

Quantitative Research on the Syntactic Structure of Tang and Song Prose Based on a Corpus — With a Discussion on the Evolutionary Laws of Classical Chinese Special Sentence Patterns

Huayang Shi

Department of General Education, Henan Vocational University of Science and Technology, Zhoukou 466000, Henan, China

Corresponding author: Huayang Shi (e-mail: HuayangShi1989@163.com)

ABSTRACT By applying quantitative analysis to Tang and Song prose, this study maps the syntactic evolution of classical Chinese. A 3.5- million-character syntactic corpus of Tang and Song prose is constructed, and quantitative methods such as frequency statistics and significance tests of differences are combined with vertical comparison and horizontal analysis to systematically explore the evolutionary characteristics of basic syntactic structures and special sentence patterns. The results show that basic syntactic structures exhibit a trend of significant increase in attributive-head structures and slight decrease in subject-predicate and coordinate structures. Special sentence patterns follow the laws of form simplification, marker intensification, and word-order regularization. The proportion of “bei” passive sentences rises from 8.7% in the Tang Dynasty to 15.2% in the Song Dynasty, while the proportion of “zhe...ye” judgment sentences decreases by 22.9%. Syntactic evolution is influenced by internal factors, including grammatical coordination and semantic precision demand, and external factors, including social change, the Ancient Prose Movement, and spoken-language penetration. The corpus and annotation framework provide methodological support for classical Chinese teaching, ancient book collation, and Chinese language history research.

INDEX TERMS tang and song prose, corpus, syntactic structure, quantitative linguistics, special sentence patterns

I. INTRODUCTION

A. RESEARCH BACKGROUND AND SIGNIFICANCE

1) Research Background

THE mid-Tang to the Northern Song Dynasty represents the “golden age” of prose, a period when the “Eight Great Prose Masters” refined the literary landscape much like master weavers perfecting an intricate loom. As the pinnacle of classical expression, Tang and Song prose functions as a masterfully woven cultural brocade that not only carries profound literary value but also preserves the intricate interlacing of linguistic threads— akin to the interplay of high-quality silk warp and weft—during a critical evolutionary stage of written Chinese. The structural complexity and aesthetic richness of these texts reflect a refined textual spinning process, where historical wisdom and linguistic innovation are tightly wound into a resilient fabric of national heritage, echoing the enduring durability and sophisticated patterns of historical craftsmanship [1, 2].

As the core carrier of linguistic expression, the evolutionary laws of syntactic structure are a key topic in

Chinese language history research [3]. With the development of corpus linguistics, quantitative analysis methods have provided technical possibilities to break through the limitations of traditional qualitative research, enabling accurate capture of the distribution differences and evolutionary trends of syntactic components [4, 5]. The construction of high-quality corpora is not only fundamental to theoretical linguistics but also serves as a critical technical support for the digital preservation of industrial heritage and the text mining of specialized history, such as the evolution of technical terminology.

Current academic research on Tang and Song prose mostly focuses on literary schools, literary value, and authors’ styles [6, 7], while research from a linguistic perspective mainly centers on vocabulary and rhetoric [8]. Quantitative research at the syntactic level still has gaps, especially the lack of systematic data support for the evolution of special sentence patterns.

2) Research Significance

Theoretically, this study constructs a syntactic corpus of Tang and Song prose and establishes a multidimensional quantitative analysis framework, filling the gap in quantitative research on syntactic evolution and improving the theoretical system of written Chinese evolution. Practically, the research results can provide precise references for the laws of sentence patterns in classical Chinese teaching, offer a syntactic annotation paradigm for ancient book collation, and provide reusable quantitative data and methodological reference for related research on Chinese language history.

B. REVIEW OF DOMESTIC AND FOREIGN RESEARCH STATUS

1) Research Status of Tang and Song Prose

From a literary perspective, the academic community has formed a research system centered on the Ancient Prose Movement and the Eight Great Prose Masters, focusing on issues such as the evolution of schools and the innovation of writing styles [9, 10].

From a linguistic perspective, research mostly focuses on vocabulary exegesis and rhetorical techniques, while syntactic research is dominated by qualitative description. For example, Hu sorted out the syntactic and rhetorical phenomena in argumentative prose to study the “methods” and “skills” of language use among the Eight Great Prose Masters [11]. Existing research mostly focuses on the grammatical functions of certain sentence patterns, lacking quantitative comparison based on large-scale corpora, making it difficult to reveal the overall evolutionary trend.

2) Research Status of Quantitative Chinese Syntactic Research

Corpus methods have become the high-precision looms of linguistic analysis, allowing scholars to achieve frequency statistics and distribution analysis of syntactic structures by weaving vast resources like the Comprehensive Chinese Text Corpus into a searchable textual fabric. This systematic approach enables researchers to trace the interlacing patterns of historical grammar, treating language evolution as a structurally dense where every syntactic thread can be quantified and mapped [12, 13].

However, existing research has two major shortcomings:

(1) Lack of targeted corpora, as general corpora are difficult to meet the needs of special research on Tang and Song prose;

(2) The quantitative dimension of special sentence pattern evolution is single, mostly limited to frequency statistics, lacking multi-dimensional analysis of form evolution and function expansion.

3) Research Status of the Evolution of Classical Chinese Special Sentence Patterns

Special sentence patterns such as judgment sentences and passive sentences are core research objects of classical

Chinese syntax [14]. The academic community has sorted out their basic evolutionary context, but there are many controversies, such as whether the scope of “bei” passive sentences expanded in the Song Dynasty and whether inverted sentences showed a simplification trend [15].

Existing research mostly relies on typical example sentences for argumentation, lacking quantitative verification based on large-scale corpora, making it difficult to accurately locate key evolutionary nodes.

C. RESEARCH CONTENT AND METHODS

1) Research Content

The core research content includes four aspects:

(1) Construct a syntactic corpus of prose covering major schools and themes of the Tang and Song Dynasties, and formulate a unified annotation system;

(2) Quantitatively analyze the frequency and proportion of basic syntactic structures such as subject-predicate, verb-object, and attributive-head structures, and compare their distribution differences between the Tang and Song Dynasties;

(3) Focus on judgment sentences, passive sentences, inverted sentences, and elliptical sentences to analyze the evolutionary trajectory of their specific forms;

(4) Explain the internal and external motivations and core laws of syntactic evolution combined with quantitative data.

2) Research Methods

Adopt interdisciplinary methods:

(1) Corpus method: Construct a special corpus and complete syntactic annotation and retrieval;

(2) Quantitative analysis method: Use frequency statistics, distribution density calculation, significance test of differences ($P < 0.05$ for significance), and other methods to process data;

(3) Comparative method: Conduct vertical comparison between the Tang and Song Dynasties and horizontal comparison across different themes and authors.

II. RESEARCH FOUNDATION: CORPUS CONSTRUCTION AND SYNTACTIC ANNOTATION SYSTEM

A. SELECTION OF TANG AND SONG PROSE CORPUS

1) Principles of Corpus Selection

Adhere to the principles of representativeness, authority, diversity, and timeliness:

(1) Cover major schools such as the Ancient Prose Movement and parallel prose;

(2) Prioritize classic masterpieces and works of the Eight Great Prose Masters from Complete Prose of the Tang Dynasty and Complete Prose of the Song Dynasty;

(3) Take into account themes such as political commentaries, historical biographies, and essays, as well as authors at different levels (including both renowned masters and lesser-known writers);

(4) Strictly distinguish between the Tang and Song periods: Tang Dynasty corpus is limited to the early to late Tang (618-907 AD), and Song Dynasty corpus is limited to the Northern to Southern Song (960-1279 AD).

2) Sources and Scale of the Corpus

Tang Dynasty corpus is selected from selected works of Complete Prose of the Tang Dynasty, Collected Works of Han Yu, Collected Works of Liu Zongyuan, etc.; Song Dynasty corpus is selected from selected works of Complete Prose of the Song Dynasty, Collected Works of Ouyang Xiu, Seven Collections of Su Shi, etc. The total scale of the corpus is 3.5 million characters, including 1.5 million characters from the Tang Dynasty and 2 million from the Song Dynasty. The specific composition is shown in Table 1.

B. CORPUS CONSTRUCTION PROCESS

1) Corpus Preprocessing

Adopt the “manual proofreading + machine assistance” method for preprocessing: first, correct text errors and unify variant characters and punctuation marks by referring to the Zhonghua Book Company annotated version; then use the jieba word segmentation tool for word segmentation, and optimize the segmentation results combined with Ancient Chinese Dictionary;

finally, complete part-of-speech tagging using a part-of-speech tagging system, with an accuracy rate of over 92% verified manually.

2) Corpus Structure Design

The corpus is divided into four layers:

- (1) Basic information layer: Record metadata such as author, dynasty, theme, and creation date;
- (2) Text layer: Store original text and preprocessed text;
- (3) Annotation layer: Annotate basic syntactic structures and special sentence patterns, and store annotation information in XML format;
- (4) Retrieval layer: Support precise retrieval by sentence pattern type, dynasty, author, and other multidimensional criteria.

C. CONSTRUCTION OF SYNTACTIC ANNOTATION SYSTEM

1) Annotation Principles

Adhere to the principles of scientificity, consistency, and practicality: based on Wang Li's Ancient Chinese syntactic theory, refer to existing classical Chinese syntactic annotation systems; formulate Manual for Syntactic Annotation of Tang and Song Prose to clarify the definition standards of various sentence patterns; take into account the needs of quantitative analysis and annotate core syntactic components and sentence pattern markers.

2) Core Annotation Content

Annotation of basic syntactic structures includes subject-predicate, verb-object, attributive-head (attributive + headword, adverbial + headword), predicate-complement, and coordinate structures, clarifying the definition standards and annotation symbols of each structure; annotation of special sentence patterns covers judgment sentences (“zhe...ye” pattern, “nai/ze” pattern, unmarked pattern [defined as judgment meaning conveyed through semantic logic and context without explicit structural markers], etc.), passive sentences (“yu” pattern, “wei...suo” pattern, etc.), inverted sentences (object fronting, attribute postposing, adverbial postposing), and elliptical sentences (subject ellipsis, object ellipsis, preposition ellipsis), annotating their specific forms and core marker words.

3) Annotation Quality Control

Form a 3-person annotation team to carry out annotation work after systematic training; randomly select 100,000 characters of corpus for consistency inspection, evaluated using Cohen's Kappa coefficient, with a Kappa value of 0.86 (>0.8 is excellent), meeting annotation standards; establish an error correction mechanism to collectively review and revise annotation differences.

III. QUANTITATIVE ANALYSIS OF BASIC SYNTACTIC STRUCTURES IN TANG AND SONG PROSE

A. DEFINITION AND CLASSIFICATION OF BASIC SYNTACTIC STRUCTURES

Refer to Chinese syntactic structure theory to define core syntactic structures [16]:

Subject-predicate structure refers to the declarative relationship of “subject + predicate” (e.g., “孔子曰”); verb-object structure refers to the dominant relationship of “verb + object” (e.g., “治国”); attributive-head structure includes attribute modifying headword (e.g., “贤臣”) and adverbial modifying headword (e.g., “甚善”); predicate-complement structure refers to the supplementary relationship of “predicate + complement” (e.g., “说之尽”); coordinate structure refers to the parallel/alternative relationship between components (e.g., “君臣父子”).

Clarify exclusion rules: exclude non-core structures such as parentheticals and independent elements, and merge synonymous heterogeneous forms.

B. QUANTITATIVE CHARACTERISTICS OF BASIC SYNTACTIC STRUCTURES IN TANG PROSE

1) Overall Distribution Characteristics

Statistics on 1.5 million characters of Tang Dynasty corpus show that among basic syntactic structures, the attributive-head structure has the highest proportion (32.1%), followed by the subject-predicate structure (25.3%) and verb-object structure (20.7%), while the coordinate structure (14.5%) and predicate-complement structure (7.4%) have relatively low proportions.

TABLE 1. Composition of the Corpus

Dynasty	Theme	Representative Authors	Number of Articles	Word Count (10k characters)	Proportion
Tang	Political Commentary	Han Yu, Liu Zongyuan, etc.	82	52	34.7%
Tang	Historical Biography	Liu Su, Li Zhao, etc.	45	48	32.0%
Tang	Essay	Liu Zongyuan, Liu Yuxi, etc.	73	50	33.3%
Song	Political Commentary	Ouyang Xiu, Su Shi, etc.	105	72	36.0%
Song	Historical Biography	Sima Guang, Ouyang Xiu, etc.	68	65	32.5%
Song	Essay	Su Shi, Wang Anshi, etc.	97	63	31.5%
Total	-	-	470	350	100%

The ranking of high-frequency syntactic structures is: attributive-head > subject-predicate > verb-object > coordinate > predicate-complement, reflecting the expressive characteristics of Tang prose that emphasizes modification and logical relationships.

2) Syntactic Differences in Tang Prose of Different Themes

The proportion of attributive-head structures in political commentaries reaches 36.2% (4.1 percentage points higher than the average), as precise expression of political views relies on attributes and adverbials to strengthen semantics; the subject-predicate structure has the highest proportion in historical biographies (28.5%), which is consistent with the basic narrative logic of “subject + predicate” in narrative texts; the proportion of predicate-complement structures in essays is slightly higher (9.3%), which is mostly used to describe scenery and emotions, requiring complements to improve expression.

C. QUANTITATIVE CHARACTERISTICS OF BASIC SYNTACTIC STRUCTURES IN SONG PROSE

1) Overall Distribution Characteristics

Statistics on 2 million characters of Song Dynasty corpus show that the proportion of attributive-head structures rises to 35.7%, the proportion of subject-predicate structures drops to 23.5%, the proportion of verb-object structures slightly rises to 21.2%, and the proportions of coordinate structures (13.1%) and predicate-complement structures (6.5%) slightly decrease.

Compared with the Tang Dynasty, the proportion of attributive-head structures increases significantly (an increase of 3.6 percentage points), and the proportion of subject-predicate structures decreases slightly, reflecting the more precise expression of Song prose.

2) Syntactic Differences in Song Prose of Different Themes

The distribution trend of syntactic structures in various themes of Song prose is consistent with that of the Tang Dynasty, but the difference range narrows.

The proportion of attributive-head structures in political commentaries is 38.1% (still the highest), the proportion of subject-predicate structures in historical biographies is 26.3%, and the proportion of predicate-complement structures in essays is 8.7%.

Compared with the same themes in the Tang Dynasty, the proportion of attributive-head structures in Song political

commentaries increases more significantly, reflecting more rigorous logical expression.

D. EVOLUTIONARY TRENDS OF BASIC SYNTACTIC STRUCTURES IN TANG AND SONG PROSE

1) Vertical Differences of Core Syntactic Structures

The comparison of the distribution of basic syntactic structures between the Tang and Song Dynasties is shown in Table 2. Through significance test ($p < 0.05$), the differences in attributive-head structures, subject-predicate structures, and coordinate structures between the Tang and Song Dynasties are statistically significant. The evolutionary trends are as follows: attributive-head structures continue to rise, subject-predicate structures slowly decline, verb-object structures remain basically stable, and coordinate structures and predicate-complement structures slightly decline.

2) Preliminary Exploration of Influencing Factors of Evolutionary Trends

Internally, the rise of attributive-head structures conforms to the inherent demand for “precise expression” of language—the strengthening of attributes and adverbials achieves semantic clarity without unnecessary redundancy, and does not simply increase sentence complexity;

Externally, the advocacy of “literature carrying Dao” in the Song Dynasty’s Ancient Prose Movement required prose to have rigorous logic, promoting the widespread use of attributive-head structures;

At the same time, the interactive penetration of spoken and written language in the Song Dynasty made the expression of subject-predicate structures more concise, leading to a slight decrease in their proportion.

These two trends are complementary rather than contradictory: attributive-head structures ensure semantic precision, while the streamlined expression of subject-predicate structures maintains conciseness, jointly optimizing the expressive efficiency of Song prose.

IV. QUANTITATIVE RESEARCH ON CLASSICAL CHINESE SPECIAL SENTENCE PATTERNS IN TANG AND SONG PROSE

A. CLASSIFICATION AND ANNOTATION RULES OF CLASSICAL CHINESE SPECIAL SENTENCE PATTERNS

The core research focuses on four types of special sentence patterns:

TABLE 2. Comparison of the Distribution of Basic Syntactic Structures Between the Tang and Song Dynasties

Type of Syntactic Structure	Proportion in Tang Dynasty	Proportion in Song Dynasty	Change Range	Significance Test (P-value)
Subject-predicate Structure	25.3%	23.5%	-1.8%	0.021 (significant)
Verb-object Structure	20.7%	21.2%	+0.5%	0.357 (not significant)
Attributive-head Structure	32.1%	35.7%	+3.6%	0.008 (significant)
Predicate-complement Structure	7.4%	6.5%	-0.9%	0.183 (not significant)
Coordinate Structure	14.5%	13.1%	-1.4%	0.042 (significant)

(1) Judgment sentences are divided into “zhe...ye” pattern, “nai/ze” pattern, unmarked pattern (judgment meaning conveyed through semantic logic, context, or word order without explicit structural markers), etc.; passive sentences are divided into “yu” pattern, “wei...suo” pattern, “bei” pattern, etc.;

(2) Inverted sentences are divided into object fronting, attribute postposing, and adverbial postposing;

(3) Elliptical sentences are divided into subject ellipsis, object ellipsis, and preposition ellipsis.

Annotation rules clarify: judgment sentences need to annotate judgment markers, passive sentences need to annotate passive markers and agents, inverted sentences need to annotate original word order and inversion type, and elliptical sentences need to annotate omitted components and restoration basis.

B. QUANTITATIVE ANALYSIS OF SPECIAL SENTENCE PATTERNS IN TANG PROSE

1) Overall Distribution Characteristics of Various Special Sentence Patterns

Statistics on Tang Dynasty corpus show that among special sentence patterns, elliptical sentences have the highest proportion (42.3%), followed by judgment sentences (28.7%) and inverted sentences (21.5%), while passive sentences have the lowest proportion (7.5%).

The “zhe...ye” pattern dominates judgment sentences (65.2%), the “wei...suo” pattern accounts for 58.1% of passive sentences, object fronting is the main type of inverted sentences (52.3%), and subject ellipsis is the most common in elliptical sentences (61.7%).

2) Differences in Special Sentence Patterns Among Tang Prose by Different Authors

It should be noted that the Eight Great Prose Masters, as advocates of the Ancient Prose Movement, had in tentional literary pursuits of archaism, but this study’s corpus also includes works by non-mainstream authors and different literary levels to ensure the representativeness of evolutionary conclusions. The proportion of inverted sentences in Han Yu’s prose reaches 25.1% (3.6 percentage points higher than the Tang average), and he particularly prefers object fronting (accounting for 62.4% of inverted sentences), reflecting his bold and exotic writing style. For example, “Bu zhi ju dou, bu jie huo” (Ignorant of sentence and pause, unclear of doubts) in *On Teachers* strengthens the emphasis through object fronting.

The proportion of elliptical sentences in Liu Zongyuan’s prose is 45.7%, mostly used in the description of scenery in essays. Continuous subject ellipsis creates a concise artistic conception, such as in *The Little Stone Pool*; in his historical biography *Biography of Captain Duan*, subject ellipsis accelerates the narrative rhythm. Wang Bo’s early Tang prose, which belongs to the “Essay” theme in the corpus, has a “zhe...ye” judgment sentence proportion as high as 72.5%, highlighting the stylistic feature of the combination of prose and parallel prose (a cross-dimensional characteristic that does not conflict with thematic classification) ;

The proportion of inverted sentences in Chen Ziang’s high Tang prose is 22.8%, adjusting the sentence rhythm through adverbial postposing.

In general, the choice of special sentence patterns by famous Tang authors is deeply bound to their writing styles and themes. The syntactic innovation of Han Yu and Liu Zongyuan laid the linguistic foundation for the Ancient Prose Movement, while early and high Tang authors reflected the transitional traces from parallel prose to ancient prose.

C. QUANTITATIVE ANALYSIS OF SPECIAL SENTENCE PATTERNS IN SONG PROSE

1) Overall Distribution Characteristics of Various Special Sentence Patterns

Statistics on Song Dynasty corpus show that the proportion of elliptical sentences drops to 37.6%, the proportion of judgment sentences rises to 31.2%, the proportion of inverted sentences drops to 18.9%, and the proportion of passive sentences rises to 9.3%.

Compared with the Tang Dynasty, the proportion of “zhe...ye” judgment sentences drops to 42.3%, the proportion of “nai/ze” pattern rises to 28.5%; the proportion of “bei” passive sentences rises from 8.7% in the Tang Dynasty to 15.2%, and the proportion of object fronting in inverted sentences drops to 41.8%.

2) Differences in Special Sentence Patterns Among Song Prose by Different Authors

Similar to the Tang Dynasty, the Song Dynasty corpus includes both the Eight Great Prose Masters and other authors to avoid overgeneralization from literary giants’ intentional choices. The distribution of special sentence patterns in Su Shi’s prose is the most balanced, with elliptical sentences accounting for 35.2% and judgment sentences for 32.1%. The “nai/ze” judgment sentences and “bei” passive

sentences are used more frequently than the Song average, reflecting his plain and natural writing style. For example, “Ci fei Meng De zhi kun yu Zhou Lang zhe hu” (Is this not the place where Meng De was trapped by Zhou Lang) in Ode to the Red Cliff integrates judgment and passive functions, and the unmarked judgment sentence at the end of Note on a Night Out at Chengtianshi Temple shows a plain style.

The proportion of passive sentences in Ouyang Xiu’s prose reaches 10.5%, with equal emphasis on the “wei...suo” pattern and “bei” pattern. “Shen si guo mie, wei tian xia xiao” (Lost his life and destroyed his country, laughed at by the world) in Preface to the Biography of the Musicians in the Five Dynasties and the simplified judgment sentences in The Old Drunkard’s Pavilion balance narrative accuracy and fluency.

The proportion of judgment sentences in Wang Anshi’s political commentaries is as high as 34.7%, and the “nai/ze” pattern accounts for 31.2%, clarifying his stance through judgment sentences.

The proportion of elliptical sentences in Zeng Gong’s prose is 36.5%, enhancing the compactness of the text through subject ellipsis.

The proportion of inverted sentences among famous Song authors is generally lower than that of Tang authors, reflecting the trend of sentence pattern regularization, while highlighting the diversity of writing styles.

D. QUANTITATIVE EVIDENCE OF THE EVOLUTION OF SPECIAL SENTENCE PATTERNS BETWEEN THE TANG AND SONG DYNASTIES

The comparison of the evolution of core special sentence patterns between the Tang and Song Dynasties is shown in Tables 3 and 4. Quantitative evidence shows three major trends:

(1) Judgment sentences shifted from being dominated by the “zhe...ye” pattern (65.2% in the Tang Dynasty) to the “nai/ze” pattern and unmarked pattern (57.7% in total in the Song Dynasty). For example, the doublemarked judgment sentence in Han Yu’s On Teachers was simplified to an unmarked pattern in Su Shi’s Note on a Night Out at Chengtianshi Temple, which meets the demand for plain writing style in the Song Dynasty.

(2) The proportion of “bei” passive sentences rose from 8.7% in the Tang Dynasty to 15.2% in the Song Dynasty, while the proportion of the “wei...suo” pattern slightly decreased. “Wu mu bei lue” (My mother was captured) in Ouyang Xiu’s Epitaph on the Longgang reflects the absorption of spoken language by written language.

(3) The overall proportion of inverted sentences dropped from 21.5% to 18.9%, with a significant decrease in the proportion of object fronting (-3.3%) and a slight increase in adverbial postposing (+1.3%). The adverbial postposing in Su Shi’s Ode to the Red Cliff balances rhythm and expression.

(4) The proportion of elliptical sentences dropped from 42.3% to 37.6%, with a decrease in subject ellipsis. Zeng Gong’s Note on the Ink Pool uses pronouns to supplement subjects, improving clarity.

Overall, Tang sentence patterns are both complex and innovative, while Song sentence patterns are more concise and clear, achieving a balance between the standardization and practicality of written language.

Unmarked judgment sentences refer to structures where judgment meaning is conveyed through semantic logic, context, or word order without explicit markers such as “zhe”, “ye”, or “nai”.

V. LAWS AND MOTIVATIONS OF SYNTACTIC EVOLUTION IN TANG AND SONG PROSE

A. CORE LAWS OF SYNTACTIC STRUCTURE EVOLUTION

1) Evolutionary Laws of Basic Syntactic Structures

Based on quantitative data, basic syntactic structures present the law of “one rise, three stabilities, and one decline”:

Attributive-head structures continue to rise, becoming the dominant syntactic structure, reflecting the trend of precise expression; verb-object structures remain basically stable, with a solid function as the core dominant structure; subject-predicate structures and coordinate structures slightly decline, while predicate-complement structures remain basically stable; overall, it presents the characteristics of “structural balance and functional specialization”, with clearer division of labor among various structures.

2) Evolutionary Laws of Special Sentence Patterns

The evolution of special sentence patterns follows three major laws: “simplification of forms, intensification of markers, and regularization of word order”:

Judgment sentences are simplified from double markers (“zhe...ye”) to single markers (“nai/ze”) or unmarked; passive sentences are concentrated from multiple formats to the “bei” pattern, with more explicit marker words; the word order of inverted sentences is gradually close to modern Chinese, with reduced use of irregular word order; elliptical sentences change from high-frequency use to moderate use, balancing conciseness and clarity.

B. INTERNAL MOTIVATIONS OF SYNTACTIC EVOLUTION

The coordination of the grammatical system is the core internal motivation. The evolution of syntactic structures needs to adapt to the consistency between semantic expression and the grammatical system. For example, the rise of attributive-head structures is consistent with the demand for semantic precision. The precision of semantic expression promotes the intensification of special sentence pattern markers. For example, the emergence of “bei” passive sentences makes passive relationships clearer.

TABLE 3. Comparison of the Distribution of Core Forms of Judgment Sentences and Passive Sentences in Tang and Song Prose

Type of Sentence	Specific Form	Proportion in Tang Dynasty (within the type)	Proportion in Song Dynasty (within the type)	Change Range
Judgment Sentence	“zhe...ye” Pattern	65.2%	42.3%	-22.9%
Judgment Sentence	“nai/ze” Pattern	18.7%	28.5%	+9.8%
Judgment Sentence	Unmarked Pattern	16.1%	29.2%	+13.1%
Passive Sentence	“wei...suo” Pattern	58.1%	52.4%	-5.7%
Passive Sentence	“yu” Pattern	33.2%	32.4%	-0.8%
Passive Sentence	“bei” Pattern	8.7%	15.2%	+6.5%

TABLE 4. Comparison of the Distribution of Inverted Sentences and Elliptical Sentences in Tang and Song Prose

Type of Sentence	Specific Type	Proportion in Tang Dynasty (in overall sentences)	Proportion in Song Dynasty (in overall sentences)	Change Range
Inverted Sentence	Object Fronting	11.2%	7.9%	-3.3%
Inverted Sentence	Attribute Postposing	6.3%	5.7%	-0.6%
Inverted Sentence	Adverbial Postposing	4.0%	5.3%	+1.3%
Elliptical Sentence	Subject Ellipsis	25.9%	22.9%	-3.0%
Elliptical Sentence	Object Ellipsis	10.8%	9.2%	-1.6%
Elliptical Sentence	Preposition Ellipsis	5.6%	5.5%	-0.1%

The internal analogy mechanism of language promotes the unification of sentence patterns. For example, judgment sentences analogy to single-marker formats.

The grammaticalization process promotes the functional solidification of marker words. For example, “nai/ ze” gradually grammaticalizes from adverbs to judgment markers.

C. EXTERNAL MOTIVATIONS OF SYNTACTIC EVOLUTION

Social changes in the Tang and Song Dynasties drove changes in language expression needs. The development of commodity economy in the Song Dynasty expanded the scope of written language use, requiring sentence patterns to be more understandable and promoting the regularization of word order.

The advocacy of “literature carrying Dao” and “abandoning hackneyed expressions” in the Ancient Prose Movement not only promoted the innovation of prose language but also retained core classical Chinese characteristics, forming a unique evolutionary trajectory.

The interactive penetration of spoken and written language is significant. The use of “bei” passive sentences in Song spoken language influenced written language, simplifying the format of passive sentences. Cultural exchange and the inheritance of ancient books ensured the continuity of syntactic evolution, allowing the continuation and innovation of core sentence patterns.

D. MUTUAL VERIFICATION BETWEEN QUANTITATIVE DATA AND TRADITIONAL QUALITATIVE CONCLUSIONS

The quantitative data of this study verifies the rationality of traditional qualitative conclusions. For example, the view that “Song prose is more plain” can be supported by data such as the decrease in the proportion of inverted sentences and the simplification of special sentence patterns.

At the same time, it supplements the detailed gaps in traditional research. For example, it accurately locates the

increase range of “bei” passive sentences in the Song Dynasty (+6.5%), clarifying the key stage when it became the core passive format. It corrects some qualitative biases. For example, traditional research often emphasizes that the “zhe...ye” judgment sentence was “absolutely dominant” in the Tang Dynasty, implying an overwhelming proportion close to 100%. However, this study’s quantitative data shows that its proportion was 65.2%—though it was still the dominant form, it was not “absolute” as it accounted for less than two-thirds of all judgment sentences, with other formats such as “nai/ze” and unmarked patterns accounting for 34.8% in total.

VI. RESEARCH CONCLUSIONS AND PROSPECTS

A. MAIN RESEARCH CONCLUSIONS

(1) Constructed a 3.5-million-character syntactic corpus of Tang and Song prose and established a scientific syntactic annotation system, providing basic resources for related research; This system provides a reusable quantitative paradigm for the linguistic analysis of thematic archives, such as ancient records of techniques and craft evolution.

(2) Clarified the evolutionary characteristics of basic syntactic structures in Tang and Song prose, with a significant increase in the proportion of attributive-head structures and a slight decrease in subject-predicate and coordinate structures;

(3) Revealed the evolutionary laws of classical Chinese special sentence patterns, presenting the trend of “simplification of forms, intensification of markers, and regularization of word order”, with an increase in the proportion of “bei” passive sentences and unmarked judgment sentences;

(4) The internal motivations of syntactic evolution involve the structural coordination of the grammatical loom and the demand for semantic precision, while external motivations—such as social changes, the Ancient Prose Movement, and evolving pragmatic needs—act as the environmental tensions that reshape the tension and density

of the linguistic fabric. These forces work in tandem to ensure that the interlacing of syntax remains both functional and resilient, allowing the language to adapt its textural complexity to the shifting requirements of each era.

B. LIMITATIONS OF THE RESEARCH

(1) The corpus coverage is limited, not including works by minor authors and non-mainstream themes;

(2) The annotation system focuses on core sentence patterns, not covering marginal types such as fixed sentence patterns;

(3) Quantitative indicators are mainly frequency and proportion, lacking multi-dimensional analysis combining semantic and pragmatic dimensions;

(4) The analysis of evolutionary motivations focuses on the macro level, with insufficient discussion on micro mechanisms.

(5) The cross-domain applicability of the current quantitative model, particularly its performance on highly technical or specialized historical texts, requires further empirical verification.

C. FUTURE RESEARCH PROSPECTS

(1) Expand the corpus scale to include more authors and themes in the future;

(2) Refine annotation dimensions and carry out comprehensive analysis combining semantics and pragmatics;

(3) Conduct cross-genre comparative research, comparing syntactic characteristics with poetry, parallel prose, and other genres;

(4) Expand the time span by combining unearthed documents to sort out the complete trajectory of syntactic evolution;

(5) Use advanced technologies such as machine learning to optimize the annotation and analysis process and deepen the research on evolutionary mechanisms.

AUTHOR CONTRIBUTIONS

Huayang Shi designed, collected and analyzed the data, and drafted the manuscript. Huayang Shi conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Huayang Shi participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

- [1] Y. Xie, "Self-regulation: Stylistic criticism in Tang and Song prose," *Central Plains Culture Research*, vol. 12, no. 4, pp. 89-98, 2024, doi: 10.16600/j.cnki.41-1426/c.2024.04.002.
- [2] X. Y. Zheng, "A study on disyllabic words in Liu Zongyuan's prose," *Hebei Normal University*; 2014.
- [3] S. Gao and L. M. Zhang, "Diachronic evolution research of Chinese syntax based on dependency treebank," *Heilongjiang Social Sciences*, no. 6, pp. 86-92, 2025, doi: 10.20230/j.cnki.hljshkx.2025.06.011.
- [4] R. T. Zhou, "Enlightenment of corpus linguistics on the construction of ancient Chinese character corpus," *Journal of Ancient and Modern Chinese Creation*, no. 22, pp. 124-127, 2025, doi: 10.20024/j.cnki.CN42-1911/I.2025.22.037.
- [5] S. Q. Wu and Y. Y. Jiang, "New development of the quantitative turn in cognitive linguistics: Cognitive corpus linguistics," *Foreign Language and Literature Research (Series)*, no. 1, pp. 76-90, 2024.
- [6] Y. D. Peng, "Contentment with circumstances and attachment to one's native land - A study on the motif of frustration in official careers among the four great prose masters of the Tang and Song Dynasties," *Masterpieces*, no. 20, pp. 88-90, 2025.
- [7] Y. Q. Li, "A study on the Eight Great Prose Masters of the Tang and Song Dynasties from the perspective of holism," *Journal of Jiamusi Vocational Institute*, vol. 36, no. 7, pp. 96-97, 2020.
- [8] F. Guo, "Ci poetry of the Song Dynasty and poetics of the Tang and Song Dynasties," *Peking University Press*; 2023. 313 p.
- [9] J. C. Chen and X. Q. Lü, "The similarities and differences in stylistic innovation between the Ancient Prose Movement and the New Yuefu Movement," *Journal of Guangdong Open University*, vol. 33, no. 5, pp. 72-76, 2024.
- [10] J. Duan and Z. L. Yang, "Viewing Tang and Song prose through the "Eight Great Masters"," *Journal of Hebei North University*, no. 2, pp. 7-10, 2006.
- [11] J. Hu, "A study on syntactic rhetoric in argumentative prose of the Eight Great Prose Masters of the Tang and Song Dynasties," *Journal of Ningxia University (Humanities and Social Sciences Edition)*, vol. 42, no. 4, pp. 76-80, 2020.
- [12] S. Gao, L. Qian, and Q. Ding, "Construction of a general Chinese academic word list driven by corpus," *China Terminology*, vol. 26, no. 3, pp. 84-92, 2024.
- [13] X. Li, "Construction of a Chinese rhetorical question corpus and research on recognition methods," *Soochow University*; 2022, doi: 10.27351/d.cnki.gszzhu.2022.002039.
- [14] L. X. Ma, "A brief discussion on four common sentence patterns in classical Chinese," *Language, Mathematics and Foreign Language Learning (Senior High School Edition, Late Issue)*, no. 2, pp. 28-29, 2024.
- [15] Y. J. Zhang, "Passive and disposal constructions in the Southern Song ci collection *White Clouds over the Mountains*," *Journal of Xinyang Normal University (Philosophy and Social Sciences Edition)*, vol. 33, no. 4, pp. 83-86, 2013.
- [16] B. J. Zhang, "Syntactic structure and pragmatic structure of Chinese," *Chinese Language Learning*, no. 2, pp. 3-12, 2011.