

## Post-print of Cluster Analysis and Thematic Deconstruction of Projects from the Columbia Center for Oral History Research

**Authors:** Wang Ruan, Deng Jun, Sun Shaodan, Song Xianzhi, Song Xueyan

**Date:** 2023-04-01T16:16:00+00:00

### Abstract

[Purpose/Significance] This study examines and uncovers the project characteristics of the Columbia Center for Oral History (CCOH) from a novel perspective, provides an in-depth deconstruction of project themes, and offers a reference benchmark for enriching the themes of oral history projects in China. [Method/Process] The study selected 9,586 oral history projects from the Columbia Center for Oral History as research subjects, utilized Python to crawl data, and performed cluster analysis and project theme deconstruction on the filtered 3,589 projects (filtering out 5,997 projects without abstracts) using Word2Vec and K-means. [Results/Conclusion] The projects of the Columbia Center for Oral History encompass four thematic dimensions:

People-centered themes—comprehensive and inclusive, covering groups from all walks of life; Institution-centered themes—embracing diverse strengths, leveraging their functional advantages; Region-centered themes—coordinated planning, linking regional system construction; Themes centered on specific periods, movements/events—innovative paradigms, reflecting on national and era themes.

### Full Text

## Cluster Analysis and Theme Deconstruction of Columbia Center for Oral History Projects

**Wang Ruan, Deng Jun, Sun Shaodan, Song Xianzhi, Song Xueyan**

School of Management, Jilin University, Changchun 130022

### Abstract:

[Purpose/Significance] This study examines and excavates the project characteristics of the Columbia Center for Oral History (CCOH) from a novel perspective.

tive, deconstructing project themes in depth to provide a reference benchmark for enriching oral history project themes in China.

**[Method/Process]** We selected 9,586 oral history projects from CCOH as research objects, captured data using Python, and performed cluster analysis on 3,589 filtered projects (excluding 5,997 projects without introductions) through Word2Vec and K-means to conduct thematic deconstruction.

**[Results/Conclusion]** CCOH projects encompass four thematic dimensions: People-themed projects—inclusive and radiating to all walks of life; Institution-themed projects—drawing on diverse strengths and leveraging functional advantages; Region-themed projects—coordinated planning with linked regional system construction; and Specific periods/movements/events-themed projects—innovative paradigms reflecting national and era themes.

**Keywords:** CCOH; project theme; Word2Vec; K-means

As the origin of oral history research, the establishment of the Columbia University Oral History Research Office laid the foundation for the substantial development of oral history in the United States. Subsequently, oral history rapidly developed in Canada, the United Kingdom, France, Singapore, Japan, and other countries, producing a large number of oral history researchers and professional research groups. On June 1, 2011, the research office was officially renamed the Columbia Center for Oral History (CCOH). After 70 years of development, CCOH's research scope has continuously expanded, penetrating almost every field, gradually forming its own research characteristics and becoming a benchmark institution for oral history research worldwide. Together with the Oral History Office at the University of California, Berkeley, and the Oral History Research Center at the University of California, Los Angeles, it is known as one of the three major oral history centers in the United States and a world-leading center for oral history practice and teaching.

China's earliest oral history practice can be traced back to the oral science and technology history project initiated by the Institute of History of Natural Sciences of the Chinese Academy of Sciences in 1980—lagging nearly 30 years behind the United States at the starting point. In recent years, with rising attention to intangible cultural heritage, oral history has received continuous emphasis. China has successively established institutions such as the Wenzhou University Oral History Research Institute and the Cui Yongyuan Oral History Research Center, aiming to better inherit and protect oral history resources. However, China's oral history project practice has progressed relatively slowly, with resource storage richness and description granularity being somewhat deficient, and project themes remaining relatively singular.

Based on this context, this study selects 3,589 oral history projects from CCOH (out of 9,586 total projects, excluding 5,997 without introductions) as research objects, uses Python to crawl project introductions, and employs Word2Vec (Word to Vector, also called “Word Embedding”) and K-means for cluster analysis and thematic deconstruction. Through reviewing and 梳理 ing themes such as people, institutions, regions, and specific periods/events, we examine and

deeply excavate project thematic characteristics from a fresh perspective, hoping to gain insights for optimizing China's oral history resource allocation and enriching project themes, which will have far-reaching significance for formulating oral history project practice strategies in China.

## 2. Origins and Development of the Columbia Center for Oral History

The Columbia University Oral History Research Office was founded by historian A. Nevins in 1948. In its early days, the office primarily collected documentary materials such as diaries and letters. With continuous development of information technology, oral interviews became the main method. Initial interviews focused on political figures, then gradually expanded to other fields including charitable organizations, media, business, arts, and social groups. Since the research office was renamed CCOH in 2011, in November 2013, to further distinguish the center's research and archival functions, CCOHR (Columbia Center for Oral History Research) and CCOHA (Columbia Center for Oral History Archives) were established separately. Currently, CCOH has collected over 8,000 textual and audio interviews, with its resource richness and thematic diversity broadening research perspectives on oral history both domestically and internationally.

Domestic scholars have extensively researched CCOH. From a developmental perspective, scholars such as Chen Shuixiang, Wang Yulong, and Liu Jiamin have briefly 梳理 ed CCOH's development overview. From a functional analysis perspective, Zhang Hongtao conducted detailed investigations on five aspects: position settings and personnel functions, resource construction and processing organization, resource archiving and use, education training and interdisciplinary cooperation, and funding sources and research surveys. From a micro perspective, at the resource collection level, CCOH adopts a model of "independent interviews as the main approach, supplemented by collaborative collection and donations," with staff not directly participating in oral history collection but recruiting experienced interviewers to plan and complete the entire interview process. At the organizational level, CCOH has established an advisory committee composed of leaders and experts from relevant Columbia University departments, research institutes, and libraries, responsible for oral history education, training, and research. At the management and operation level, CCOH is particularly rigorous and standardized, strictly adhering to detailed regulations regarding interview operations and legal ethics from the Oral History Association, Institutional Review Board, and Creative Commons throughout project construction. At the digital storage and retrieval level, because CCOH is located in Columbia University's Butler Library, the library serves as the digital center for American oral history resources, laying the foundation for oral history project database construction in processing oral audio, transcripts, and related materials. Additionally, the Oral History Portal jointly designed by CCOH and Columbia University Library's digital preservation team enables

keyword retrieval of oral history collections, projects, and individual interviews. At the dissemination and promotion level, CCOH not only provides open access to Columbia University Library's precious manuscripts but also sells oral history-related microform products through a dedicated online store, or presents project results through stage plays, exhibitions, animations, and other forms to actively interact with the public. In terms of funding, most CCOH projects receive support and awards from the U.S. National Science Foundation.

Although foreign scholars have also documented CCOH's development, they mostly focus on specific projects, either revealing complete project processes or exposing social realities reflected in projects, such as attention to labor, race, class, and gender, or preservation of special events. However, these studies mostly employ traditional methods like web surveys and literature analysis, with somewhat scarce innovative research methods. While existing literature provides a good theoretical foundation for CCOH research and offers some 启发 significance for advancing CCOH practice, domestic scholars focus too much on operational processes and neglect attention to CCOH oral history project themes, lacking fine-grained interpretation of project themes. Foreign scholars, while also recording CCOH's development, emphasize project themes but with single project samples that cannot encompass project overviews or radiate project breadth.

Therefore, this study hopes to examine CCOH projects from a new perspective, introducing Word2Vec and K-means methods to cluster analyze 3,589 filtered projects and deconstruct themes, attempting to deeply analyze project thematic characteristics from the four-dimensional level of people-events-regions-institutions, excavate project connotations, provide reliable basis for optimizing China's oral history resource allocation and enriching project themes, and offer profound implications for formulating China's oral history project practice strategies.

### 3. Research Methods

#### 3.1 Word2Vec

Word2Vec is a word vector training tool open-sourced by Google in 2013. The algorithm includes two models: CBOW (continuous bag-of-words model) and Skip-Gram (continuous skip-gram model). The former predicts feature words based on context, while the latter predicts context probability through feature words. Word2Vec has strong applicability and flexibility in thematic clustering, efficiently generating word vectors from large-scale unannotated corpora, and demonstrating significantly higher accuracy in keyword extraction for longer articles compared to other algorithms.

#### 3.2 K-means

K-means originated in 1957, proposed by S. Lloyd. Compared to other algorithms, it is suitable for processing large-capacity datasets, with more efficient

cluster integration. It depends on parameter values, especially the selection of K value, which determines experimental results. It supports cluster analysis of multiple attributes without any prior knowledge, including four steps:

**Step 1:** Randomly select k sample mean points, denoting the ith mean point as  $u_i$ .

**Step 2:** Calculate the distance from each sample point to each mean point, grouping the shortest distances together;  $j = 1, 2, \dots, k$ ,  $i = 1, 2, \dots, n$ , where  $d$  is the distance from the ith point to the jth mean, and  $x_i$  is the ith data point:

$$d = \sum_{i=1}^n (x_i - u_j) \quad \text{Formula (1)}$$

**Step 3:** Calculate the mean of each of the k new classes obtained in Step 2 to get new mean points:

$$\text{New mean} = \frac{\sum_{i=1}^n x_i}{n} \quad \text{Formula (2)}$$

**Step 4:** Repeat Steps 2 and 3 until the sample mean points obtained no longer change significantly.

Given the simplicity and efficiency of Word2Vec and K-means algorithms, many scholars have combined them for thematic cluster analysis in various fields. For example, H. Tao et al. used the Word2Vec model to enhance topic semantic information expression and set up a weighted K-means algorithm for topic identification. Shen Si et al. applied Word2Vec and K-means for emergency handling of food safety event-related Weibo and information data, exploring specific processes, methods, and evaluation of category knowledge mining results. This demonstrates that the combined use of these two methods has been widely applied in various fields with good experimental results.

This study first uses Python to capture data sources, performs NLTK tokenization preprocessing on the text, then combines Word2Vec and K-means algorithms for CCOH project cluster analysis. Specifically, Word2Vec converts text into vector representation, which is then input into the K-means integrated model as data to obtain the K value, i.e., clustering results, thereby determining the number of project themes for further thematic deconstruction.

## 4. Clustering of Columbia Center for Oral History Projects

### 4.1 Data Sources and Preprocessing

This study used Python to capture 9,586 oral history projects collected by CCOH. The author organized the projects, with results shown in [Figure 1: see original paper]. The crawled data was stored in the text editor SublimeText,

using the Scrapy crawler framework to collect project background introductions and save them in .CSV format. Partial code is shown in [Figure 2: see original paper].

## 4.2 High-Frequency Words and Word Cloud

Python was used for word frequency statistics on the crawled data source. After case merging and synonym merging of the text data preprocessed by the NLTK tokenization package, a high-frequency word list was obtained (see ). As shown in the table, high-frequency words ranking at the top facilitate identification of CCOH oral history project themes and are 启发 for research. CCOH oral history projects are conducted in “interview” form, suggesting each project collects, organizes, describes, stores, utilizes, shares, and promotes oral history resources around this basis to deepen the depth and breadth of CCOH oral history project research. Meanwhile, word frequencies such as “family,” “childhood,” “life,” “education,” and “background” revolve around interview backgrounds, while “president,” “Eisenhower,” “engineer,” “worker,” “Black,” and “White” demonstrate changes in interviewee identities. Word frequencies like “corporation,” “administration,” “university,” and “association” reflect organizational diversity; “Africa,” “Turkey,” and “Washington” reflect geographical breadth; and “war” and “campaign” highlight war backgrounds. Therefore, changes in high-frequency words reflect CCOH oral history project thematic characteristics from different levels, embedded in the semantic revelations of numerous projects.

To more clearly present project theme words and enhance visual effects, the author used WordArt to generate a high-frequency word cloud, as shown in [Figure 3: see original paper].

## 4.3 Clustering of CCOH Oral History Projects

Cluster analysis aims to more clearly understand CCOH project themes and explore project characteristics. This study used NLTK for tokenization preprocessing of the data, then applied Word2Vec from the Gensim package and K-means for cluster analysis. The experimental tool was Jupyter Notebook.

**4.3.1 Data Cleaning** Tokenization and stop-word processing are critical steps in text analysis data processing, essential for accuracy. Given differences between Chinese and English writing formats, before extracting feature words, this study first performed tokenization processing, removed useless symbols, and used NLTK tokenization preprocessing. A new “split” column was added to store data after tokenization and stop-word processing to adapt to text precision mode, as shown in [Figure 4: see original paper].

**4.3.2 Word Vector Construction** Word2Vec was used to convert text into vectors, with the average value of word vectors in the text representing the text

vector, denoted as  $x_1, x_2, x_3, x_4, x_5, \dots$ , converted to DataFrame format (see [Figure 5: see original paper]).

**4.3.3 K-value Selection** Regarding random selection of  $k$  values in  $k$ -means algorithm classification, two methods (average silhouette coefficient and elbow method) can currently evaluate optimal cluster numbers. This study adopts the elbow method, expressed by the following formula:

SSE is the sum of squared errors, representing clustering result quality, where  $C_i$  is the  $i$ th cluster,  $p$  is a sample point in  $C_i$ , and  $m_i$  is the centroid of  $C_i$ :

$$SSE = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2 \quad \text{Formula (3)}$$

As the clustering number  $k$  increases, sample division becomes more refined, and as the aggregation degree of each cluster improves, SSE gradually decreases. That is, when  $k$  is less than the true cluster number (as  $k$ 's increase substantially improves each cluster's aggregation degree), SSE's decline magnitude is large. When  $k$  reaches the true cluster number, the aggregation return from further increasing  $k$  rapidly diminishes, and SSE's decline magnitude sharply decreases, then tends to flatten as  $k$  continues increasing. In other words, the relationship graph between SSE and  $k$  forms an elbow shape, and the elbow point corresponds to the data's true cluster number.

For this sample clustering evaluation, the changing relationship between  $k$  value and SSE is shown in [Figure 6: see original paper]. Clearly, from the elbow graph, the dataset's optimal cluster number is 4, so CCOH project themes can be summarized into four aspects.

## 5. Thematic Analysis of Columbia Center for Oral History Projects

Based on clustering results, CCOH project themes can be clustered into four categories. This section provides in-depth interpretation of CCOH project research themes, revealing implicit thematic features.

### 5.1 People as Theme—Inclusive, Radiating to All Walks of Life

People-themed CCOH projects center on individuals or groups with certain characteristics or social influence, focusing on vivid character portrayals. Through interviews, they explore interviewees' occupations, genders, ages, deeds, and achievements, discovering their contribution trajectories in respective fields and future development directions of those fields, radiating to all industries. The author classified results by identity as follows:

**5.1.1 Political Figures** Such figures are mostly upper-class individuals, including authoritative political figures such as presidents and senators. The Eisenhower Administration Project records President Eisenhower's term (1953-1961), with interviewees including Eisenhower and his family, White House staff, cabinet members, political advisors, congress members, and ambassadors. Since July 1, 2019, CCOH has officially undertaken the Obama Presidency Oral History Project, the second U.S. presidential oral history project after Eisenhower's, which will comprehensively record decisions and measures during Obama's term and reveal former First Lady Michelle Obama's work and impact. The project is currently ongoing. For oral histories beyond its own construction, CCOH has promoted Chinese notable figures' oral memoirs since 1958, selecting important Chinese figures from various fields such as military strategists, philosophers, and writers (e.g., Liu Jingshan, He Guoguang, Mei Qiaolin). Meanwhile, historian Tang Degang's memoir of Li Zongren, compiled and edited, also belongs to China's oral history projects.

**5.1.2 Science and Technology Figures** Science and technology figures mainly include scientists, architects, engineers, pilots, and other groups. U.S. industrial development is renowned for advanced technology, complete categories, abundant resources, strong production capabilities, and high labor productivity. Aviation industry development has benefited from rapid technological innovation. The Aviation Project's interview memories trace back to the rapid development of aviation during the two world wars, interviewing engineers and pilots to record technical transformations and development histories of gliders, light aircraft, jet aircraft, and supersonic aircraft in manuscripts and archives. Similar to aviation, mining development also stems from new technological inventions. The Mining Engineers Project delves into miners' lives in Alaska and South Africa, vividly depicting labor intensity and character traits to portray unique laborer images and spiritual inspiration. As a crucial engine of U.S. economic development, the construction industry holds equally important status. Columbia University's Architecture Project assisted civil engineering development by selecting four great architectural founders (Charles-Edouard Le Corbusier, Walter Gropius, Ludwig Mies van der Rohe, Frank Lloyd Wright) as interviewees to interpret foundational architectural scenes and explore integration points between architecture and urban planning.

As material life becomes increasingly satisfied and enriched, people's spiritual worlds also continuously change. With the Nobel Prize recognized worldwide as the highest award in all fields, more scientists have entered public view. The Nobel Laureates on Scientific Research Project focuses on scientists' work logs and family life, while the Thomas Alva Edison Project portrays Edison's personal qualities and research journey, forming 1,048 pages of oral interview manuscripts that become precious historical materials for future generations to recall Edison.

**5.1.3 Folk Figures** Folk figures are most prominently craftspeople. At the beginning of American independence, almost no one mentioned arts and crafts treasures. With the centennial independence anniversary (1876), research methods commemorating art through oral interviews became popular. The American Craftspeople Project interviews craftspeople in five specific fields (ceramics, fiber, glass, metal, and wood), with terms like “gallery,” “furniture making,” “jewelry,” “pottery,” “wood carving,” and “weaving” interpreting American crafts development (especially over the past 30 years), attempting to define crafts’ social status in American society, revealing how this role has changed over time and why, helping people understand America’s rich cultural background.

**5.1.4 Performing Arts Figures** Performing arts figures involve actors, directors, and photographers. The film industry produces and sells cultural products in relatively independent material forms. Film is a performing, visual, and auditory art that captures images and sound through film, videotape, or digital media, supplemented by post-production editing. In 1971, the Hollywood Film Industry Project conducted a series of interviews with Hollywood actors, directors, photographers, and technicians, covering production editing, distribution, and film development. From silent to sound films, the U.S. gradually gave birth to numerous film industry giants such as Broadway, Hollywood, Disney, and Time Warner, driving rapid development of related industries and forming distinctive American culture.

**5.1.5 Social Figures** Entering the 21st century, people-themed projects remain the main research thread. Social figures focus on entrepreneurs, founders, social activists, and philanthropists. The first phase of the Atlantic Philanthropies Oral History Project (2005) included interviews with 139 people, encompassing Atlantic Philanthropies Foundation leaders, Feeney’s business partners, institutional partners, and grantees, focusing on recording philanthropist Feeney’s life. The second phase selected 85 interviewees, covering social leaders and business elites, obtaining little-known stories behind philanthropy development through their memories.

**5.1.6 Marginalized Figures** Moreover, this study should note that people-themed projects are not only voice channels for upper classes and various industry figures but also historical witnesses to the development and continuation of ordinary people. CCOH timely adjusts its perspective, shifting focus to marginalized and lower-class groups such as African Americans, Mexican immigrants, Blacks, women, and laborers. In alternating research 重心, CCOH gradually recognizes the importance of maintaining and preserving oral history materials and recording different classes’ development histories. The Black Journalists Oral History Collection project reports on civil rights events and racial segregation by African Americans from the late 19th century to the early 1970s. The Women’s History and Population Issues Project includes women related to

maternal and child health services, marriage counseling, and family planning. The Venezuelan Labor Leaders Oral History Collection upholds the philosophy of being rooted in life and singing for laborers, revealing lower-class laborer images from unique perspectives.

Thus, people-themed projects reflect the diversification of all walks of life with an inclusive attitude, radiating broad fields. They are rich and full without lacking depth and vitality, emphasizing harmony between individuals and groups, reflecting CCOH's non-limitation to local figures while also focusing on excavating rich Chinese historical materials, providing strong supplements for interested researchers. Interview audiences have transitioned from upper-class political figures to marginalized classes without discourse power, vertically excavating historical contexts and inheriting national memory. This unique perspective switching from flattening to vertical stretching helps CCOH better grasp social trends and maintain stable and balanced development.

## **5.2 Institutions as Theme—Drawing on Diverse Strengths, Leveraging Functional Advantages**

Institution-themed CCOH projects include universities, companies, associations, research institutes, foundations, and other organizations, typically using institutional evolution as the main thread to record institutional changes and development history.

**5.2.1 University/Company Projects** University projects focus on academic research and cultural inheritance. The New York University Project records New York University's development history through book publication. Company projects often focus on their own functions. As America's largest department store, the Federated Department Stores Project investigates consumer groups and purchasing intentions, archiving organizational development. The World Bank Project analyzes bank management processes, including policy formulation, project evaluation, loan management, equity, and risk analysis.

**5.2.2 Association/Research Institute Projects** Association projects focus on depicting trajectories of change. The American Association of Physics Teachers Project records the association's evolution since 1930. As associations continuously innovate and develop, research institutes also actively launch project research based on development needs. As America's first academic center dedicated to Russian and Soviet interdisciplinary research, the Harriman Institute Oral History Collection records 70 years of changes at the institute against the background of Russian academic research and policy conditions. Notably, religious activities are indispensable in Western life, so projects reflecting religious history are also important topics for research institutes, such as the China Missionaries Project launched by the Claremont Graduate University to assist Christian missionaries in recalling missionary activities.

**5.2.3 Foundation Projects** Typically, foundation projects are the strongest, broadest, and largest due to powerful financial group support. Established in 1911, the Carnegie Corporation is one of America's oldest and most influential foundations. The Carnegie Corporation Project has three parts: Part 1 records the foundation's development from 1966-1970; Part 2 records economic research on impoverished Blacks and democratic economic development funded from 1996-2000, outlining 25 years of organizational development structure and the impact of American philanthropy's transition to democracy; Part 3 studies the corporation's philanthropic plans and measures, more broadly discussing charitable responses to emergencies (such as Hurricane Katrina) and donation management during economic recessions.

Evidently, universities, companies, associations, research institutes, foundations, and other organizational institutions have all developed unique project construction styles. From their own functional perspectives—whether focusing on academic research and cultural inheritance, functional analysis, depicting change trajectories, or recording evolution history—they draw on diverse strengths to complete specialized field project construction around their own development needs. It is precisely due to complete organizational systems and intact organizational teams that CCOH research has gradually moved toward prosperity, with broader knowledge dissemination and audience reach.

### **5.3 Regions as Theme—Coordinated Planning with Linked Regional System Construction**

Region-themed CCOH projects typically limit oral history collection to specific geographical areas, focusing on regional politics, economy, culture, security, livelihood, education, and other fields to document regional development history.

The New York Political Studies (A) Project discusses New York's political development issues and achievements, while also focusing on the 1949 New York election, recording and covering all candidates' party histories and political election impacts (New York Political Studies (B) Project, New York Political Studies (C) Project). Beyond focusing on local regional history, CCOH also pays attention to oral history research in other regions, such as the Middle East and Central Europe. The Middle East focuses on Tunisia and Iran. The Tunisian Transition Oral History Collection records Tunisia's history during the revolution (2010-2011) and transitional government period (2011-2014), particularly highlighting regional politics, education, security, and economic development during Prime Minister Mehdi Jomaa's term. The Iranian-American Relations Project traces post-WWII Iran-U.S. diplomatic relations, presenting Iran's regional history and political structure during King Mohammad Reza Pahlavi's reign (1941-1979). Central Europe focuses on Hungary, Argentina, and Austria. The Hungary Project records regional art, politics, and international relations development; the Argentina Project records economic, industrial, and manufacturing-related technological changes; the Austria Project records music, literature, painting, and writing development.

In other words, region-themed oral history projects are carried out systematically, not only broadening regional historical research perspectives but also enhancing local oral history research value to some extent. CCOH has made important contributions to rescuing and protecting local oral history resources, recording regional political, economic, cultural, and educational characteristics through oral history, effectively linking local development layouts with other regional mechanisms, and fully playing a coordinated planning role.

#### **5.4 Specific Periods, Movements/Events as Theme—Innovative Paradigms Reflecting National and Era Themes**

CCOH projects themed around specific periods, movements, or events typically collect fixed 标志性 or public opinion events occurring in certain years or time periods, investigating and interviewing social hotspots or highly publicized issues to deeply excavate causes, processes, and results, often war-background-oriented.

To commemorate the only civil war in American history (April 12, 1861 - April 9, 1865), the Civil War Centennial Project records issues concerning slavery abolition and farmer land resettlement. The project not only maintained national unity and cleared the path for American capitalist development but also laid the foundation for America's rise as a world power. The Flying Tigers Project represents a brilliant chapter of China-U.S. joint combat against Japanese fascist aggression in WWII, occupying an important position in China-U.S. relations history and providing important enlightenment and guidance for enhancing mutual trust and inheriting friendly cooperation spirit.

Beyond commemorating specific period movements and development events, focusing on thematic historical event collection and 锁定 major historical events is another unique CCOH perspective. As an unprecedented historical commemorative event, the September 11, 2001 Oral History Narrative and Memory Project consists of five sub-projects with diversified interviewees. The September 11, 2001 Telling Lives Oral History Project covers survivors including escaped civilians, firefighters, paramedics, street vendors, bereaved families, and psychologists, maximizing restoration of the scene where terrorists hijacked five civilian aircraft to crash into New York's World Trade Center and Washington's Pentagon on September 11, 2001 (six buildings including the iconic World Trade Center twin towers were completely destroyed). To record the event's response and recovery work (September 11, 2001 Response and Recovery Oral History Project), a national memorial, memorial plaza, and museum were subsequently established. Since establishing the "9/11 Memorial Day" in 2002, the project has accumulated over 900 hours of audio materials (including 23 hours of video) from more than 600 interviewees by the 10th anniversary. To date, 687 hours of materials from 351 people have been opened to the public for archival access.

In summary, given the uniqueness of specific periods, movements, and events themselves, people can hardly interpret real history relying solely on documents and physical materials. In such cases, innovative paradigms using personal

memories and oral narratives from relevant witnesses and experiencers become particularly important. Although oral history itself has fluidity and uncertainty, making people doubt its content's authenticity and objectivity, witnesses' identity backgrounds and the fusion and transformation between personal memories and historical events 注定 that witnesses' oral history can provide true and relatively objective historical event restoration, just like documentary materials.

## 6. Implications for China's Oral History Practice

From the above CCOH project thematic analysis, the richness of American oral history project themes benefits from profound American cultural accumulation. CCOH project construction has formed a relatively mature system, broadly divided into four categories: people-themed, institution-themed, region-themed, or specific periods/movements/events-themed, breaking through the traditional narrative paradigm that previously emphasized politics and upper classes with a "top-down" approach. Remarkably, besides including some Chinese, Middle Eastern, and Central European histories, CCOH has always focused locally for years, taking rescuing local history and cultural folklore as primary goals, with all walks of life actively carrying out actions for comprehensive multi-field penetration, activating historical and cultural genes, and showing era development pulses, each with its own focus.

Specifically: **"Gathering People"**—Character richness. People are history's creators, and oral interviews with various industry groups reflect "people-oriented" humanistic thinking. **"Gathering Institutions"**—Institutional inheritance. Institutions, as important carriers of cultural inheritance, can awaken and reshape social memory, multi-dimensionally constructing social memory systems. **"Gathering Regions"**—Regional memory. As a tool that can relatively completely record stories, history, and memory, oral history has broad application space in recording regional history and strengthening regional memory. **"Gathering Events"**—Event reflection. Oral history does not directly "enter" daily life history but reshapes and recalls key historical periods and events in historical evolution, awakening event experiences through event records.

Although China's practice started nearly 30 years later than America's and remains in its initial growth stage, as guardians and inheritors of Chinese culture, China has considerable rich and vast human resources awaiting excavation. It is foreseeable that China has great potential in oral history project system construction, theme richness and granularity, and result transformation. Optimization can proceed from the following aspects:

**(1) Form professional teams and optimize system construction.** Compared with other literature resources, oral history resources have extensive content, diverse carrier forms, and value in rich collections and protecting human cultural heritage. With unique humanistic attributes inheriting historical memory, they require strong professionalism in collection, organization, dissemina-

tion, and utilization. CCOH recruits professionally knowledgeable interviewers to plan and participate in the entire interview process and establishes an expert advisory committee responsible for project preparation. Currently, China's oral history practice has not yet formed complete theories and methods, lacking systematic work norms and relevant personnel allocation. Therefore, China's oral history project development entities should strengthen human resource integration and optimize practice systems. At the education and training level, organize knowledge theory and skills training through professional training, field research, special lectures, and conference discussions, covering organization, interviewing, technical processing, and project coordination. At the personnel allocation level, focus on improving interviewers' professional knowledge literacy and thematic research capabilities, forming professional oral history teams to deliver talents for oral history research, actively carrying out and organizing various tasks with professional means to ensure smooth project implementation.

**(2) Activate cultural genes and show era pulses.** System construction is the foundation of project practice, and topic planning is the key to oral history services. CCOH covers American humanities, politics, economy, education, military, culture, and other fields from the four-dimensional level of "people-events-regions-institutions," diversely interpreting public life, institutional changes, regional development, and event perception, enriching project research depth. Currently, China's oral history project topic scope is relatively narrow, mostly focusing on local characteristics and rescuing regional culture, easily causing resource gaps, single content, lack of innovation, and repeated construction, which is not conducive to deep diffusion development of oral history construction. Therefore, expanding topic scope and carrying out oral history projects from multiple angles and dimensions to enrich theme granularity is crucial. At the regional allocation dimension, accept commissions from local governments, enterprises, professional associations, and other organizations, or cooperate across regions and even internationally. At content radiation breadth, not only focus on basic content like politics, economy, religion, art, and livelihood, but also serve as carriers and disseminators of Chinese culture, combining era hotspots and cultural genes to conduct good theme classification and mining, finding project practice entry points. On the basis of enriching China's oral history research content, gradually add other countries' oral history research to increase content richness and promote international research "grand integration."

**(3) Multi-subject collaboration and linked development mechanisms.** CCOH combines online and offline promotion, derivative product sales, multimedia interaction, and regular special activities with relevant departments, research institutes, and libraries, integrating new media and digital network technologies to demonstrate American oral history services' humanistic care and the original intention of providing quality oral history services for users.

Currently, China's oral history project practice remains in the initial stages of collection, storage, and publication dissemination, lacking deep development and knowledge mining. Although there is enthusiasm for database construc-

tion, there is a lack of digital humanities technical support, only providing simple project retrieval and online exhibition services with conservative service methods lacking appeal. Therefore, China's oral history project development entities such as governments, universities, LAM (Library, Archive, and Museum), and social organizations should actively learn from CCOH's linked participation mechanism, achieve multi-subject collaboration, break through traditional operation and development models, break resource silos, and broaden result display and dissemination channels.

In short, China's oral history project practice should continue to deeply build a Chinese-characteristic oral history project practice system on one hand, and on the other hand create a social environment conducive to oral history sustainable development, establish linked mechanisms for oral history resource development and utilization, enrich project research themes, and use physical exhibitions, AR/VR and other digital humanities technologies to share personal memories with the public, promote cultural heritage value dissemination, and assist cultural heritage sustainable development.

## References

- [1] Yang Xiangyin. Research on American Modern Oral History [M]. Beijing: China Social Sciences Press, 2016: 1.
- [2] Columbia University Interdisciplinary Center for Innovative Theory and Empirics (INCITE) [EB/OL]. [2020-02-08]. <https://www.incite.columbia.edu/ccohr/>.
- [3] Columbia Center for Oral History Projects [EB/OL]. [2019-12-08]. <https://oralhistoryportal.library.columbia.edu/project>.
- [4] Yang Xiangyin. On the Function and Difficulties of Oral History [J]. *Historiography Theory and Research*, 2000 (3): 37-46.
- [5] Zhang Hongtao. Research and Enlightenment on Oral History Work at Columbia University [J]. *Library and Information Service*, 2019, 63(3): 123-128.
- [6] Columbia Center for Oral History Research (CCOHR) [EB/OL]. [2019-12-08]. <https://www.ccohr.incite.columbia.edu/>.
- [7] Columbia Center for Oral History Archives (CCOHA) [EB/OL]. [2019-12-08]. <https://oralhistoryportal.library.columbia.edu/project?rq=CCOHA>.
- [8] Columbia Center for Oral History Research Guide [EB/OL]. [2019-12-08]. <https://library.columbia.edu/content/dam/libraryweb/locations/ohro/The%20Oral%20History%20Bibliography%20A%20CCOH%20Publication.pdf>.
- [9] Chen Shuixiang. Investigation and Enlightenment of Oral History Work in American Universities [J]. *Library Development*, 2018(5): 34-41.
- [10] Wang Yulong. Network Development and Utilization of American Oral History Archive Resources Based on Case Analysis [J]. *Archives and Construction*, 2017(2): 21-24.
- [11] Liu Jiamin. Characteristics of American Oral History and Its Reference [J]. *Lantai World*, 2017(1): 28-31.
- [12] Zhong Yuan, Wu Zhenhuan. Investigation and Analysis of Oral History Work in 10 American University Libraries [J]. *Library Journal*, 2019, 38(6):

83-91.

[13] Wang Peng, Fan Zhixin. Characteristics and Enlightenment of American Oral History Work [J]. China Archives, 2019(6): 76-77.

[14] Zhang Yu. Research on Legal and Ethical Issues in Oral History Projects [D]. Nanjing: Nanjing University, 2018.

[15] Ji Shuangqi. Characteristics and Enlightenment of American Library Oral History Resource Database Construction [J]. Library Work and Research, 2018(11): 40-45.

[16] Yang Xiangyin. Digital Revolution and American Oral History [J]. Social Science Front, 2016 (3): 106-120.

[17] Chen Fangrui. Characteristics and Enlightenment of American Library Oral History Services [J]. Library Work and Research, 2018, 1(4): 5-10.

[18] Wei Qiaoming. Research on Funding Sources of American Science and Technology Oral Archives Projects [J]. Lantai World, 2016(10): 32-35.

[19] Boyd DS. Stories from the collection: Columbia University Oral History Research Office [J]. Oral history review, 2001, 28(2): 137-138.

[20] Levy JA. Apollo Theater Oral History Project and the Apollo Theater Oral History Project at C.S. 154 by Harlem Apollo Theater (review) [J]. Oral history review, 2014, 41(1): 134-136.

[21] Vaughan AG. African American cultural history and reflections on jung in the African diaspora [J]. Journal of analytical psychology, 2019, 64(3): 320-348.

[22] Davis MP. Mexican voices/American dreams: an oral history of Mexican immigration to the United States [M]. New York: Henry Holt & Co., 1990.

[23] Absher A. The black musician and the white city: race and music in Chicago, 1900-1967 [J]. Journal of the Illinois State Historical Society, 2015, 109(5): 280.

[24] Lindsey Robinette Whitley. African American Oral History Project [EB/OL]. [2019-12-08]. [https://jewlscholar.mtsu.edu/bitstream/handle/mtsu/4893/Whitley\\_{{mtsu}}{{017](https://jewlscholar.mtsu.edu/bitstream/handle/mtsu/4893/Whitley_{{mtsu}}{{017)

[25] Miller KA. Reading and writing the post-9/11 cop: trauma, personal testimony, and Jess Walter's the zero [J]. Arizona quarterly: a journal of American literature culture and theory, 2014, 70(1): 29-52.

[26] Li Yuepeng, Jin Cui, Ji Junchuan. Keyword Extraction Algorithm Based on word2vec [J]. e-Science Technology & Application, 2015, 6(4): 54-59.

[27] Zeng Zhixiong. An Effective Hybrid Clustering Algorithm Based on Partition and Hierarchy [J]. Computer Applications, 2007, 27(7): 1692-1695.

[28] Oliveira G, Coutinho F, Campello R, et al. Improving k-means through distributed scalable metaheuristics [J]. Neurocomputing, 2017, 246: 45-57.

[29] Tao H, Li J, Luo T, et al. Research on topic trends based on weighted K-means [C]//2017 7th IEEE international conference on electronics information and emergency communication. Macau: IEEE, 2017: 457-460.

[30] Shen Si, Liang Xiaojing. Research on Automatic Clustering of Food Safety Events Based on word2vec and k-means Algorithms [J]. Information & Communications, 2018, 191(11): 8-10.

[31] Columbia University Library. Eisenhower Project [EB/OL]. [2020-02-08]. <https://clio.columbia.edu/catalog/4074583>.

[32] Columbia Center for Oral History Research. Obama Presidency Oral His-

tory Project [EB/OL]. [2020-02-08]. <https://www.ccohr.incite.columbia.edu/obama-presidency-oral-history-project>.

[33] Columbia Center for Oral History. Academia Sinica Oral History Collection [EB/OL]. [2020-02-08]. [https://oralhistoryportal.library.columbia.edu/results.php?component\\_{text}=&proj](https://oralhistoryportal.library.columbia.edu/results.php?component_{text}=&proj)

[34] Columbia Center for Oral History. Academia Sinica Oral History Collection - Liu Jingshan [EB/OL]. [2020-02-08]. [https://oralhistoryportal.library.columbia.edu/document.php?id=ldpd\\_](https://oralhistoryportal.library.columbia.edu/document.php?id=ldpd_)

[35] Columbia Center for Oral History. Academia Sinica Oral History Collection - He Guoguang [EB/OL]. [2020-02-08]. [https://oralhistoryportal.library.columbia.edu/document.php?id=ldpd\\_](https://oralhistoryportal.library.columbia.edu/document.php?id=ldpd_)

[36] Columbia Center for Oral History. Academia Sinica Oral History Collection - Mei Qiaolin [EB/OL]. [2020-02-08]. [https://oralhistoryportal.library.columbia.edu/document.php?id=ldpd\\_{](https://oralhistoryportal.library.columbia.edu/document.php?id=ldpd_{)

[37] Columbia University Library. Aviation Project [EB/OL]. [2020-02-08]. <https://clio.columbia.edu/catalog/4072718>.

[38] Columbia University Library. Mining Engineers Project [EB/OL]. [2020-02-08]. <https://clio.columbia.edu/catalog/4072675>.

[39] Columbia University Library. Architecture Project [EB/OL]. [2020-02-08]. <https://clio.columbia.edu/catalog/4076593>.

[40] Columbia Center for Oral History. Nobel Laureates on Scientific Research Project [EB/OL]. [2020-02-08]. [https://oralhistoryportal.library.columbia.edu/results.php?component\\_{text}="](https://oralhistoryportal.library.columbia.edu/results.php?component_{text}=)

[41] Columbia University Library. Thomas Alva Edison Project [EB/OL]. [2020-02-08]. <https://clio.columbia.edu/catalog/4074211>.

[42] Columbia Center for Oral History. American Craftspeople Project [EB/OL]. [2020-02-08]. [https://oralhistoryportal.library.columbia.edu/document.php?id=ldpd\\_{4072539}](https://oralhistoryportal.library.columbia.edu/document.php?id=ldpd_{4072539}).

[43] Columbia University Library. Hollywood Film Industry Project [EB/OL]. [2020-02-08]. <https://clio.columbia.edu/catalog/4072703>.

[44] Columbia Center for Oral History. Atlantic Philanthropies Oral History Project [EB/OL]. [2020-02-08]. [https://oralhistoryportal.library.columbia.edu/results.php?component\\_{text}="](https://oralhistoryportal.library.columbia.edu/results.php?component_{text}=)

[45] Columbia Center for Oral History. Black Journalists Oral History Collection [EB/OL]. [2020-02-08]. [https://oralhistoryportal.library.columbia.edu/document.php?id=ldpd\\_{11561600](https://oralhistoryportal.library.columbia.edu/document.php?id=ldpd_{11561600)

[46] Columbia University Library. Women's History and Population Issues Project [EB/OL]. [2020-02-08]. <https://clio.columbia.edu/catalog/4074553>.

[47] Columbia Center for Oral History. Venezuelan Labor Leaders Oral History Collection [EB/OL]. [2020-02-08]. [https://oralhistoryportal.library.columbia.edu/document.php?id=ldpd\\_{13](https://oralhistoryportal.library.columbia.edu/document.php?id=ldpd_{13)

[48] Columbia University Library. New York University Project [EB/OL]. [2020-02-08]. <https://clio.columbia.edu/catalog/4076397>.

[49] Columbia University Library. Federated Department Stores Project [EB/OL]. [2020-02-08]. <https://clio.columbia.edu/catalog/4074565>.

[50] Columbia University Library. World Bank Project [EB/OL]. [2020-02-08]. <https://clio.columbia.edu/catalog/4074552>.

[51] Columbia University Library. American Association of Physics Teachers Project [EB/OL]. [2020-02-08]. <https://clio.columbia.edu/catalog/4072421>.

[52] Columbia University Library. Harriman Institute Project [EB/OL]. [2020-02-08]. <https://clio.columbia.edu/catalog/13273504>.

[53] Columbia Center for Oral History. China Missionaries Project [EB/OL]. [2020-02-08]. [https://oralhistoryportal.library.columbia.edu/document.php?id=ldpd\\_{4074476}](https://oralhistoryportal.library.columbia.edu/document.php?id=ldpd_{4074476}).

[54] Columbia Center for Oral History. Carnegie Corporation Project - Part I [EB/OL]. [2020-02-08]. [https://oralhistoryportal.library.columbia.edu/document.php?id=ldpd\\_{4075191}](https://oralhistoryportal.library.columbia.edu/document.php?id=ldpd_{4075191}).

- [55] Columbia University Library. Carnegie Corporation Project - Part II [EB/OL]. [2020-02-08]. <https://clio.columbia.edu/catalog/5459604>.
- [56] Columbia University Library. Carnegie Corporation Project - Part III [EB/OL]. [2020-02-08]. <https://clio.columbia.edu/catalog/10488425>.
- [57] Columbia University Library. New York Political Studies (A) [EB/OL]. [2020-02-08]. <https://clio.columbia.edu/catalog/4076400>.
- [58] Columbia University Library. New York Political Studies (B) [EB/OL]. [2020-02-08]. <https://clio.columbia.edu/catalog/4076398>.
- [59] Columbia University Library. New York Political Studies (C) [EB/OL]. [2020-02-08]. <https://clio.columbia.edu/catalog/4076399>.
- [60] Columbia University Library. Tunisian Transition Oral History Collection [EB/OL]. [2020-02-08]. <https://clio.columbia.edu/catalog/13275174>.
- [61] Columbia Center for Oral History. Iranian-American Relations Project [EB/OL]. [2020-02-08]. [https://oralhistoryportal.library.columbia.edu/document.php?id=ldpd\\_{4077015}](https://oralhistoryportal.library.columbia.edu/document.php?id=ldpd_{4077015}).
- [62] Columbia University Library. Hungary Project [EB/OL]. [2020-02-08]. <https://clio.columbia.edu/catalog/4076118>.
- [63] Columbia University Library. Argentina Project [EB/OL]. [2020-02-08]. <https://clio.columbia.edu/catalog/4072586>.
- [64] Columbia University Library. Austria Project [EB/OL]. [2020-02-08]. <https://clio.columbia.edu/catalog/4072697>.
- [65] Columbia University Library. Civil War Centennial Project [EB/OL]. [2020-02-08]. <https://clio.columbia.edu/catalog/4074005>.
- [66] Columbia University Library. Flying Tigers Project [EB/OL]. [2020-02-08]. <https://clio.columbia.edu/catalog/4074563>.
- [67] Columbia Center for Oral History. September 11, 2001 Oral History Narrative and Memory Project [EB/OL]. [2020-02-08]. <https://oralhistoryportal.library.columbia.edu/results.php?component=9&document=13275174>.
- [68] Columbia Center for Oral History. September 11, 2001 Telling Lives Oral History Project [EB/OL]. [2020-02-08]. <https://oralhistoryportal.library.columbia.edu/results.php?component=9&document=13275174>.
- [69] Columbia Center for Oral History. September 11, 2001 Response and Recovery Oral History Project [EB/OL]. [2020-02-08]. <https://oralhistoryportal.library.columbia.edu/results.php?component=9&document=13275174>.

**Author Contributions:**

Wang Ruan: Responsible for data analysis, paper writing and revision;  
Deng Jun: Proposed overall research ideas and framework;  
Sun Shaodan: Responsible for data processing;  
Song Xianzhi: Responsible for data collection;  
Song Xueyan: Responsible for data collection.

**Author Affiliations:**

School of Management, Jilin University, Changchun 130022

*Note: Figure translations are in progress. See original paper for figures.*

*Source: ChinaXiv — Machine translation. Verify with original.*