	後面沒聽出來
L1_2	D103	could not hear "series"	series 沒聽出來
L1_2	D104	could not hear "every"	沒聽出來 every
L1_2	D105	could not hear "and the"	and the 我也沒聽到
L1_2	D106	heard it as "the"	聽成 the 了
L1_2	D107	could not hear "fear of"	fear of 沒聽出來
L1_2	D108	could not hear "how to", so frustrating	how to 沒聽出來 氣
L1_2	D109	could not hear the second "ve"	第二個"ve 沒聽出來
L1_2	D110	could not hear "could" before "that"	that 前 could.沒聽出來
L1_2	D111	really could not hear "it"	真的沒聽出來 it
L1_2	D112	heard "an" as "the"	把 an 聽成 the 了
L1_2	D113	heard "a" as "the"	我把 a 聽是 the
L1_2	D114	I also wrote "the" and couldn't understand why there was a "the"	我也寫了 the 還想不通為什麼會有個 the
L1_2	D115	could not hear the second half of the sentence	後半句完全沒聽出來
L1_2	D116	could not hear "a"	沒聽出來 a
L1_2	D117	could not hear "bound"	bound 沒聽出來

Table A7. Cont.

Community	Label	English Translation	Original Content
L1_2	D118	could not hear 0.5	0.5 我都聽不出來
L1_2	D119	always hear “a” as “the”	老是吧 a 聽成 the
L1_2	D120	completely could not hear it	完全沒聽出來
L1_2	D121	could not hear “figure”	figure 沒聽出來
L1_2	D122	could not hear “the what” with “the”	the what 的 the 聽不出
L1_2	D123	really could not hear it	really 竟然沒聽出來
L1_2	D124	could not hear “and”	and 沒聽出來
L1_2	D125	could not hear “that”	that 沒聽出來
L1_2	D126	heard “the” as “a”	the 聽成了a
L1_2	D127	could not hear “been”	been 完全沒聽出來
L1_2	D128	could not hear “big”, don’t understand the meaning	big 沒聽出來, 不懂啥意思
L1_2	D129	completely could not hear the last sentence	最後一句完全沒聽出來
L1_2	D130	could not hear “into”	into 沒聽出來
L1_2	D131	heard it as “open the”	我聽成 open the
L1_2	D132	could not hear anything...	啥也聽不出來...
L1_2	D133	could not hear the part at the end	後面聽不出來
L1_2	D134	was wondering why I couldn’t hear it	我說咋聽不出來
L1_2	D135	could not hear the place name	地名沒聽出來

Appendix H. Visualization of Social Network and the Detailed Information in S1

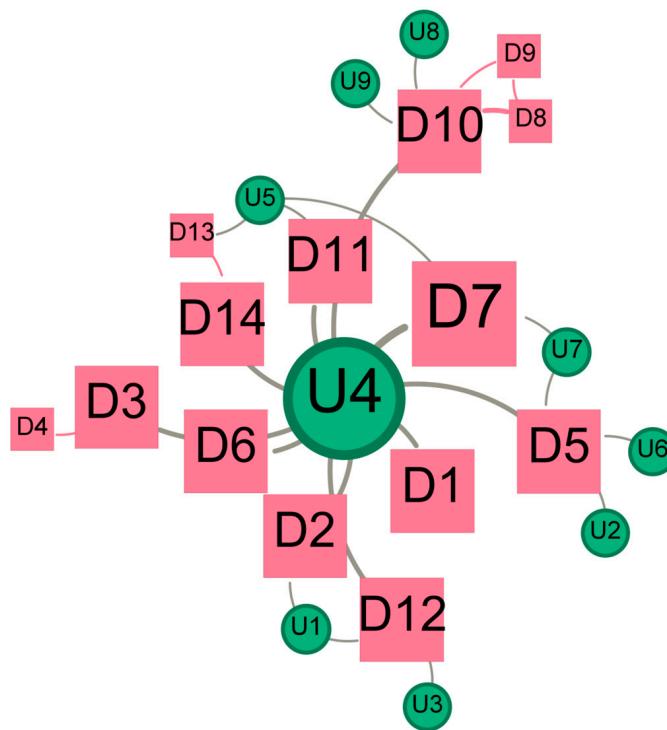
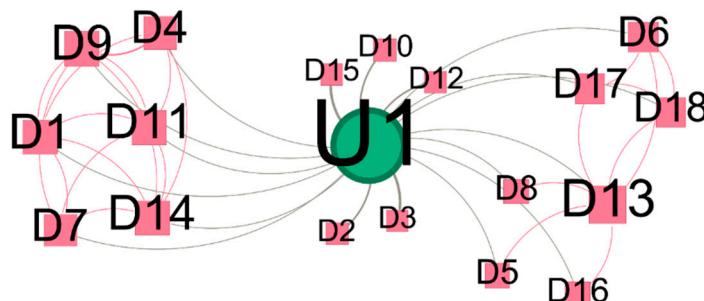
**Figure A1.** Visualization of social network in S1: vocabulary-focused comments.

Table A8. The detailed information in Figure 1.

Community	Label	English Translation	Original Content
S1_2	D1	bear endure; tolerate;	bear 承受; 忍受;
S1_2	D2	botulism botulinum poisoning	botulism 肉毒中毒
S1_2	D3	collaborate cooperate	collaborate 合作
S1_2	D4	collaborate cooperate, collaborate	collaborate 合作 協作
S1_2	D5	glucose	glucose 葡萄糖
S1_2	D6	likelihood probability; possibility;	likelihood 可能; 可能性;
S1_2	D7	molecule	molecule 分子
S1_2	D8	notorious infamous	notorious 臭名昭著
S1_2	D9	notorious infamous	notorious 臭名昭著
S1_2	D10	notorious infamous	notorious 臭名昭著的
S1_2	D11	quantum	quantum 量子
S1_2	D12	respiration breathing	respiration 呼吸
S1_2	D13	synthetic artificially synthesized	synthetic 人工合成的
S1_2	D14	synthetic synthetic	synthetic 合成的

Appendix I. Visualization of Social Network and the Detailed Information in S3

**Figure A2.** Visualization of social network in S3: vocabulary-focused comments.**Table A9.** The detailed information in Figure 2.

Community	Label	English Translation	Original Content
S3_3	D1	recognition means “the act of identifying, acknowledging, or showing appreciation for something”.	recognition n. 認出; 承認, 讚賞
S3_3	D2	buffalo n.	buffalo n. 水牛
S3_3	D3	mosquito n.	mosquito n. 蚊子
S3_3	D4	recognition means “the act of identifying, acknowledging, or showing appreciation for something”. (Spelling error case)	recognition n. 認出; 承認; 賞識
S3_3	D5	opposite means “completely different; situated on the other side” (adj), “the other side” (n), “across from” (prep), or “to co-star” (v).	opposite a. 迥然不同的; 對面的 n. 對面 prep. 在...對面; 與...合演
S3_3	D6	opposite means “situated on the other side or contrary to something” (adj), “a counterpart or antonym” (n), or “facing something” (prep).	opposite a. 對面的; 相反的 n. 對立面; 反義詞 prep. 與...相對
S3_3	D7	recognition means “the act of identifying, acknowledging, or showing appreciation for something”.	recognition n. 認出; 認可; 賞識

Table A9. Cont.

Community	Label	English Translation	Original Content
S3_3	D8	opposite means “completely different or situated on the other side” (adj), “a counterpart” (n), or “facing something” (prep).	opposite a. 截然不同的;對面的 n. 對立面 prep. 與...相對與.....合演
S3_3	D9	recognition means “the act of identifying, acknowledging, or showing appreciation for something”.	recognition n. 認出; 承認; 賞識
S3_3	D10	fragrant a.	fragrant a. 香的, 芳香的
S3_3	D11	recognition means “the act of identifying, acknowledging, or showing appreciation for something”.	recognition n. 認出, 認識, 識別; 承認, 認可; 讀賞, 賞識
S3_3	D12	vessel means “a large ship, a container, or a blood-carrying structure in the body”.	vessel n. 大船; 容器; 血管
S3_3	D13	opposite means “situated on the other side; completely different” (adj), “a counterpart or contrary thing” (n), or “facing something” (prep).	opposite adj. 對面的; 迥然不同的 n. 對立面 prep. 與...相對; 與...合演
S3_3	D14	recognition means “the act of identifying, acknowledging, or showing appreciation for something”.	recognition n. 讀賞; 認出; 承認
S3_3	D15	remind v.	remind v. 提醒, 使想起
S3_3	D16	opposite means “on the other side; very different” (adj), “a counterpart” (n), or “facing something” (prep).	oppsite adj. 對面的, 另一邊的; 迥異的 n. 對立面 prep. 與...相對的
S3_3	D17	opposite means “contrary or completely different” (adj), “a counterpart” (n), or “facing something” (prep).	oppssite adj. 相反的, 迥然不同的; 對面的 n. 對立面 prep. 與...相對; 與...聯袂演出
S3_3	D18	opposite means “on the other side or completely contrary” (adj), “a counterpart or opposite thing” (n).	opposite adj. 對面的, 相對的, 相反的 n. 對立面

Appendix J. Visualization of Social Network and the Detailed Information in R3

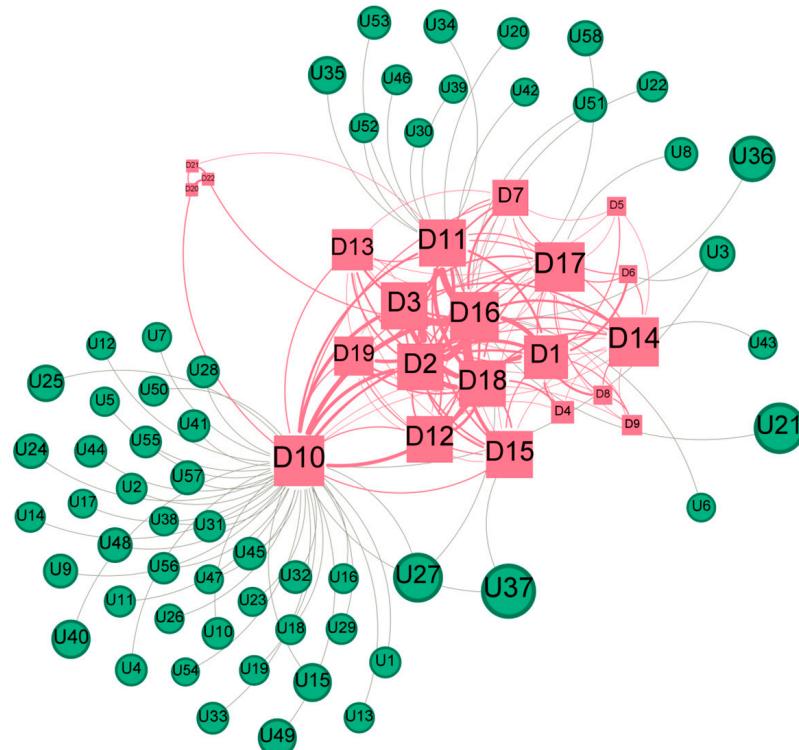
**Figure A3.** Visualization of social network in R3: vocabulary-focused comments.

Table A10. The detailed information in Figure 3.

Community	Label	English Translation	Original Content
R3_1	D1	radical big	radical 大的
R3_1	D2	radical big, extreme	radical 大的, 極端的
R3_1	D3	radical big, extreme	radical 大的極端的
R3_1	D4	radical huge	radical 巨大的
R3_1	D5	radical thorough	radical 徹底的
R3_1	D6	radical fundamental	radical 根本的
R3_1	D7	radical fundamental, thorough, extreme	radical 根本的, 徹底的, 極端的
R3_1	D8	radical extreme	radical 極端
R3_1	D9	radical extremely big	radical 極端大的
R3_1	D10	radical extreme	radical 極端的
R3_1	D11	radical extreme, big	radical 極端的, 大的
R3_1	D12	radical extreme, huge	radical 極端的, 巨大的
R3_1	D13	radical extreme, thorough	radical 極端的, 徹底的
R3_1	D14	radical extreme radical	radical 極端的 radical
R3_1	D15	radical extreme, big	radical 極端的, 大的
R3_1	D16	radical extreme, big	radical 極端的大的
R3_1	D17	radical extreme, big radical	radical 極端的 大的 radical
R3_1	D18	radical extreme, huge	radical 極端的 巨大的
R3_1	D19	radical extremely huge	radical 極端的極大的
R3_1	D20	radical extreme	radical 極端的
R3_1	D21	radical extreme, big	radical 極端的, 大的
R3_1	D22	radical extremely big	radical 極端的大的

Appendix K. Visualization of Social Network and the Detailed Information in W3

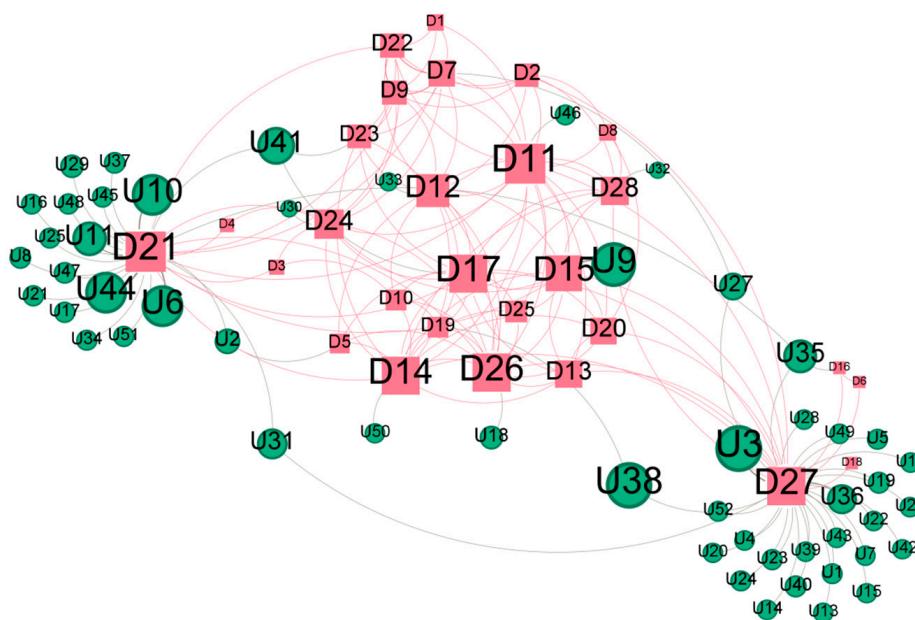
**Figure A4.** Visualization of social network in W3: vocabulary-focused comments.

Table A11. The detailed information in Figure 4.

Community	Label	English Translation	Original Content
W3_2	D1	foster cultivate, raise	foster 培養, 扶養
W3_2	D2	cultivate	cultivate 培養
W3_2	D3	foster develop	foster 養成
W3_2	D4	foster develop	foster 發展
W3_2	D5	form	form
W3_2	D6	caltivate	caltivate
W3_2	D7	foster cultivate	foster 培養
W3_2	D8	create cultivate foster	營造 培養 forster cultivate
W3_2	D9	foster, cultivate	foster, 培養
W3_2	D10	foster cultivate form develop	foster cultivate form 養成
W3_2	D11	cultivate, foster	cultivate, foster 培養
W3_2	D12	foster. form. cultivate develop	foster. form. cultivate 培養
W3_2	D13	foster cultivate atmosphere	foster cultivate atmosphere
W3_2	D14	cultivate form foster	cultivate form foster
W3_2	D15	foster and cultivate	foster and cultivate
W3_2	D16	ciltivate	ciltivate
W3_2	D17	form foster cultivate	form foster cultivate
W3_2	D18	cultivate develop	cultivat 養
W3_2	D19	form cultivate foster a habit	form cultivate foster a habit
W3_2	D20	foster cultivate create	foster cultivate 營造
W3_2	D21	foster	foster
W3_2	D22	foster cultivate	foster 培養
W3_2	D23	form develop foster cultivate	form 形成 foster 培養
W3_2	D24	foster foster foster	foster foster foster
W3_2	D25	foster cultivate develop	foster cultivate develop
W3_2	D26	foster cultivate form	foster cultivate form
W3_2	D27	cultivate	cultivate
W3_2	D28	cultivate develop cultivate	cultivate 培養 cultivate

Appendix L

Table A12. The detailed information in Figure 7.

Community	Label	Original Content
L2_1	D1	he should apologize for being rude to the guests.
L2_1	D2	it's more blessed to give than to receive
L2_1	D3	i told you not to talk about the matter in her presence.
L2_1	D4	can you wake me up at 7 o'clock tomorrow morning?
L2_1	D5	there must be a way to arrive at a diplomatic solution.
L2_1	D6	she should apologized for being rude to guests
L2_1	D7	i haven't had time to look for what you wanted.

Table A12. *Cont.*

Community	Label	Original Content
L2_1	D8	mary didn't refer to the accident she had seen
L2_1	D9	the teacher lined the children up in order of height
L2_1	D10	he should apologize for being rude to the guests
L2_1	D11	he should apologize for being rude to the guest
L2_1	D12	there must be a way to arrive at a diplomatic solution
L2_1	D13	we teacher lined the children up in order of the height
L2_1	D14	mary didn't refer to the accident she had seen
L2_1	D15	can you wake me up in the seven o'clock tomorrow morning
L2_1	D16	it is more blessed to give than to receive.
L2_1	D17	he should apolpgize for being rude to the guests
L2_1	D18	i have a friend whose father is a teacher.
L2_1	D19	is this what you have wanted for a long time
L2_1	D20	the teacher lined the children up in order of height.
L2_1	D21	our music teacher advised me to visit vienna.
L2_1	D22	our music teacher advised me to visit a vienna.
L2_1	D23	i told you not to talk about her in prensence
L2_1	D24	i tould you not to tell about the matter inher presence

References

1. Jiang, J.; Shi, S.; Luan, Z. Making Online English Learning More Engaging for Chinese College Students: A Comparative Analysis of MOOCs and Bilibili. In Proceedings of the 14th International Conference on Education Technology and Computers, Barcelona, Spain, 28–30 October 2022.
2. Yang, B.; He, H. Learning alone yet together: Enhancing between-learner social connectivity at scale. In Proceedings of the Ninth ACM Conference on Learning@ Scale, New York, NY, USA, 1–3 June 2022.
3. Hew, K.F. Unpacking the strategies of ten highly rated MOOCs: Implications for engaging students in large online courses. *Teach. Coll. Rec.* **2018**, *120*, 1–40. [[CrossRef](#)]
4. Zhang, L.-T.; Vázquez-Calvo, B.; Cassany, D. The emerging phenomenon of L2 vlogging on Bilibili: Characteristics, engagement, and informal language learning. *Prof. Inf.* **2023**, *32*, 1–16. [[CrossRef](#)]
5. Yang, Y. The danmaku interface on Bilibili and the recontextualised translation practice: A semiotic technology perspective. *Soc. Semiot.* **2020**, *30*, 254–273. [[CrossRef](#)]
6. Zeng, L.; Tan, Z.; Ke, Y.; Xia, L. Danmaku-Based Automatic Analysis of Real-Time Online Learning Engagement. *Int. J. Interact. Mob. Technol.* **2024**, *18*, 127–139. [[CrossRef](#)]
7. Peng, T.; Wang, T. The effect of using the danmaku mechanism in video learning: A pilot study. In Proceedings of the 13th International Conference on Education and New Learning Technologies, Online, 5–6 July 2021; EDULEARN21 Proceedings.
8. Li, S.; Zhu, H.; Qian, Y.; Ren, S.; Fang, B. Classification and quantification of Danmaku Interactions in online video lectures: An exploratory study. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 5656669. [[CrossRef](#)]
9. Khalid, H. Systematic Literature Review on Social Network Analysis. In Proceedings of the 2019 International Conference on Innovative Computing (ICIC), Lahore, Pakistan, 1–2 November 2019; pp. 1–7. [[CrossRef](#)]
10. Sikarwar, R.; Shakya, H.K.; Singh, S.S. A Review on Social Network Analysis Methods and Algorithms. In Proceedings of the 2021 13th International Conference on Computational Intelligence and Communication Networks (CICN), Lima, Peru, 22–23 September 2021; pp. 1–5. [[CrossRef](#)]
11. Bouhali, S.; Ellouze, M. Community detection in social network: Literature review and research perspectives. In Proceedings of the 2015 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI), Yasmine Hammamet, Tunisia, 15–17 November 2015; pp. 139–144. [[CrossRef](#)]
12. Majeed, S.; Uzair, M.; Qamar, U.; Farooq, A. Social Network Analysis Visualization Tools: A Comparative Review. In Proceedings of the 2020 IEEE 23rd International Multitopic Conference (INMIC), Bahawalpur, Pakistan, 5–7 November 2020; pp. 1–6. [[CrossRef](#)]

13. Choudhary, S.; Sharma, K.; Bajaj, M. Social Networks Analysis and Machine Learning: An Overview of Approaches and Applications. In Proceedings of the 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India, 14–16 June 2023; pp. 123–128. [[CrossRef](#)]
14. Oliveira, M.; Gama, J. An overview of social network analysis. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2012**, *2*, 99–115. [[CrossRef](#)]
15. Bonacich, P. Factoring and weighting approaches to status scores and clique identification. *J. Math. Sociol.* **1972**, *2*, 113–120. [[CrossRef](#)]
16. Bonacich, P. Some unique properties of eigenvector centrality. *Soc. Netw.* **2007**, *29*, 555–564. [[CrossRef](#)]
17. Newman, M. *Networks: An Introduction*; Oxford University Press: New York, NY, USA, 2010. [[CrossRef](#)]
18. Cui, Y.; Che, W.; Liu, T.; Qin, B.; Wang, S.; Hu, G. Revisiting pre-trained models for Chinese natural language processing. *arXiv* **2020**, arXiv:2004.13922.
19. Devlin, J. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
20. Cui, Y.; Che, W.; Liu, T.; Qin, B.; Yang, Z. Pre-training with whole word masking for chinese bert. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2021**, *29*, 3504–3514. [[CrossRef](#)]
21. MacQueen, J. Some methods for classification and analysis of multivariate observations. In Proceedings of the 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, 21 June–18 July 1965 and 27 December 1965–7 January 7 1966; University of California Press: Berkeley, CA, USA, 1967.
22. Arthur, D.; Vassilvitskii, S. K-means++: The advantages of careful seeding. In Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithm, SODA 2007, New Orleans, LA, USA, 7–9 January 2007.
23. Jain, A.K. Data clustering: 50 years beyond K-means. *Pattern Recognit. Lett.* **2010**, *31*, 651–666. [[CrossRef](#)]
24. Chowdhury, G.G. *Introduction to Modern Information Retrieval*; Facet Publishing: London, UK, 2010.
25. Sparck Jones, K. A statistical interpretation of term specificity and its application in retrieval. *J. Doc.* **1972**, *28*, 11–21. [[CrossRef](#)]
26. Salton, G.; Fox, E.A.; Wu, H. Extended boolean information retrieval. *Commun. ACM* **1983**, *26*, 1022–1036. [[CrossRef](#)]
27. Levenshtein, V. Binary codes capable of correcting deletions, insertions, and reversals. *Proc. Sov. Phys. Dokl.* **1966**, *10*, 707–710.
28. Serva, M.; Petroni, F. Indo-European languages tree by Levenshtein distance. *Europhys. Lett.* **2008**, *81*, 68005. [[CrossRef](#)]
29. Berger, B.; Waterman, M.S.; Yu, Y.W. Levenshtein distance, sequence comparison and biological database search. *IEEE Trans. Inf. Theory* **2020**, *67*, 3287–3294. [[CrossRef](#)] [[PubMed](#)]
30. Bastian, M.; Heymann, S.; Jacomy, M. Gephi: An open source software for exploring and manipulating networks. In Proceedings of the International AAAI Conference on Web and Social Media, San Jose, CA, USA, 17–20 May 2009.
31. Traag, V.A.; Waltman, L.; Van Eck, N.J. From Louvain to Leiden: Guaranteeing well-connected communities. *Sci. Rep.* **2019**, *9*, 5233. [[CrossRef](#)]
32. Blondel, V.D.; Guillaume, J.-L.; Lambiotte, R.; Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, *2008*, P10008. [[CrossRef](#)]
33. Squires, L. Enregistering internet language. *Lang. Soc.* **2010**, *39*, 457–492. [[CrossRef](#)]
34. Zhang, L.T.; Cassany, D. ‘Is it always so fast?’ Chinese perceptions of Spanish through danmu video comments. *Span. Context* **2019**, *16*, 217–242. [[CrossRef](#)]
35. Ginther, A.; Yan, X. Interpreting the relationships between TOEFL iBT scores and GPA: Language proficiency, policy, and profiles. *Lang. Test.* **2018**, *35*, 271–295. [[CrossRef](#)]
36. Paradowski, M.B.; Whitby, N.; Czuba, M.; Bródka, P. Peer Interaction Dynamics and Second Language Learning Trajectories During Study Abroad: A Longitudinal Investigation Using Dynamic Computational Social Network Analysis. *Lang. Learn.* **2024**, *74*, 58–115. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.