Acta Linguistica Asiatica

TABLE OF CONTENTS

Foreword

Dear Linguistic Enthusiasts and Scholars,

It is with our great pleasure to announce that the new issue of Acta Linguistica Asiatica is now available. As you pursue this issue, you will notice a refreshed look inspired by the new outlines of the University of Ljubljana, reflecting our commitment to growth and innovation.

We take pride in our dedicated team whose expertise drives our exploration of linguistic intricacies and cultural nuances. This issue features four scientific articles and one review article, offering diverse perspectives on Chinese and Japanese as the East Asian languages in focus. We extend our heartfelt gratitude to all contributors and reviewers whose scholarly rigor enriches our journal.

To begin with, **Tereza SLAMĚNÍKOVÁ** and **David UHER** wrote an article entitled "Rhythmical Features of Spontaneous Spoken Narrative in Mandarin Chinese", in which they introduce and expand a notation system for teaching Chinese rhythm pioneered by Oldřich Švarný, a Czech phonetician. They analyze spontaneous speech to outline rhythmical features in personal narratives and compare these features with a contemporary understanding of Chinese speech rhythm, showcasing the applicability of Švarný's framework in diverse spoken discourses.

Following is the work by **KUANG Lincai** "A Colexficational Analysis of *Chū* in Modern Chinese" that deals with directional verbs in modern Chinese, in particular with *chū*. The author delves into colexification and mechanisms surrounding *chū*, using the CLICS3 database. It explores colexified pairs of *chū* and constructs a colexificational network, emphasizing conceptual metaphor and metonymy as core mechanisms.

Yet another scientific paper was authored by **YANG Yongzhong** and is entitled "The Structure of Chinese Compounds: The Perspective of Predicative Implicitness", which explored Chinese nominal compounds' structure, relations, generation, and constraints, emphasizing predicate implicitness. Findings reveal varying constituent functions aligning morphological and syntactic structures, and the author offers a detailed account of their hierarchy.

Trang PHAN and **Giang VU** submitted the paper "*Wh*-questions in Japanese: Challenges for Vietnamese L2 learners" in which they explored how Vietnamese learners handle structural differences in Japanese wh-questions. Results showed a correlation between Japanese proficiency and performance. Though learners were evaluated as more accurate with mirrored word orders, advanced learners also struggled with native-like interpretations. The study suggests explicit instruction to enhance accuracy in teaching.

Last but not least is the survey article by **Duc Du HUYNH**, "Sino-Vietnamese Bilingual Dictionaries from the Middle Ages to the Present", in which the author thoroughly surveys Sino-Vietnamese bilingual dictionaries from the Middle Ages to the Modern Ages, often overlooked in scholarly discourse. It delineates their evolution, regional differences, and historical context, aiming to catalyze further research in this field.

Editors and Editorial Board invite the regular and new readers to engage with the content, to question, challenge, and reflect. We hope you have a pleasant read full of inspiration and a rise of new research ideas inspired by these papers.

Editors

RESEARCH ARTICLES

Rhythmical Features of Spontaneous Spoken Narrative in Mandarin Chinese

Tereza SLAMĚNÍKOVÁ
Palacký University Olomouc, Czech Republic
tereza.slamenikova@upol.cz

David UHER
Palacký University Olomouc, Czech Republic
david.uher@upol.cz

Abstract

Every oral communication is shaped by a series of closely linked factors and rhythm is one of them. The rhythmical flow in colloquial Mandarin Chinese was thoroughly described by the Czech phonetician Oldřich Švarný, who invented a unique notation system dedicated to facilitating the teaching of Chinese as a foreign language. His prosodic transcription parses speech into a two-level arrangement of prosodic units. It displays the speech rhythm as an alternation of rhythmically significant and insignificant syllables of unequal tone prominence. The paper contributes to Švarný's findings by shifting the attention from prefabricated teaching materials to an example of unrehearsed dynamic speech. Specifically, it attempts to outline the basic rhythmical features of personal narrative and confront them with the current understanding of speech rhythm in colloquial Chinese. It also demonstrates how Švarný's methodological framework can be exploited to investigate different types of spoken discourse.

Keywords: Mandarin Chinese, speech rhythm, Oldřich Švarný, spontaneous speech, personal narrative

Povzetek

Vsako ustno komunikacijo oblikuje serija tesno povezanih faktorjev in ritem je eden izmed njih. Ritmični tok pogovorne kitajščine je temeljito opisal že češki fonetik Oldřich Švarný, ki je izumil edinstven sistem notacije z namenom poučevanja kitajščine kot tujega jezika. Njegova prozodična transkripcija razčlenjuje govor v dvonivojsko urejenost prozodičnih enot. Prikazuje govorjeni ritem kot izmenjavo ritmično pomembnih in nepomembnih zlogov neenake tonske pomembnosti. Naš članek Švarnýjeve ugotovitve, zasnovane na vnaprej pripravljenih učnih materialih, preverja na primeru spontanega dinamičnega govora. V njem popišemo osnovne ritmične značilnosti osebne pripovedi in jih soočimo z obstoječim razumevanjem govornega ritma v pogovorni kitajščini. V članku tudi prikažemo, kako je moč Švarnýjev metodološki okvir uporabiti pri raziskovanju različnih vrst govornega diskurza.

Ključne besede: kitajščina, ritem govora, Oldřich Švarný, spontani govor, osebna pripoved

1   Introduction

Prosodic research indicates that colloquial[1] Mandarin Chinese is characterized by tonal assimilation and neutralization, a phenomenon already pointed out by the Chinese-American linguist Chao Yuen Ren (1892–1982) in his widely respected study *A Grammar of Spoken Chinese* (1968). Although other researchers have further developed his observations, one cannot fail to notice that their considerations are either limited to a word or, eventually, a phrasal level (e.g., Duanmu, 2007; Lin, 2007) or are primarily concerned with the issue of intonation in the tonal language (e.g., Shen, 1989; Tao, 1996). As such, they do not pay adequate attention to how tonal weakening affects the overall rhythmical segmentation of spoken texts. The findings of the Czech sinologist and phonetician Oldřich Švarný (Švarný, 1991a, 1991b, 1998, 2000; Švarný & Rusková Tang, 1991;[2] Švarný & Uher, 2014) and scholars adopting his line of research (e.g., Uher & Slaměníková, 2015; Uher, 2016; Pospěchová, 2016a, 2016b, 2022; Slaměníková, 2016, 2018; Třísková, 2017, 2019; Slaměníková & Uher, 2021), however, clearly indicate the importance of this phenomenon. To provide a basis for teaching Chinese as a foreign language,[3] Švarný developed a unique prosodic labeling system. It is accompanied by a complex theory on speech rhythm in Mandarin Chinese grounded on the empirical and instrumental examination of acoustic cues.

Simply speaking, the philosophy underlying Švarný's system is built on an observation that rhythmical features are organized in terms of categorically distinct entities (e.g., stressed and unstressed syllables) and relations (e.g., a relatively stronger or weaker tone prominence). Syllables with different properties are grouped and create different rhythmical patterns. Švarný et al. (1998, p. 8) identify two basic rhythmical sections of the so-called linear segmentation: cola (sg. colon), i.e., basic units characterized as compact

---

[1] There is a great discrepancy between the spoken (kǒuyǔ 口语) and written languages (shūmiànyǔ 书面语) in Mandarin Chinese. The term 'colloquial' refers to the first mentioned variant. Gajdoš (2011) outlines the differences between these two language variants considering choosing language data for synchronous linguistic research.

[2] To increase their accessibility, Švarný's articles written in foreign languages (including Švarný, 1991a, 1991b, 2000; Švarný & Rusková Tang, 1991) were gathered together and published as a collected reprint under the title *Oldřich Švarný: Prosodia linguae Sinensis* (Uher & Slaměníková, 2019). The edition also includes the English chapter Rhythmical Features and Prosodic Transcription of Chinese included in *Hovorová čínština v příkladech III* [Colloquial Chinese in Sentence Examples] (Švarný et al.,1998, p. 7ff).

[3] Although the situation is changing, prosodic aspects of speech are still largely ignored in the practice of L2 teaching. This seems to be a mistake, especially in cases when L1 and L2 are typologically different, which is often the case with native Slavic speakers learning East Asian languages. One researcher who pays attention to this phenomenon is Golob (2011, 2022), targeting the (dis)similarities between Slovene and Japanese.

semantic sections with intonation features separated by a pause, and segments, i.e., minimal units with a certain rhythmical flow. Their research on speech rhythm has shown that the rhythmical features are "variable according to the speed of speech, its emotional coloring, the overall length of the predicative and nominal complexes and according to the possible cases of logical and contrastive stresses" (Švarný et al., 1998, p. 7).

For understandable reasons, Švarný focused on the practical ends of the developed theory. He analyzed the prosodic interpretation of an extensive set of example sentences in learning materials for Chinese language students. The most influential is *Učební slovník jazyka čínského* [Learning Dictionary of Modern Chinese] (1998–2002, abbrev. *LDMC*). In light of this, his works systematically describe rhythmical features of read speech. The developed theory and methodology are universal, however, and can therefore be applied to any spoken Chinese discourse. Taking a personal narrative as an example, this paper shifts the attention to a speech performance that represents a continuous flow of clauses created spontaneously at the moment of recording.

2 Theoretical background and research approach

Spontaneous spoken language is often defined in opposition to written language. Using such a comparison, Miller and Weinert (1998, p. 22) provide a list of key properties describing this type of discourse as a language that is produced in real-time by people talking face-to-face, is accompanied by suprasegmental and paralinguistic information, and is constrained by the capacity of both the speaker's and hearer's short-term memory. Halliday (1989, p. 81) focuses on the way reality is represented and argues that spoken language shows phenomena as dynamic processes, whereas written language reproduces them as ready-made products. Similarly, Crystal (2000, p. 181) assigns qualities of time-boundedness, dynamism, and transience to the speech and sets them in contrast with the space-bounded, static, and permanent nature of writing. A different point of view distinguishes (written or spoken) spontaneous forms of communication that are "created on the fly" from the manufactured varieties that are "prepared, reworked, and turned into an artifact or performance" (Clark, 2014, p. 292). This article analyzes a sample of verbal communication that meets the criteria of 'spontaneity' in both of these approaches: it is a story delivered ad hoc by a speaker.

The various general properties of spontaneous spoken communication are reflected in certain linguistic features on both grammatical and lexical levels. The overall fragmented and unintegrated syntax is characterized by an

occurrence of relatively simple phrases and clausal constructions and a preference for paratactic over hypotactic patterns (Miller & Weinert, 1998, p. 22ff). The structure of clause complexes is intricate in a way that it is "able to represent not one process after another in isolation but whole configurations of processes related to each other in a number of different ways" (Halliday, 1989, p. 86). As such, the grammatical intricacy can compensate for the otherwise lower lexical density per clause (Halliday, 1989, p. 87). To help ensure messages are understandable, speakers "carefully stage" the sections with a higher amount of lexical information provided in the long strings of words (Miller & Weinert, 1998, p. 22ff). The prosodic effects are one of the main tools used for this purpose. They also have the ability to reflect elements of language that are not provided through the choice of vocabulary or encoded in the grammar (cf. e.g. Crystal, 2000, p. 179).

Despite the dominant role of conversational talk, spontaneous speech may also include numerous monologues. One of its forms is a spoken narrative, e.g., speeches in the shape of recounting an event, telling a joke, or narrating a story in dialogues (Clark, 2014, p. 303). Brazil (1993, p. 155) mentions two fundamental aspects that distinguish narratives from conversation. First, storytellers usually engage the other participants in active listenership through clearly marked openings and endings. Second, their linguistic choices are goal-oriented in that they carefully set the story's endpoint. In contrast, Brazil also emphasizes that narrative is, just like a conversation, an interactive and contextualized phenomenon. Speakers adjust their choices to the moment-to-moment needs emerging during the course of communication.

From a broader perspective, the spoken narrative is an interdisciplinary subject of study that is in the interest of several major groups of scholars, including literary theorists, psychologists, folklorists, and anthropologists, as well as sociolinguists (cf. Juzwik, 2012, p. 331ff). According to Labov (1972, p. 359), the narrative is a "method of recapitulating past experience by matching a verbal sequence of clauses to the sequence of events which (it is inferred) actually occurred." The definition implies a certain interplay between several key elements. The framework developed by Bauman (1986) identifies the event (event recounted in the narrative, i.e., narrated event, be it an actually occurred or narratively constructed event), story (i.e., narrated text), and performance (the act of telling the story including the situated context, in which the narrative is told, i.e., narrative event). A similar threefold distinction was drawn by Blum-Kulka (1993), however, she leaves the narrated event from the model and separates the performer from the act of storytelling, who may or may not be the person accountable for the story. All in all, the so-called "dimensions of narrativity" distinguished by Blum-Kulka include the 'tale,' the

'teller,' and the 'telling.' This short introduction to several basic concepts briefly indicates the complexity of spoken narrative discourse and demonstrates why it can offer researchers from different disciplines a great deal of space to target different segments of its manifold organism.

The sociolinguistic approach has been influenced by the structural approach of Labov (1972, p. 354ff; Labov & Waletzky, 1967), who, among others, defined the formal units of personal narratives and established a six-category typology of clauses constituting the structure of the narrated text, i.e., a) abstract (opening clauses capturing the point of the narrative); b) orientation (descriptive clauses setting the scene); c) complicating action (sequentially ordered narrative clauses forming the plot); d) evaluation (any markers revealing the attitude of the narrator toward the narrative); e) result or resolution (the termination of the complicating action); f) coda (clauses signaling the end of the narrative). The conversation analytic scholars shifted the attention, from elicited, interview-style narratives collected from informants, to the stories occurring in everyday communication (Thornborrow, 2012). Their interest centers around the nature of narrative events, addressing issues such as the typology of social actions in which they occur, the role that the situated context plays, the functions that the narratives fulfill, the complexity of the participant roles structure, the conditions enabling participants to gain the storytelling rights or the mechanisms underpinning the transition from non-narrative talk to storytelling (cf. Norrick, 2007; De Fina, 2008; Thornborrow, 2012). Elements observed in the process of narrative organization analyses include formulaic sequences characteristic for specific parts of the narrative (i.e., openings, closings, transition or turning points), disfluencies (e.g., false starts, restarts, self-corrections, hesitation markers or pause-fillers), discourse markers, or attention signals. In contrast to the Labovian approach, the approach is also not limited to personal narratives but examines how these aspects play out in a wide range of storytelling types (cf. Norrick, 2000, p. 135ff).

The sociologically motivated interest in storytelling significantly contributed to analyzing their language structure from which the other branches of linguistics can benefit. This paper addresses the issue of speech rhythm in colloquial Mandarin. The narrative was chosen as one of the forms that the spontaneous spoken discourse can take. As mentioned earlier, Švarný's findings are grounded on an abundant but strikingly homogeneous corpus mainly composed of prefabricated example sentences read by a Pekinese native speaker. In contrast, the theoretical works on spontaneous spoken narrative mention two opposing tendencies relating to the chunking of messages: a) creation of long strings of words (with a lower amount of information); and b) sequencing of the speech into shorter units as a result of

disfluencies. The previous research also indicates that narratives are usually composed of several types of clauses differing in the function they fulfill. It is therefore reasonable to assume that the heterogeneous composition of the narrative impacts the overall rhythmical segmentation of speech. In this paper, we attempt to depict its rhythmical characteristics emerging from a comparison with findings based on an analysis of read sentences. Apart from Švarný's (1991a, 1998, 2000) theoretical works, we base our description on analyses of the sentences listed in the *LDMC* (i.e., Švarný, 1991b; Švarný & Rusková Tang, 1991; Slaměníková, 2016, 2018). The term "*LDMC* sample" used in the analysis refers to the investigation of sentences included in entries 420–454 and 1926–1969 (Slaměníková, 2018).

3   Research corpus and methodology

The analysis was undertaken with a speech by a Pekinese woman uploaded on the *Phonemica* website, an online platform for stories told in local vernaculars (van Dam et al., 2011–2021). The storyteller (Mrs. Qin, born in 1947) is a retired doctor (Zheng, 2013) sharing her emotionally intense experience during medical practice. The story is entitled *Bǎ tǔdòur gěi mǎile cái yǒu qián shūyè* 把土豆儿给卖了才有钱输液 *Selling Potatoes to Get Money for an Infusion* (Mrs. Qin, 2012). Considering the platform's objective, it is evident that the recipient elicited the narrative event in a way similar to how a host asks a participant in a TV talk show (cf. Thornborrow, 2012, p. 58ff). The storyteller takes the floor for an extended time without needing to set the stage for the story. While speaking, she is not interrupted by the hearer. His reactions are limited to a few minimal displays of recipiency delivered in the background and one question produced partly in the overlap. What is different, however, from the talk show practices is that the narrative event does not show any signs suggesting that the hearer (unlike the talk show moderator) would be familiar with the tale before the recording. It is unclear, however, whether he elicited a talk on a specific topic from the speaker. The general conception of the *Phonemica* website, covering, e.g., nursery rhymes, humorous anecdotes, or local history (van Dam et al., 2011–2021), suggests that it might not be the case and that the speaker most likely decided to tell a story of personal experience and accomplishment. All in all, the delivery style is natural, without offering any evidence of being prepared in advance.

The story's plot comprises three events (E1, E2, E3) that are told chronologically. Each of them is introduced by a formulaic opener (E1 and E2 yǒu yí cì 'once'; E3 zài yǒu 'furthermore'). The E2 ended with a closing cliché zhèi gùshi wǒ yí bèiz yě wàngbuliǎo "I will not forget this story for the rest of

my life" (cf. Norrick, 2000, p. 49ff). It is also one of the evaluative comments through which the storyteller enhances the tellability of her story. Another one is located at the end of the E3, i.e., wǒ jiù tèbié gǎndòng "I was very moved." Apart from this, the analyzed narrative displays a set of other discourse-specific features, including the following formal verbal means:

1) In accordance with the observations made by Li (2019), the advancement of the story is highlighted by different perfective markers. One of them is the optional grammatical marker -le following a dynamic verb[4] (e.g., E1.02 lái 'come' + le; E1.23 bāo 'wrap' + le; E1.26 zǒu 'go' + le). Apart from this, the devices of the so-called situation aspect also prove themselves productive. The two most frequent include: a) resultative and directional verb complements (e.g., E1.13 pèishang 'join with'; wèiwan E1.15 'finish to feed'); and b) constructions with coverbs occurring after the main verb (e.g., E2.04 zǒudao 'walk to').

2) The discourse markers manage the flow of the narrative. Most of them may perform different functions that are not mutually exclusive and operate on several linguistic levels simultaneously (Jucker & Ziv, 1988, p. 3). In this respect, Mandarin is no exception (e.g., Liu, 2009). Using the classification established by Y. Maschler (2009, p. 22ff), they can be divided into four prototypical groups. The textual structure of the analyzed narrative is dominated by two of them, i.e., the referential and cognitive discourse markers. The sequence markers ránhòu and hòulái 'then,' result marker suǒyǐ 'therefore,' causal marker yīnwéi 'because' and non-coordination marker kěshì 'but' are the most frequent examples of the first of them. As for the second mentioned, the speaker often uses the marker jiùshì 'precisely be' to rephrase her previous words. The expressions nèi ge/ zhèi ge 'that is' also signal her thought process. Another one is the interjection wa, indicating the processing of the information as a surprising fact.

3) As expected, the discourse markers belong to the most productive lexical items of the analyzed narrative (cf. Jucker & Ziv, 1998, p. 3). As a natural consequence of telling a personal experience, the pronoun wǒ 'I' is the most frequent word in the analyzed narrative. The story's plot involves more people whose appearance is associated with high use of the third-person deixis. The speaker also often utilizes dialogue as a dynamic device to advance the plot. This, in other words, means that the narrative contains

---

[4] Similarly, M. Petrovčič (2009) argues that the so-called operator le "contributes to change in dynamics" (p. 4) and "tends to appear the most frequently in narrative sentences and further narrative passages" (p. 192). It should be noted that Petrovčič adopts the 'one-le approach' and uses the term operator le as a hypernym for both of the syntactic positions of le, i.e., a post-verbal and a sentence-final particle.

a relatively high proportion of clauses that indicate the upcoming act of speaking. Specifically, the speaker repeatedly uses the dialogue tag shuō 'say,' which represents the most frequent verb of the narrative.

The processing of the audio recording followed the procedure developed by Švarný. First, it was transcribed into Švarný's prosodic labeling system. After parsing the clauses into rhythmically relevant units, we mark each syllable's relative degree of tone prominence with respect to their significance in the rhythmical flow. We adopt the graphic marking introduced in *Učebnice čínské konverzace 2* [Textbook of Chinese Conversation 2] (Uher et al., 2016), which provides a more user-friendly version of the graphic labeling used in *LDMC*. Second, the transcript was analyzed according to the method established by Švarný. The following section provides basic guidelines for the understanding of both the prosodic transcription and the evaluation of rhythmical patterns. For more details, we advise the reader to consult the original texts written by Švarný listed in the previous section.

The prosodic transcript enclosed in Appendix 1 is, for clarity, divided into three parts, each represented by one event and supplemented by a brief plot summary.[5] Apart from this, we used the segmentation into sections as published on the *Phonemica* website. This is because the timing of the *Phonemica* made it possible to provide each section with its speaking ratio, one of the variables monitored by O. Švarný. The sections of each event are numbered chronologically (e.g., E1.01, E1.02, etc.). Speech errors resulting in meaningless strings of phonemes or rhythmically hardly identifiable units were excluded from the analysis. These expressions are struck through in the transcript. Moreover, the transcript does not record the reactions signaling the hearer's attention. Apart from the transcript, Appendix 1 also documents the rhythmical patterns of cola.

4   Analytical framework of speech rhythm in Chinese by Švarný

Švarný's[6] prosodic transcription depicts a clause complex[7] as a combination of stressed and unstressed syllables with a certain tonal prominence. The first

---

[5] This procedure was inspired by V. Štefek (2021).

[6] This chapter is based on Švarný's description of rhythmical features and prosodic transcription as provided in Švarný et al. (1998, p. 7ff); Švarný & Uher (2014, p. 9ff, 147ff).

[7] Considering the nature of the analyzed corpus, Švarný uses the term 'sentence.' In our paper, we prefer Halliday's (1989, p. 66) term 'clause complex' as an alternative to identify a sentence in the spoken language. Its component part is a clause, i.e., "the grammatical unit in which semantic constructs of different kinds are brought together and integrated into a whole."

mentioned, called ictus-bearing syllables or arses, are more prominent (in terms of the expansion of pitch range and time duration) than the immediately adjacent syllables. The other syllables are called non-ictus-bearing syllables or theses. The three types of ictus-bearing and four types of non-ictus-bearing syllables establish in all a seven-degree scale of syllable prominence hierarchy. As mentioned above, the so-called segments are the minimal units carrying a rhythmical flow. Each segment contains at least one arsis of any type, and the syllable immediately preceding or following an arsis is always the thesis.[8] The following table summarizes the types of syllables, including their graphic marking in the transcript. Appendix 2, listed at the end of the paper, provides example descriptions of syllable prominence and linear segmentation of clauses occurring in three selected sections.

Table 1: Types of syllables

Rhythmical significance	Syllable prominence		Marking
Ictus-bearing syllable, i.e., arsis (abbrev. A)	Stressed-prominence syllable*		**mǎ**
	Full-tone syllable**		mǎ
	Weakened-tone syllable		ma^3
Non-ictus-bearing syllable, i.e., thesis (abbrev. T)	Full-tone syllable		mǎ
	Weakened-tone syllable		ma$_3$/ ma***
	Void-tone syllable****	Neutralized	ma
		Toneless	

Notes to the table:

* It either represents the most prominent syllable of a colon or indicates a logical or contrastive stress, putting certain aspects of the clause in the foreground. The last ictus-bearing full-tone syllable is the most prominent in cola without any stressed-prominence syllable.

** The traditional tone marks label both ictus-bearing and non-ictus-bearing full-tone syllables. Their distinction is based on their position within a frame of one segment. The full-tone syllables, surrounded by weakened-tone syllables or void-tone syllables, are arses. The so-called acronymic rule determines the rhythmical significance in an immediate sequence of full-tone syllables: a) in a two-syllable structure, the first of them is a thesis and the second arsis; b) in a three- or four-syllable structure, arses are located at the edges, i.e., at the last and the first syllable.

---

[8] This means, in other words, that two arses can never occur in immediate proximity to one other in the same segment.

*** Non-ictus-bearing weakened-tone syllables, located in the middle or at the end of a segment, are marked by a number in a lower index. Those at the beginning of a segment are transcribed without marking, e.g., ma. The fact that the first syllable of a segment is always a bearer of at least a weakened prominence eliminates the confusion with a void-tone syllable.

A rhythmical flow arises by joining syllables together. There are two basic types of connections: a) a descending sequence of two or more syllables with the arsis resting on the first of them; and b) an ascending sequence of two or more syllables with the arsis at the end. Some syllables remain single and join as proclitics or enclitics of the already existing descending or ascending sequences. Alternatively, they behave as the so-called odd ictus-bearing syllables. The two types of sequences and the odd ictus-bearing syllable represent the basic structural components of the rhythmical flow. Segments with a simple structure contain one, whereas segments with a more complex structure combine two or more of these items. In the transcript, hyphens connect syllables within a segment in cases where the syllables are not part of one word.

Cola, the highest units of the rhythmical flow, can be formed by one or more segments. The colon boundaries are labeled with punctuation marks. In the transcript provided in Appendix 1, we add a comma in cases when the original transcription does not depict the boundaries of cola. In other cases, we keep the original marks provided on the *Phonemica* website when they do not interfere with the prosodic segmentation of the audio recording.

Apart from the transcription, Švarný established the metalinguistic apparatus used for the quantitative analysis of rhythmical patterns of cola. These are as follows: / an ascending sequence; /° an ascending sequence followed by an enclitic syllable(s); \ a descending sequence; °\ a descending sequence preceded by a proclitic syllable; | an odd ictus-bearing syllable; V a trisyllabic two-arsis acronymic sequence of the so-called 'cretic' type (i.e., a combination of a descending sequence and an odd syllable; or less often, a combination of an odd syllable and an ascending sequence). In Appendix 1, we add these marks in a separate column on the right side of the transcribed text. The cola of each section are numbered, e.g., c1, c2, etc. The letters A–D preceding the description of rhythm indicate the type of colon as described below.

Švarný's classification of cola takes into account two aspects. The first of them is the number of rhythmical sequences. Considering this, the category of one-sequence cola is separated (hereinafter referred to as type A). More

than one sequence cola are then classified based on the position of arses into:

- cola without an immediate sequence of two arses (type B)

- cola with an immediate sequence of two arses (type C)[9]

- cola with an enclitic or proclitic syllable preventing the immediate sequence of two arses (type D)

As can be seen, the model accentuates whether two ictuses can be found in immediate contact next to each other. By its nature, this can happen strictly at the edges of two neighboring segments, represented by the type C cola in the model. The contact of two arses is indicated by the symbol +.

All types include a spectrum of rhythmical patterns, as will be shown in the analysis. At this point, the so-called acronymic rhythm principle must be explained. The simple acronymic structure of the so-called 'cretic' type comprises three syllables. Its arrangement has been described above. A slightly more complex structure combines a descending and an ascending sequence \ /, i.e., the so-called 'choriamb' type. Both of these patterns have two arses, the first located on the initial syllable and the second on the last syllable. Apart from these, three- and more arses acronymic structures are also quite common. They are created through the repetition of the descending sequences or ascending sequences. The core of these structures can be of both a cretic or a choriamb type (i.e., three-arses acronymic structures \V, \\ /, V/, \ //; four-arses acronymic structures \\V, \\\ /, V//, \ ///, \V/, \\ //, etc.).

5 Analysis and discussion of the findings

5.1 Speaking rate

The duration of the analyzed narrative is 336 sec. Compared to previous research samples, the speech of the storyteller differs in one crucial respect: the speaking rate varies widely during her speech. The up to now most extensive corpus of approximately 16,000 transcribed sentences collected in *LDMC* shows a certain acceleration of the speech tempo between the initial and the final part of the large corpus (Švarný, 1991b, p. 132). The complexes of sentences located in particular parts manifest, however, a high degree of homogeneity. In contrast, the speech analyzed in this paper is characterized

---

[9] An immediate sequence of arses occurs when an ascending sequence of the first segment is immediately followed by a descending sequence of the second segment, including the ascending and descending sequences in any acronymic structure.

by significant tempo differences within a short amount of time that are associated with the speaker's emotional involvement as the story evolves. She manipulates the speech tempo to contextualize the clauses in a certain way. Specifically, her speaking rate varies between 1.9 to 5.7 syl/sec through the different sections.

Since Švarný et al. (1998, p. 7) identify the speech tempo as the variable influencing the rhythmical features, we pay attention to it while analyzing individual aspects and, once it is relevant, draw a comparison between slower and faster sections. In order to be able to determine this, we set the average speaking rate of 3.5 syl/sec as the central point and focus on the linear segmentation of the sections with a rate slower than 3.0 syl/sec and faster than 4.0 syl/sec. The first mentioned includes 53 cola composed of 238 syllables, and the second mentioned 58 cola composed of 436 syllables. In the transcript, the slower sections are presented with the background color blue and the faster with orange.

At this point, it is important to comment on the logic behind the speech segmentation into sections provided on the *Phonemica* website. They basically divide the speech into small information units. Their length ranges from one to six clauses. Having in mind the Labovians classification, multi-clauses sections are composed of both a series of one-type clauses or a combination of different types of clauses. A previous study on the relationship between speech rate and typology of clauses indicates that there is "no statistically significant effect" between these two variables on the global level (Oliviera, 2012, p. 640). The analyzed narrative seems to support this finding, as it can be determined from the inconsistent speech tempo of the sections containing one type clause. In light of this, the fact that a section might be composed of different types of clauses is not considered while observing the fluctuation in speaking rate.

5.2   Linear segmentation

As mentioned above, Švarný defines two rhythmical units of linear segmentation. Compact semantic sections with intonation features, i.e., cola, are composed of minimal rhythmical units, i.e., segments. When observing the different parts of the *LDMC* corpus, Švarný et al. (1988, p. 24) found out that while the length of segments is considerably dependent on the speech tempo, the length of cola does not seem to be influenced in this way. The syntactic complexity of the sentences can be, to a certain measure, significant in this way: the length of cola in longer sentences tends to be slightly higher.

Švarný also calculated the average length of both linear units: 2.5 to 4.5 syllables per segment and 6 to 7 syllables per colon.

The analyzed sample contains 197 cola, 367 segments, and 1,168 syllables. Considering the average length of segments, the number of 3.2 syllables places the analyzed narrative in the middle part of the spectrum determined by Švarný et al. (1998, p. 8). Just like it was reported in previous research, most segments contain two to four syllables (cf. Švarný et al., 1998). Specifically, two- and three-syllable segments account for almost 30% each, and four-syllable segments for almost 24% in the narrative. In addition, the average length of cola, i.e., 1.9 segments and 5.9 syllables, is essentially consistent with previous reports. What is different is the distribution of cola with a certain number of segments. About one-half of the *LDMC* sample (Slaměníková, 2018, p. 157) comprises two-segment cola, whereas a similar proportion belongs to one-segment cola in the narrative. In addition, the lower occurrence of two-segment cola is compensated for by the higher occurrence of three- and more-segment cola.

This change in proportion arises from the characteristic nature of the spontaneous spoken language. Disfluencies such as false starts with abrupt cut-offs, pause-fillers, or discourse markers break the speech flow into shorter cola. Phrases and clauses advancing the plot of the story tend to, in contrast, rhythmically speaking, connect together and form longer cola. The impact of spontaneity is also reflected in the relation between the average length of cola and the speaking rate (see Table 2). The average length of 4.5 syllables per colon is significantly lower than the value of 6 to 7 syllables determined by Švarný et al. (1998, p. 24), and, more importantly, it is inconsistent with the assumption that the colon is a linear unit whose average length is not to be influenced by speech tempo. Against expectation, the average length of segments that was reported to be directly proportional to the speaking rate shows insignificant variations in average numbers. A certain difference, however, can be observed: the occurrence of two- to four-syllable segments in faster sections is reduced in favor of five-syllable segments.

Table 2: Linear segmentation

Sample	The average length of segments	The average length of cola		
Whole narrative	3.2 syllables	5.9 syllables	/	1.9 segments
Slower sections (< 3,0 syl/sec)	3.1 syllables	4.5 syllables	/	1.5 segments
Faster sections (> 4,0 syl/sec)	3.3 syllables	7.5 syllables	/	2.3 segments

5.3 Syllable prominence

Table 3 summarizes the occurrence of different types of syllables in the analyzed narrative. Compared to the *LDMC* sample (Slaměníková, 2018, p. 160), the narrative contains more weakened-tone and void-tone syllables. The most noteworthy difference is the amount of stressed-prominence syllables. As far as we can say, their occurrence in the different parts of the *LDMC* sample varies between 12.1 to 19.1% (cf. Slaměníková, 2016, p. 80; Slaměníková, 2018, p. 160). The reason for their lower occurrence in the narrative can be explained as follows. While recording the isolated sentences of *LDMC*, the speaker needed to emphasize a piece of certain information in each sentence or even a colon. In contrast, the speaker telling the continuing story places the stress with the aim to contrast entities or focus attention on entities within more complex textual units.

Table 3: Title of this table

Syllable prominence	Occurrence
Stressed-prominence syllable	4.3%
Full-tone syllable – arsis	28.9%
Full-tone syllable – thesis	5.3%
Weakened-tone syllable – arsis	12.8%
Weakened-tone syllable – thesis	8.8%
Void-tone syllable	39.9%

In agreement with Švarný's observations (Švarný & Rusková Tang, 1991, p. 241ff), the changing speaking rate of the narrative affects the proportion of syllables with different rhythmical significance. Table 4 demonstrates that the

occurrence of arses in the faster sections is about 4% lower than in the slower sections. As Švarný explains, clauses with faster speech tempo are rhythmically organized into larger groups of syllables within which some of the ictus-bearing units tend to be weakened or even neutralized and, thus, become theses.

Table 4: Occurrence of arses

Syllable prominence	Occurrence
Whole narrative	46.0%
Slower sections (< 3,0 syl/sec)	47.9%
Faster sections (> 4,0 syl/sec)	43.8%

5.4   Rhythmus of segments

The rhythmus of the segments relies on the number and position of the ictus-bearing syllables. In the analyzed speech, more than one-half of the segments contain one arsis. As can be seen in Table 5, the occurrence of the ascending sequence is almost twice as low as the occurrence of the descending sequence, which represents the most frequent rhythmical pattern overall. Descending sequences, or, to be specific, the combinations of the two of them, also play an essential role in the two-ictus-bearing segments. The acronymic structure prevails, however, in this type of segment and embodies the second most frequent pattern. Almost all tree-ictus segments are acronymic; only one segment comprises three descending sequences.

Table 5: Rhythmus of segments

Arses	Occurrence	Rhythmical pattern	Occurrence
One	57.6%	Odd syllable	3.8%
		Descending sequence	35.3%
		Ascending sequence	17.7%
		Others	0.8 %
Two	38.9%	Acronymic structure	24.5%
		Two descending sequences	13.0%
		Others	1.4%
Three	3.5%	Acronymic structure	3.3%
		Three descending sequences	0.2%

Generally speaking, the occurrence of the rhythmical patterns does not deviate from our current understanding of the matter. The numbers of patterns range more or less in the middle between their occurrence in the slower and faster speech in the *LDMC* sample (cf. Slaměníková, 2018, p. 166). This does not, however, fully apply to one-arsis segments, in which case the occurrence of odd syllables and ascending sequences is slightly lower.

It should be further pointed out that both descending and ascending sequences are far less often composed of lexical compounds. The relatively lower lexical density of spontaneous speaking style is the reason for a higher occurrence of rhythmically joined groups of monosyllabic words. As was described by Švarný, a significant number of these monosyllabic words tie in with the rhythmical sequences as clitics. The most common proclitics, preceding a descending sequence, include singular personal pronouns and multifunctional adverbs dōu and jiù. The attributive marker d and sentence-final particles le and ne are the most frequent enclitics following the ascending sequences (cf. Švarný et al., 1998, p. 39).

Finally, a note has to be added concerning the lexical structure of the segments. The use of discourse-specific elements can be, of course, observed all over the narrative. It is primarily, however, the group of two-syllable segments, in which case they occur with a significant concentration. Four categories of these elements can be distinguished. The first of them is related to the first person point of view and consists of two combinations: a) pronoun wǒ + verb shuō 'I said' (e.g., E2.21); b) pronoun wǒ + adverb jiù 'so I' (e.g., E1.11). The second category includes temporal nouns and conjunctions that indicate the sequence and causality, e.g., hòulai (e.g., E2.22). The third category is the connection of the demonstrative nèi with the general measure word ge carrying the meaning 'that,' used by the speaker both as a reference device (e.g., E2.02) or as a pause-filler for lexical or content search (e.g., E1.03). The last one is the discourse marker composed of the adverb jiù and the verb shì, together carrying the meaning 'precisely be' and used by the speaker as a rephrasing tool (e.g., E2.27).

5.5   Rhythmus of cola

The analysis has shown that most of the one-sequence cola (type A) are characterized by a descending rhythm. Consistent with previous findings, cola without an immediate sequence of two arses (type B) show a preference for two patterns, i.e., the repetition of descending sequences and different forms of acronymic structures. Generally speaking, type C and D cola share the same patterns. The only difference lies in that the first mentioned are

characterized by an occurrence of an immediate sequence of two arses, whereas the second mentioned contain a clitic preventing such a contact. From the variety of patterns these two types can be organized into, the analyzed narrative uses only one more often, i.e., a combination of two acronymic structures. The following table provides the relative occurrence of each pattern and lists some examples occurring in the narrative in the case of the most frequent poly-sequence patterns.

Table 6: Rhythmus of cola

Type	Rhythmical pattern	Occurrence
A	An odd syllable |	5.1%
	A descending sequence \	14.7%
	An ascending sequence /	7.1%
	A three-syllable acronymic structure V	5.1%
B	An acronymic structure e.g., \ /, \\ /, \V, \ //, V//, \\\V° etc.	23.9%
	Two or more descending sequences e.g., \\, °\\, \\\, \\\\ etc.	17.8%
	Others	2.5%
C	A combination of two acronymic structures e.g., \ /+\ /, V+\ /, V+\\ /, \\V+V etc.	5.1%
	Others	7.6%
D	A combination of two acronymic structures e.g., V°\ /, \ /°|/, V°\V, \ //°\\ /	3.6%
	Others	7.6%

Compared to the *LDMC* sample (Slaměníková, 2018, p. 169), the narrative is characterized by an almost twice as high number of the rhythmically simple type A cola. Considering the above-discussed nature of the dependency between the average length of cola and the speaking rate, it is logical that this type is especially productive in the slower sections. It also has to be mentioned that, just as with the case of the segments, the rhythmically most simple cola are often formed by the referential and cognitive discourse markers, occasionally joined with a personal pronoun or sentence-final particle, e.g., E2.27 hòulai, E3.10 suǒyi, E1.22 ránhòu-ne, E2.07 yīnwei-tā, E1.26 jieguǒ-ne, E1.05 jiu⁴shi. Similarly, the most frequent pause-filler also tends to remain rhythmically isolated, e.g., E2.10 nèi-ge.

Naturally, all the types of more-sequences cola can be found more often in the faster sections, as seen in the table. Just as in the case of the *LDMC* sample (Slaměníková, 2018, p. 168), they are dominated by type B cola without an immediate sequence of two ictuses. As for the rhythmical patterns, it is worth mentioning that acronymic structures, consisting of at least one descending and one ascending sequence, represent the most frequent pattern in all the samples. The second most common pattern is of a descending nature in both rate-related samples, however, its form reflects the different complexity of slower and faster cola. The first mentioned bear more often one ictus, i.e., it is composed of one descending sequence (type A), while the second mentioned combines two descending sequences, each with its ictus (type B).

Table 7: Types of cola

Syllable prominence	Type A cola	Type B cola	Type C cola	Type D cola
Whole narrative	32.0%	44.2%	12.7%	11.2%
Slower sections (< 3,0 syl/sec)	45.3%	39.6%	11.3%	3.8%
Faster sections (> 4,0 syl/sec)	17.2%	51.7%	17.2%	13.8%

6   Conclusion

In 2000, O. Švarný (p. 149) argued that the prosodic features of today's Mandarin Chinese are "more and more encroaching on the lexically distinctive function of 'tones.'" He supported this statement with an extensive analysis of transcribed sentences through which he demonstrated how unequal syllable prominence generates specific rhythmical patterns. This paper builds upon the methodological and theoretical framework established by Švarný and extends his description of prosodic features of modern colloquial Chinese in a way that targets a different type of spoken discourse.

The analyzed narrative differs from Švarný's corpus in several main aspects. First, it is determined by the different circumstances of their origin. Švarný's corpus collects oral manifestations of isolated sentences. The fact that the speaker was encouraged to perform them as naturally as possible does not change the essential characteristics of a text being read out loud. In contrast, the available information about the analyzed storyteller's speech suggests that she was invited to choose a topic of her preference and use her

own words to deliver a speech. It is evident that each of the corpus is composed of a different content. Švarný's corpus collects illustrative sentence examples, each of which is tied with a specific agenda, i.e., explaining the grammatical unit in question. The narrative is a variable flow of syntactically not necessarily correct clauses that, as a whole, aim at achieving a particular communicative goal.

This article is a pilot study that attempted to depict general tendencies of speech rhythm in a spontaneous spoken narrative in colloquial Mandarin without a claim to exhaust the topic. A large amount of data is necessary to yield more conclusive findings. In any case, some preliminary conclusions can be drawn based on the comparison with the read speech. As expected, the narrative is composed of parts that differ from each other. The storyteller chunks the messages in constituents of variable length. We can also observe a significant fluctuation in speaking rate differences occurring within a short amount of time. In accordance with Švarný's et al. (1998, p. 7) observation that speech tempo is a variable influencing the rhythmical features, we have detected differences between slower and faster speech sections, especially on the level of linear segmentation, number of ictus-bearing units (arses) and typology of cola.

The analysis establishes two findings that seem germane to the issue of speech rhythm in spontaneous Mandarin Chinese communication. First, slower sections show a lower average length of cola, which are, according to Švarný, supposed to resist changes in speech tempo. This is not surprising, however, since the greater frequency of disfluencies results in shorter syntactic units and, consequently, shorter prosodic units. Moreover, discourse-specific markers, in particular, are often separated by longer pauses indicating the borderline between cola. Second, the narrative contains a considerably lower amount of stressed-prominence syllables. In this respect, it should be explained that their use in the *LDMC* sample seems to be a strong habit of routine: the speaker systematically emphasizes at least one syllable in each sentence example. In contrast, their use in the narrative is more selective: the storyteller highlights essential pieces of information in the context of the whole story plot, often occurring in longer strings of words.

On a global level, the rhythmical organization of the spontaneous narrative is similar to the read sentences of the *LDMC* sample. This feature is stronger in the case of the smaller linear units, i.e., segments. We can observe a high degree of regularity in the proportion of particular rhythmical patterns within the scope of the two samples. The issue of cola is more complicated. The ratio of the four basis rhythmical types in the narrative is greatly influenced by the heterogeneous nature of the syntactic structures

comprising both fragmented and highly elaborated clauses. Moreover, as has already been mentioned above, the occurrence of the first one is higher in the slower sections and of the latter in the faster section. The arrangement of rhythmical sequences within a certain type of cola shows, however, a similar proportional distribution. As far as we can conclude from the present data, the rhythmical features in general seem to be governed by universal principles, regardless of the speech mode in question. This is not to say, however, that there is no mutual connection between the spoken genre and the speech rhythm. Higher constituents of linear segmentation, i.e., cola, are the units that seem to exhibit certain discourse-specific rhythmical features that are worthy of attention in future research.

This paper opens a wide area of research in the field of speech rhythm in Mandarin Chinese. From the micro perspective, the present analysis, for example, points out the specific rhythmical function of discourse markers that is worthy of further investigation. Apart from this, examining the relationship between clause and colon might bring interesting findings regarding the position of their mutual boundaries. From the macro perspective, spoken communication offers numerous spoken genres. The scope of research can be expanded on finding the interrelationships between speech rhythm and various criteria used to classify them, including, but not limited to, the number of participants, degree of preparedness, communicative purpose, level of structural and context complexity, selection of lexical units, degree of elaboration of syntactic structures and occurrence of expressive elements.

Acknowledgments

References

Bauman, R. (1986). *Story, Performance and Event: Contextual Studies of Oral Narrative*. Cambridge: Cambridge University Press.

Blum-Kulka, S. (1993). "You gotta know how to tell a story": Telling, Tales, and Tellers in American and Israeli Narrative Events at Dinner. *Language in Society*, *22*(3), 361–402.

Brazil, D. (1993). Telling tales. In G. Fox, M. Hoey, & J. M. Sinclair (Eds.), *Techniques of Description* (pp. 154–169). London: Routledge.

Chao, Y. R. (1968). *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

Clark, H. H. (2014). Spontaneous Discourse. In M. Goldrick, V. Ferreira, & M. Miazzo (Eds.), *The Oxford Handbook of Language Production* (pp. 292–307). Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199735471.013.023

Crystal, D. (2000). *The Cambridge Encyclopedia of Language*. Beijing: Foreign Language Teaching and Research Press.

De Fina, A., & Georgopoulou, A. (2008). Introduction: Narrative Analysis in the Shift from Texts to Practices. *Text & Talk*, *28*(3), 275–281. https://doi.org/10.1515/TEXT.2008.013

Duanmu, S. (2007). *The Phonology of Standard Chinese*. New York: Oxford University Press.

Gajdoš, Ľ. (2011). The Discrepancy between Spoken and Written Chinese – Methodological Notes on Linguistics. *Studia Orientalia Slovaca*, *10*(1), 155–159.

Golob, N. (2011). Acoustic Prosodic Parameters in Japanese and Slovene: Accent and Intonation. *Acta Linguistica Asiatica*, *1*(3), 25–44. https://doi.org/10.4312/ala.1.3.25-44

Golob, N. (2022). Word-prosodic Typology and the Traps of False Similarity: Japanese and Slovene. *Acta Linguistica Asiatica*, *12*(2), 115–123. https://doi.org/10.4312/ala.12.2.115-123

Halliday, M. A. K. (1989). *Spoken and Written Language*. Oxford: Oxford University Press.

Jucker, A. H., & Ziv, Y. (1998). Discourse Markers: Introduction. In A. H. Jucker & Y. Ziv (Eds.), *Discourse Markers* (pp. 1–12). Amsterdam: John Benjamin Publishing. https://doi.org/10.1075/pbns.57

Juzwik, M. M. (2012). Spoken Narrative. In J. P. Gee, & M. Handford (Eds.), *The Routledge Handbook of Discourse Analysis* (pp. 326–341). London: Routledge.

Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov & Waletzky. (1967). Narrative Analysis: Oral Versions of Personal Experience. In J. Helm (Eds.), *Essays on the Verbal and Visual Arts* (pp. 12–44). Seattle: University of Washington Press.

Li, W. (2019). Aspect in Chinese discourse. In Ch. Shei (Ed.), *The Routledge Handbook of Chinese Discourse Analysis* (pp. 160–174). London: Routledge.

Lin, Y.-H. (2007). *The Sounds of Chinese*. Cambridge: Cambridge University Press.

Liu, B. (2009). Chinese Discourse Markers in Oral Speech of Mainland Mandarin Speakers. In Y. Xiao (Eds.), *Proceedings of the 21st North American Conference on Chinese Linguistics* (Vol. 2, pp. 358–374.). Smithfield, Rhode Island: Bryant University.

Maschler, Y. (2009). *Metalanguage in Interaction: Hebrew Discourse Markers*. Amsterdam: John Benjamins Publishing Company.

Miller, J., & Weinert, R. (1998). *Spontaneous Spoken Language: Syntax and Discourse*. Oxford: Clarendon Press.

Norrick, N. R. (2000). *Conversational Narrative: Storytelling in Everyday Talk*. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/cilt.203

Norrick, N. R. (2007). Conversational Storytelling. In D. Herman (Eds.), *The Cambridge Companion to Narrative* (pp. 127–141). Cambridge: Cambridge University Press. https://doi.org/10.1017/CCOL0521856965.009

Oliviera, M. (2012). A Study on Speech Rate as a Prosodic Feature in Spontaneous Narrative. *Alfa: Revista de Linguística*, *56*(2), 623–651. DOI: https://doi.org/10.1590/S1981-57942012000200012

Petrovčič, M. (2009). *Operator LE in Chinese: Complexity within Simplicity and Simplicity within Complexity*. VDM Verlag. Retrieved July 17, 2023, from https://www.researchgate.net/publication/325828254_Operator_Le_in_Chinese

Pospěchová, Z. (2016a). *Tónové změny v moderní čínštiny ze sociolingvistické perspektivy*. Olomouc: Univerzita Palackého v Olomouci.

Pospěchová, Z. (2016b). Prosodic Transcription of Standard Chinese and its Use in Teaching. *Acta Linguistica Asiatica*, *6*(1), 35–45. https://doi.org/10.4312/ala.6.1.35-45

Pospěchová, Z. (2022). Gendering Prosody in Communication: The Case of Standard Chinese. *Journal of Asian and Pacific Communication*, *32*(1), 133–153). http://dx.doi.org/10.1075/japc.00087.pos

[Mrs.] Qin 秦女士 (2012). Ba tudou gei maile cai you qian shuye 把土豆儿给卖了才有钱输液 In K. P. van Dam, S. Hansen & J. Qi (Eds.), *Phonemica*. Retrieved October 21, 2017, from http://phonemica.net/entry?mid=543748332bd553180882ab53&story=0

Shen, X. S. (1989). *The Prosody of Mandarin Chinese*. Berkeley, Los Angeles, Oxford: University of California Press.

Slaměníková, T. (2016). Rytmické členění příkladových vět Učebního slovníku jazyka čínského (hesla 1–50). *Dálný východ*, *VI*(2), 64–82.

Slaměníková, T. (2018). An Analysis of the Correlation between Speech Tempo and Rhythmical Segmentation in Standard Chinese. *Studia Orientalia Slovaca*, *17*(2), 151–169.

Slaměníková, T., & Uher, D. (2021). Prozodická charakteristika textu: Připravenost a spontánnost mluveného projevu. *Dálný východ*, *XI*(1), pp. 104–122.

Štefek, V. (2021). Prozodická charakteristika textu: Neformální promluva filmové postavy. *Dálný východ*, *XI*(1), 49–81.

Švarný, O. (1991a). The Functioning of the Prosodic Features in Chinese (Pekinese). *Archiv orientální*, *59*(2), 208–216.

Švarný, O. (1991b). Rhythmical Features of Spoken Chinese: Quantitative and Grammatical Analysis. *Rocznik orientalistyczny*, *XLVII*(2), 132–137.

Švarný, O., et al. (1998). *Hovorová čínština v příkladech III*. Olomouc: Vydavatelství Univerzity Palackého.

Švarný, O. (1998–2002). *Učební slovník jazyka čínského I–IV*. Olomouc: Univerzita Palackého. (abbrev. *LDMC*)

Švarný, O. (2000). Prosodical Transcription of Modern Chinese: Experimental Research and Teaching Practice. In Z. Palková & H.-W. Wodarz (Eds.), *Papers in Phonetics and Speech Processing* (pp. 149–159). Frankfurt am Main: Hector Verlag.

Švarný, O., & Rusková Tang, Y. (1991). Prosodic Features in Chinese (Pekinese): Prosodic Transcription and Statistical Tables. *Archiv orientální*, *59*(3), 234–254.

Švarný, O., & Uher, D. (2014). *Prozodická gramatika čínštiny*. Olomouc: Univerzita Palackého v Olomouci.

Tao, H. (1996). *Units in Mandarin Conversation: Prosody, Discourse, and Grammar*. Amsterdam: John Benjamins Publishing Company.

Thornborrow, J. (2012). Narrative Analysis. In J. P. Gee & M. Handford (Eds.), *The Routledge Handbook of Discourse Analysis* (pp. 51–65). London: Routledge.

Třísková, H. (2017). De-stressed in Mandarin: Clitics, Cliticoids and Phonetic Chunks. In Kecskes, I., & Sun, Ch. (Eds), *Key Issues in Chinese as a Second Language Research* (pp. 29–56). New York: Routledge.

Třísková, H. (2019). Is the Glass Half-full, or Half-empty? The Alternative Concept of Stress in Mandarin Chinese. *Studies in Prosodic Grammar*, *2*(4), 64–105.

Uher, D. (2016). The Rhythmic Structure of Sentences in the Textbook of Chinese Conversation. *China & the World Cultural Exchange*, *5*, 25–28.

Uher, D., Jin, X., & Slaměníková, T. (2016). *Učebnice čínské konverzace 2*. Voznice: Leda.

Uher, D., & Slaměníková, T. (Eds). (2019). *Oldřich Švarný: Prosodia linguae Sinensis*. Olomouc: Univerzita Palackého v Olomouci.

Uher, D., & Slaměníková, T. (2015). Prozodická analýza monologu. *Dálný východ*, *V*(2), pp. 104–115.

van Dam, K. P., Hansen, S., & Qi, J. (Eds.). (2011–2021). *Phonemica*. https://phonemica.net/about, *Phonemica*. Retrieved July 19, 2023, from https://phonemica.net/

Zheng, R. 郑汝可. (2013). Meigou xiaohou ban wanzhang shoulu Zhongguo fanyan 美国小伙办网站收录中国方言. *Hanwang – Changjiang ribao 汉网-长江日报*. Retrieved from July 17, 2023, http://hb.ifeng.com/news/fygc/detail_2013_07/12/986775_0.shtml

Appendix 1: Prosodic transcript of the analyzed narrative

List of abbreviations: c = colon; A = type A colon; B = type B colon; C = type C colon; D = type D colon

List of graphic symbols: / ascending sequence; /° ascending sequence followed by an enclitic syllable(s); \ descending sequence; °\ descending sequence preceded by a proclitic syllable; | odd ictus-bearing syllable; V trisyllabic two-arsis acronymic sequence; + contact of two arses in type C cola; ~~strikethrough text~~ indicates rhythmically hardly identifiable units that were excluded from the analysis

1. Event: Mrs. Qin tells about a farmer who came to her with a sick child suffering from severe diarrhea and dehydration. His father went to sell potatoes to earn money to pay for an infusion in the hospital. The child's health was already bad then, and the storyteller gave him some home-made medicine. Unfortunately, the father did not return and the child was not getting better. His mother decided to take it back home to the mountains because of the local tradition that people passing away outside cannot be buried on the village's ground. Before leaving, the storyteller gave the mother some more of the home-made medicine to take later; since there was no way to contact the family then, she did not know what happened with the child next.

Section	Syl/sec	Transcript	Rhythm
E1.01	2.6	you³-yi-cì, zǎochen-qi₃lai², wo-hái-mei qǐchuáng,	(c1) A: V (c2) B: \ / (c3) B: °\ /
E1.02	3.5	yi²-ge-nóngmin bàozh-yi-xiǎohái jiu-láile. ta¹-yao-kànbìng.	(c1) D: \\\ /°\ (c2) B: \ /
E1.03	3.3	wo³-yi-kàn nei-xiǎoháir jiushì, nèi-ge, lā-dùzi, tuōshuǐ.	(c1) B: V// (c2) A: \ (c3) A: °\ (c4) A: /
E1.04	4.6	ta¹-d-yǎnjing yǐjing bìbushàng-le, jiu⁴shi yǎnjing-bìshang-dou, zhōngjian-dou¹-you-fèngr.	(c1) B: \\\V° (c2) B: \\\ (c3) B: \V
E1.05	3.5	yǎnjing dou¹-shi-kōu-d, jiu⁴-shi, ránhou, ne, jiu-zhèi-yangr jiu-zhěnduàn nei-xiǎoháir dou-yǐjing-shi sān-du-tuōshuǐ-le.	(c1) B: \V° (c2) B: \ (c3) A: \ (c4) A:	(c5) D: °\ //°\\ /°
E1.06	2.9	tuōshui-yihòu, ta-yīnggai-shū-yè,	(c1) B: \ / (c2) B: °\ /
E1.07	2.9	kěshì-ne, yīyuan-li³tou ni-**méi**-you-qian² mǎi-yào, ni³-gei-xiǎoháir shūbuliao-yè.	(c1) A: /° (c2) B: \\°V/ (c3) C: \ /+/
E1.08	2.8	ta-bà-ne, qù, dao⁴-nei-ge jiēshang-qu yao⁴-ba-tǔdòur gei-**mài**le cai²-you-qián.	(c1) A: °\ (c2) A:	(c3) D: \\\ /°\V
E1.09	4.6	náhui-qián-lai cai²-neng-mǎi-**yào** gei-xiǎohair-zhìshang.	(c1) D: \\\ /°\\
E1.10	3.6	hou⁴lai-wǒ kànzh-zhei-xiao₃hair², wo-shuō zhei-zěnme-bàn? ta¹-bu-tù, zhi³-shi lā-d te⁴bie-lihai.	(c1) C: V+\ / (c2) D: /°V (c3) A: V (c4) B: \\\\
E1.11	3.0	zhème-yangr⁴ wo³-jiu gei³-ta-yòng **wǒ**-d, e, táng, yán, he-nei⁴-ge sūdafén.	(c1) C: V+\V+\ (c2) A:	(c3) A:	(c4) A:	(c5) B: °\V
E1.12	4.2	zhei-**sān**-yangr-dōngxi, pèizai-yìqǐ, jiu-xiàng nei⁴-ge shūyè-d nei⁴-ge, wo-zhīdao-d nei⁴-ge-bǐli.	(c1) B: °\\ (c2) B: \ / (c3) C: /+\ /°\ (c4) B: °\\ /
E1.13	3.8	ránhou nei⁴-ge, wo³-jiu-gei-ta pèishang-shuǐ.	(c1) B: \\ (c2) B: \V
E1.14	5.1	wo³-gei-ta-yi²-ge-wǎnr, gei-ta¹-yi-xiǎo-sháor, wo-shuō ni-**yì**dianr-yìdianr manmānr-wèi, manmānr-wèi-ta.	(c1) B: \ / (c2) B: °\ / (c3) D: /°\\ (c4) B: / \
E1.15	2.7	wèiwanle-yihou⁴-ne, ta-bà, hai²-mei-mǎihui, hái-mei, hái-mei na²hui-qián-lai.	(c1) B: \ /° (c2) A: / (c3) B: \\ (c4) A: \ (c5) B: \\\
E1.16	4.4	ta-**mā**-ne jiu-juéde zhei-xiǎoháir, shuō, dàifu, wo³men-bu-zhìle, shuō zhei-xiaohair² bù-xíng-le.	(c1) B: °\°\ / (c2) A:	(c3) A: \ (c4) B: \\ (c5) B:	//°
E1.17	2.9	hen-xiǎo, ~~yi liang sui~~? yì-liang-suì — ~~liang~~ — liǎng-sui-zuoyou⁴.	(c1) A: / (c2) A: V (c3) B: \ /
E1.18	5.5	ta-bàozh-xiaohair². ta¹men-**dāng**di-d fēngsu-ne jiu⁴shi-shuo zhei-rén bù-neng si³zai-wàibianr yao⁴-sizai **jiā**-libianr.	(c1) B: °\ / (c2) C: \\\\ /+\\\\\
E1.19	4.7	ta¹-xiang-neng-máizai tāmen-d dāngdi-d fěn-di-li³bianr.	(c1) B: \\\\\
E1.20	4.3	yao³shi-sǐzai **wài**bianr-d zhei⁴-xie, rén-ne, jiu⁴-bu-neng jiu-huí-cūn, jiu-zhi³-neng zai-cūn-**wài**bianr-d dìshang-mai²le.	(c1) B: \\\\ (c2) A: \ (c3) B: \ / (c4) B: °\ / \\
E1.21	3.5	suo³yi tāmen-d-ren², bu-zhì-dao zuihou⁴.	(c1) B: \\ / (c2) B: °\ /
E1.22	3.3	ránhòu-ne, ta¹-jiu-**bào**zh zhè-xiaoháir jiu⁴-yao-zǒu.	(c1) A: /° (c2) C: \\V+V
E1.23	3.2	ranhòu-ne, wo³-jiu gei-tā-**bāo**le táng he-yán, zhei⁴-xie ji-yàngr-dōngxi.	(c1) A: /° (c2) D: \ /°|/ (c3) B: \ / \

E1.24	4.1	wo³-rang-tā, wo-shuo¹ jiā³ru-huíjiā xiaohair²-hai-kěyi, ~~wo wo jiu~~ ~~wo~~ wo³-rang-ta-chī zhei-yào.	(c1) A: V (c2) C: /+\ //°\ (c3) B: \ //
E1.25	4.9	ranhòu ba-wo₃-zìjǐ-d shi-huánglian**sù**-ma shi-shen²me-dōngxi, wo³-zai-gěi-ta yi²-pianr-liǎng-pianr₄, yin¹wei-xiaoháir hen-xiǎo.	(c1) D: //°°V°°\\ (c2) B: \\\\ (c3) B: \ //
E1.26	2.9	jieguǒ-ne, ta¹-jiu-zǒule.	(c1) A: /° (c2) B: \\
E1.27	5.7	zou³le-dao dà-shān-libianr. gen¹ben jiu-**méi**-you-diànhuà, méi-you-shen²me, gen¹ben jiu-**méi**far-liánxì shi⁴-bu-shi?	(c1) B: \ /°° (c2) B: \°\ / (c3) B: \\ (c4) C: \°\ /+\

2. Event: After some time, Mrs. Qin came by coincidence together with a barefoot doctor to the village where the child was from. His mother dragged them to their home and told them the child had healed from the disease. Then, the mother served them some local dish that Mrs. Qin could not eat because of its extreme spiciness. Taking advice from the barefoot doctor, she asked the child's mother to let her take it away and eat later, so as not to show a lack of respect. At the end of this part, Mrs. Qin admires the sincerity of the farmer's gratitude.

Section	Syl/sec	Transcript	Rhythm
E2.01	2.8	you-yí-ci nei⁴-ge,tā-d-nei⁴-ge,chìjiao yīshēng...	(c1) B: °\\ (c2) B: \\ (c3) B: \ /
E2.02	4.8	nèi-shihour jiu-guǎn-tamen jiao-chì jiu⁴shi guāngzh-jiǎobuyār-d nei⁴-ge, chìjiao-yīshēng.	(c1) C: \°\ /+\\V°\ (c2) B: \ /
E2.03	4.3	ran²hou-tā, wo³men-liǎng-ge, bu⁴-zhidào yao⁴-qu-**zuò**-shenme, dao-nóngcun-li³bianr.	(c1) A: V (c2) B: \\ (c3) C: V+\\ (c4) B: °\\
E2.04	3.1	zhènghaor-zǒudao zhei⁴-ge-nǚd ta¹men-jiā.	(c1) B: \\\\V
E2.05	3.4	wa! ta-**fēi**-yao ba³-wo-lādao ta¹men-jiā!	(c1) A: I (c2) B: °\\\V
E2.06	3.8	ran²hou-wǒ, wo³-bu-rènshi-**tā**,	(c1) A: V (c2) B: \V
E2.07	3.5	yin¹wei-tā, ~~nong~~ nèi-xie-ren² dou-zhǎng-d wo³-dou-juéde tèbie-yi-yàng.	(c1) A: V (c2) D: V°\\\\ /
E2.08	3.4	suo³yi-hòulai, ~~tā~~ ta¹-jiu-gēn-wo-shuō zhèi-jian-**shìr**, ta-shuō wo³-d-xiǎoháir mei²-you-**sǐ**.	(c1) B: \\ (c2) C: \V+V (c3) C: /+\ /+V
E2.09	4.2	huíjiā wo³-jiu-gei-tā-hē zhei-shuǐ, ta¹-jiu-shi-huó, jiu-hǎo, ~~jiu jiu hao le nei bing.~~	(c1) C: /+\ // (c2) B: \ / (c3) A: /
E2.10	2.2	hou⁴lai-nèi-ge, tā-gei-wo **zuò**-shenme-ne? nèi-ge	(c1) B: \\ (c2) B: \\ (c3) A: \
E2.11	2.7	~~qi~~ qīshu₄, jiùshi, yóu**qī**-d-nei⁴-ge, **shù**-keyi chū-youqi¹ wo³-bu-zhīdao.	(c1) A: \ (c2) A: \ (c3) D: /°\ (c4) C: V+\\
E2.12	3.4	ta¹men-jiao-**qī**shu₄, jiu⁴shi yóuqi-d-qī.	(c1) B: \\ (c2) B: \\ /
E2.13	2.7	qīshu-d, nei⁴-ge-**shù**-**zīr**, zhàchulai-d-yóu	(c1) A: \ (c2) B: \ / (c3) B: \ /
E2.14	3.4	gei³-wo-**zhá**-yi-ge, ~~jiu jiu~~ fan³zheng-wo-gūji-jiao, ta-shuō-shi, jiu-xiàng, tā-jiao wánz-ba, quèshi-jiushi sū-mian-gēdar	(c1) B: \\ (c2) B: \\ (c3) A: °\ (c4) A: / (c5) B: \ (c6) B: \\\
E2.15	3.8	nèi-ge-dong¹xi, ke³shi-zhèi-ge	(c1) B: \\ (c2) B: \\
E2.16	2.8	**méi**-you₃, jiu⁴-shi-miàn.	(c1) A: \\ (c2) A: V
E2.17	2.9	ránhou, nèi-ge. wa! chi¹-zhei-ge-yóu-ba, tèbie-**là** sang³z-**yān**-d-huang, **là**-d nei⁴-zhong-wèir.	(c1) A: \ (c2) A: \ (c3) A: | (c4) B: \ /° (c5) C: V+\\ (c6) B: \V
E2.18	3.3	ni-gēnben yānbuliǎo.	(c1) B: °\V
E2.19	4.7	hòulai-wo³-jiu-**gǎn**kuai toutōur gen¹-nei chìjiao-yīsheng-shuo¹, wo-shuo¹ zhei-dōngxi-wo³ zhēn-d-chi¹buliǎo.	(c1) C: \\\ /+\\V (c2) C: /°V+\V
E2.20	4.4	ke³shi-**tā**, jiu⁴shi-tāmen-jiā **zuì**-hao-d-dōngxi, ta¹-gei-ni-zuò yóu-**zhá**-d zhei-dōngxi.	(c1) A: V (c2) C: V+\\ (c3) D: //°°
E2.21	3.6	hòulai-wo ~~wo~~ wo-shuo¹ wo-**zhēn**d-chībuliǎo. ta-shuō, wo-shuo¹ ni-gǎnkuai xiang-bànfa₃.	(c1) A: \ (c2) D: /°\V (c3) A: / (c4) D: /°\^
E2.22	4.4	hòulai ta¹-jiu gen¹-zhei-nǚd-shuō (+ a question of the man)	(c1) B: \\\V
E2.23	2.6	**dāng**di-d, cūn-libianr-d	(c1) A: \ (c2) A: \
E2.24	3.1	ranhòu, ta¹-jiu-gǎnkuai gen¹-ta zěnme-shuō.	(c1) A: / (c2) B: \\\V
E2.25	3.9	shuo¹le-yihòu ke-hái-gei-wo-dàishang. dàishangle, wo³-jiu gǎnkuai-zǒu-le.	(c1) D: \ /°\\ (c2) A: \ (c3) B: \\\
E2.26	3.3	jiu-shuō hái-you bie²-d-shìr	(c1) C: /+\V
E2.27	2.2	hòulai, suǒyi jiushì, ta-**dāng**di-d nóngmin-ba, jiushì, tā...	(c1) A: \ (c2) B: \ / (c3) B: °\\ (c4) A: / (c5) A: |
E2.28	2.8	yin¹wei-gēn **wài**bianr jiēchu-bu-duō, suo³yi-tā xīn tèbie-zhēncheng.	(c1) C: V+\\ / (c2) C: V+|+\\
E2.29	1.9	**zhēn**d-gǎnxie-ni³!	(c1) B: \V
E2.30	2.0	zhei-gùshi wo-yí-beiz ye-wàngbuliǎo	(c1) B: °\°\°V

3. Event: Mrs. Qin elaborates her experience with farmers' genuine and sincere hospitality. She describes how she met an old farmer who offered her a dish that, as she assumed, consisted of ingredients valuable for the farmers, and she expressed her concerns that perhaps there might not be enough for the farmers themselves.

Section	Syl/sec	Transcript	Rhythm
E3.01	3.2	zai-yǒu, jiǎru-dao⁴ tiáojiàn shāowei-hǎo-yidianr³-d-a, nèi-ge, nei-nóngmin jiā-li₃bianr	(c1) A: / (c2) C: V/+\V°° (c3) A: \ (c4) B: °\\
E3.02	4.0	ranhòu, wo³men you³d-shíhou yě-ke₃neng gōngshe-d gen¹-wo yikuàir-qu	(c1) A: / (c2) B: \\\\\\ /°
E3.03	3.4	ránhòu jiushi⁴ lǎotàitai jiu⁴-gei-wo-zhǔ yi-wǎnr-miàn	(c1) D: //°\\ //
E3.04	4.4	shuō chī-wǎnr-miàn, nuǎnhuo-nuan³huo-ba, chī-yi-wǎnr-miàn nuǎnhuo-nuan³huo	(c1) C: |+V (c2) B: \\ (c3) C: \ /+\\
E3.05	3.5	shíjishang wo³-d-biǎomiànr yì-céng shi-miàntiáor	(c1) D: /°\ ///
E3.06	4.3	duì, ta¹men-jiā kěndìng-you-ren² zai-wàibianr gongzuo⁴, suo³yi jiu⁴-gei-ta-yǒu nèi-zhong **shǒu**gōng-d **xiǎo**-d guà-miàn.	(c1) A: | (c2) D: V//°\ / (c3) C: \\ /+\\\ /
E3.07	3.0	dōu-shi, **hěn**-bu-róngyì-d wo-zhīdao, suǒyi	(c1) A: \ (c2) D: \ /°°\ (c3) A: \
E3.08	3.4	zhèi-ge, ke³shi-díxia-ne, wo³-nei wǎn-li₃tou, ke³yi-fàng sì-**wǔ**-ge-jīdàn, biǎomiànr-yì-céng miàntiáor	(c1) A: \ (c2) B: \\ (c3) B: \\ (c4) D: V°\ / (c5) B: \ //
E3.09	3.9	**jiù**shi yào-gei-wo chī-d, ni-zhīdao-ma	(c1) B: \\\ (c2) A: °\
E3.10	2.4	suǒyi, jiùshi, kěshi tāmen biéren…	(c1) A: \ (c2) A: \ (c3) B: \\\
E3.11	4.8	kěndìng-méi-yǒu. rang-tāmen, gei-tā yidianr³ buguǎn-shénme-chī, ~~wo bu~~ wo³-dou bú-jìde-le	(c1) B: \ / (c2) A: °\ (c3) B: /// V (c4) B: \°\
E3.12	3.7	wo³-jiu-zhèi-ge, wo³-jiu **tè**bie-gǎndòng, lǎotàitai-dui-wo³ **zhēn**-xin-zhen-yì-d	(c1) B: \\ (c2) B: \\ / (c3) C: \ /+\ /°

Appendix 2: An example description of linear segmentation and syllable prominence

List of abbreviations: c = colon, s = segment, A = arsis, T = thesis

Example 1: E1.01 you³-yi-cì, zǎochen-qi₃lai², wo-hái-mei qǐchuáng,

Linear segmentation: 3 cola (c1, c2, c3); 4 segments (c1_s1; c2_s1; c3_s1, c3_s2)

Syllable prominence:

(c1_s1) you³-yi-cì: A weakened-tone + T void-tone + A full-tone

(c2_s1) zǎochen-qi₃lai²: A full-tone + T void-tone + T weakened-tone + A weakened-tone

(c3_s1) wo-hái-mei: T weakened-tone + A full-tone + T void-tone

(c3_s2) qǐchuáng: T full-tone + A full-tone

Example 2: E2.29 yin¹wei-gēn **wài**bianr jiēchu-bu-duō, suo³yi-tā xīn tèbie-zhēncheng

Linear segmentation: 2 cola (c1, c2); 6 segments (c1_s1, c1_s2, c1_s3; c2_s1, c2_s2, c2_s3)

Syllable prominence:

(c1_s1) yin¹wei-gēn: A weakened-tone + T void-tone + A full-tone

(c1_s2) **wài**bianr: A stressed-prominence + T void-tone

(c1_s3) jiēchu-bu-duō: A full-tone + T void-tone + T void-tone + A full-tone

(c2_s1) suo³yi-tā: A weakened-tone + T void-tone + A full-tone

(c2_s2) xīn: A full-tone

(c2_s3) tèbie-zhēncheng: A full-tone + T void-tone + A full-tone + T void-tone

Example 3: E3.07 dōu-shi, **hěn**-bu-róngyì-d wo-zhīdao, suǒyi

Linear segmentation: 3 cola (c1, c2, c3); 4 segments (c1_s1; c2_s1, c2_s2; c3_s1)

Syllable prominence:

(c1_s1) dōu-shi: A full-tone + T void-tone

(c2_s1) **hěn**-bu-róngyì-d: A stressed-prominence + T void-tone + T full-tone + A full-tone + T void tone

(c2_s2) wo-zhīdao: T weakened-tone + A full-tone + T void-tone

(c3_s1) suǒyi: A full-tone + T void-tone

A Colexficational Analysis of *Chū* in Modern Chinese

KUANG Lincai
College of International Studies, Southwest University, Chongqing, China
kuanglincai@163.com

Abstract

Directional verbs in modern Chinese, notably *chū*, are crucial. Previous studies on *chū* pay little attention to colexification and its mechanisms; they primarily focus on polysemy, syntactic restriction, and diachronic evolution in Mandarin Chinese. Research on Chinese topolects has also been neglected. Development of the Database of Cross-linguistic Colexifications (CLICS[3]) facilitates its cross-linguistic and topolectal exploration. This paper first discusses colexified pairs of *chū* in Modern Chinese, and constructs its colexificational network based on the CLICS[3]. It also argues for colexificational mechanisms of colexified extensions. Conceptual metaphor and metonymy are considered two core mechanisms accounting for the colexified extensions of *chū*. Metaphoric colexified extensions map onto the space, time, and state domain, following the order of "space > time > state". As for the metonymic colexified extensions, the extended concepts result from different substructures of the image schema profiled in the motion event of *chū*.

Keywords: *chū*, directional verb, colexified pair, metaphor, metonymy

Povzetek

Smerni glagoli, še posebej *chū* v sodobni kitajščini, so ključnega pomena. Dosedanje študije o *chū* namenjajo premalo pozornosti koleksifikaciji in njenim mehanizmom; raje se osredotočajo na polisemijo, sintaktične omejitve in diahroni razvoj v standardni kitajščini. Še posebej neraziskani so dialekti. Nedavno oblikovana baza podatkov o medjezikovnih koleksifikacijah (CLICS[3]) omogoča vpogled v glagol *chū* z medjezikovnega in narečnega vidika. Članek prvo obravnava njegove koleksificirane pare v sodobni kitajščini, nato pa na podlagi CLICS[3] razvije koleksifikacijsko mrežo. Obenem predlaga mehanizme koleksifikacijske razširitve koleksificiranega. Osnovna mehanizma koleksifikacije sta konceptualna metafora in metonimija, ki pojasnjujeta koleksificirane razširitve glagola *chū*. Njegove metaforične koleksificirane razširitve je moč preslikati v prostorsko in časovno domeno ter stanje, pri čemer si le-te sledijo vrstnemu redu "prostor > čas > stanje". V primeru metonimične koleksificirane razširitve pa se razširjeni koncepti pojavijo iz različnih podstruktur slikovne sheme, profilirane v gibalnem dogodku glagola *chū*.

Ključne besede: *chū*, smerni glagoli, koleksificiran par, metafora, metonimija

1   Introduction

François (2008) defines colexification as "a given language is said to colexify two functionally distinct senses if, and only if, it can associate them with the same lexical form" (2008, p. 163). He claims that whether a sense can be included in a semantic list depends on whether two senses are strictly colexified in at least one language. Interestingly, the development of the Database of Cross-linguistic Colexifications (CLICS³) (2019) makes it possible to conduct a colexficational analysis of a lexical form. Directional verbs are specific and widely used in modern Chinese, among which *chū* 出 "go out" is an important one. Previous studies of *chū* pay little attention to colexification but focus on polysemy (Liu, 1998; Shi, 2001; Wang, 2005; Hu et al., 2019), syntactic restriction (Chu et al., 1999; Hu et al., 2019), and diachronic evolution (Wu, 1997; Chu et al., 1999; Liu, 2015; Zhou, 2018) in Mandarin Chinese, while topolects only play an inconspicuous role in verifying the results. Whereas topolects are indispensable parts of modern Chinese (Chu et al., 1999; Liu, 2015) due to some characteristics different from those in modern Chinese. In this sense, topolects should be attached as much importance as Mandarin Chinese in the study of *chū*. Based on Mandarin Chinese and topolects in Chinese, this present paper intends to explore the complete colexified pairs of *chū*, and then construct its colexificational network, as well as uncover its colexficational mechanisms.

    This paper is unfolded as follows. Section 2 explains the data and methodology for this study. Section 3 discusses the colexified pairs of *chū* in Mandarin Chinese and 43 topolects from pertinent data. To justify the colexifications for *chū* in different topolects, Section 3 employs the CLICS³ to complement cross-linguistic colexifications for the concept GO OUT, and further establishes the colexificational network of *chū* in modern Chinese. Section 4 discusses the mechanisms for the colexified extensions of *chū*. And the last section 5 summarizes the whole study and concludes.

2   Data and methodology

This present paper makes use of Mandarin Chinese data from the online version of the BCC Chinese Corpus created and managed by Beijing Language and Culture University, and the Modern Chinese Dictionary (7th edition). The corpus and the dictionary consist of rich Mandarin Chinese instances, which can be utilized to testify to the common lexical semantics or concepts of *chū* in Mandarin Chinese. The data for the quantitative analysis of lexical semantics or concepts of *chū* in Chinese topolects in section 3 comprise the large dictionary of modern Chinese topolects (43 volumes) and

the Database of Cross-Linguistic Colexifications (CLICS[3]). The large dictionary is selected because it is widely considered to be a standard representative of modern Chinese topolects, including all Chinese topolectal families identified by Language Atlas of China (2nd edition) (Xi & Zhang, 1990) as well as tremendous instances of different topolects. CLICS[3] covers more than 3100 language varieties of the world, showing the concepts that share the same lexical form with the target concept, having been proven to be scientific in psychological and cognitive research (Joshua et al., 2019). The database provides us with some cross-language evidence to some degree, demonstrating the commonality between *chū* in Mandarin Chinese and other languages. In addition, this paper also collects previous studies of *chū* in Mandarin Chinese (Liu, 1998; Chu et al., 1999; Hu, 2012) to conduct a fine-grained analysis.

From the preceding description, it is clear that data have been drawn from a great variety of sources. This is inevitable because the colexificational changes are slow and proceed in small steps, whereas this study requires refined data covering detailed colexified pairs. Fortunately, the large dictionary of topolects and CLICS[3] make it possible to conduct a close-grained analysis of *chū* in Modern Chinese.

3   The colexified pairs and colexificational network of *chū*

Lehrer (1992) suggests that the typological study focuses on the specific way of how to package linguistic materials into lexemes or phrases with the purpose of uncovering the internal universality and systematicity of languages within lexical levels. Since lexicons are traditionally counted as "unique" and "idiosyncratic" without a common ground for cross-linguistic comparison, François (2008) further argues that breaking up lexemes into semantic atoms or senses can make the cross-linguistic comparison between lexical items possible. It is precisely for this reason that he incorporates the notion of colexification into cross-linguistic lexical studies to resolve the thorny issue of comparability for lexical items. The following part discusses the colexified pairs of *chū* and constructs its colexificational network in modern Chinese.

3.1   The colexified pairs of chū in Modern Chinese

As is argued by François (2008), whether a sense can be incorporated in a semantic list hinges upon the strict colexification of two senses in at least one language. Specifically, if a linguistic form in a sort of Language has <sense1>

and <sense2>, at the very least it can be seen as evidence of the view that the form strictly colexifies <sense1> and <sense2>, manifested by the colexified pairs <sense1, sense2>. Gast and Koptjevskaja-Tamm (2022) expand the view and further distinguish independent colexification from dependent one, that is, the colexification resulting from an intermediate concept. Additionally, they also add negative evidence to exclude some pairs from the colexificational database. Based on these views, this part argues the colexified pairs of *chū* in modern Chinese. Since "GO OUT" is the basic sense of *chū*, such as *chūmén* 出门 "go out of the door", and it can be found in every topolect in Chinese, this present paper selected it as the pivot node in the colexificational network, in sync with the pivot concept (François, 2008). Finally, 9 colexified pairs of *chū* in modern Chinese have been identified, including <GO OUT, HAPPEN>, <GO OUT, ARRIVE>, <GO OUT, EXCEED>, <GO OUT, TAKE OUT>, <GO OUT, PRODUCE>, <GO OUT, PAY>, <GO OUT, COME TO AN END>, <GO OUT, GO>, <GO OUT, OUTSIDE>. In what follows, each colexified pair will be elaborated further and supported with examples.

<GO OUT, HAPPEN>: *Chū* can express the concept of HAPPEN, which can be found in Mandarin Chinese, Chengdu topolect, Taiyuan topolect, Xi'an topolect, etc. Examples are as follows:

(1) 病人<u>出</u>血太多，需要立刻包扎。(Mandarin Chinese)
      Bìngrén <u>chū</u> xiě tài duō, xūyào lìkè bāozhā.
      patient – <u>happen</u> blood – too much, need a bandage immediately
      'The patient was bleeding too much and needed to be bandaged immediately.'

(2) 手上<u>出</u>咧个疙瘩。 (Xi'an topolect)
      Shǒushàng <u>chū</u> liě gè gēdá.
      on the hand – <u>happen</u> – a blister
      'There is a blister on the hand.'

<GO OUT, ARRIVE>: *Chū* can denote the concept ARRIVE, which can be found in Mandarin Chinese, Meixian topolect, Chengdu topolect, Xi'an topolect, etc. Illustrative examples are given as follows:

(3) 全体代表<u>出</u>席了今天的大会。 (Mandarin Chinese)
      Quántǐ dàibiǎo <u>chū</u>xí le jīntiān de dàhuì.
      all representatives– <u>arrive</u> –today's conference
      'All representatives attended today's conference.'

(4) 这阵应该<u>出</u>场么? (Chengdu topolect)
  Zhè zhèn yīnggāi <u>chū</u>chǎng me.
  the time should– <u>arrive</u>
  'I don't think this is the time to play.'

<GO OUT, EXCEED>: *Chū* can imply the concept of EXCEED, which can be found in Mandarin Chinese, Shanghai topolect, Xi'an topolect, Jixi topolect, etc. Examples are as follows:

(5) 他在公司的地位已经大大超<u>出</u>了他老板。(Mandarin Chinese)
  Tā zài gōngsī de dìwèi yǐjīng dàdà chāo<u>chū</u> le tā lǎobǎn.
  his status in the compapany greatly– <u>exceed</u> –his boss
  'His status in the company has greatly exceeded that of his boss.'

(6) 不<u>出</u>三年，渠肯定就是个好木匠了。 (Jixi topolect)
  Bù <u>chū</u> sān nián，qú kěndìng jiùshì gè hǎo mùjiàng le.
  not– <u>exceed</u> –three years, you must a good carpenter
  'In three years, you will be a good carpenter.'

<GO OUT, TAKE OUT>: *Chū* can indicate the concept of TAKE OUT, which can be found in Mandarin Chinese, Guangzhou topolect, Taiyuan topolect, Meixian topolect, etc. Examples are as follows:

(7) 你看你<u>出</u>的什么主意，一点用都没有。 (Mandarin Chinese)
  Nǐ kàn nǐ <u>chū</u> de shénme zhǔyì, yīdiǎn yòng dōu méiyǒu.
  you look your– <u>take out</u> – ideas, no use
  'Look, whatever ideas you came up with are of no use at all.'

(8) <u>出</u>支估仔给你估。 (Guangzhou topolect)
  <u>Chū</u> zhī gūzǎi gěi nǐ gū.
  I– <u>take out</u> –a riddle you guess
  'Take out a riddle and try to guess it.'

<GO OUT, PRODUCE>: *Chū* can signify the concept of PRODUCE, which can be found in Mandarin Chinese, Taiyuan topolect, Shanghai topolect, Jixi topolect, etc. Examples are as follows:

(9) 我们县城里今年出了个文科状元。 (Mandarin Chinese)
　　Wǒmen xiànchéng lǐ jīnnián chū le gè wénkē zhuàngyuán.
　　our county this year– produce –a liberal arts champion
　　'Our county has a liberal arts champion this year.'

(10) 爾徽州真出人才。 (Jixi topolect)
　　Ěr huīzhōu zhēn chū réncái.
　　your Huizhou indeed– produce –talents
　　'Huizhou is a place full of talents.'

<GO OUT, PAY>: *Chū* can typify the concept of PAY, which can be found in Mandarin Chinese, Chongming topolect, Danyang topolect, and Xiamen topolect. Examples are as follows:

(11) 大家尽情吃喝，今天这顿饭饭钱我出。 (Mandarin Chinese)
　　Dàjiā jìnqíng chī hē, jīntiān zhè dùn fàn fànqián wǒ chū.
　　everyone eat and drink today the bill I– pay
　　'Everyone, eat and drink to your heart's content. I'll pay for today's meal.'

(12) 我出五块钱和你交换。 (Danyang topolect)
　　Wǒ chū wǔ kuài qián hé nǐ jiāohuàn.
　　I– pay –five yuan for your exchange
　　'I'll give you five bucks in exchange.'

<GO OUT, COME TO AN END>: *Chū* can designate the concept of COME TO AN END, which can be found in Loudi topolect, Meixian topolect, Danyang topolect, etc. Examples are as follows:

(13) 我早就讲了，这个人是要出豁的。 (Loudi Chinese)
　　Wǒ zǎo jiù jiǎng le，zhè ge rén shì yào chūhuō de.
　　I early told, the guy must– come to an end
　　'I have told you, you and that guy must come to an end.'

(14) 明交出伏咧。 (Danyang topolect)
　　Míng jiāo chūfú liē.
　　tomorrow heat– come to an end
　　'The heat will come to an end tomorrow.'

<GO OUT, GO>: *Chū* can express the concept of GO, which can be found in Mandarin Chinese and Guangzhou topolect. Examples are as follows:

(15) 逢年过节都要<u>出</u>城一趟。 (Mandarin Chinese)
Féng nián guòjié dōu yào <u>chū</u>chéng yī tàng.
go through the year– <u>go</u> –to the town
'I have to go out of the town once a year.'

(16) 你<u>出</u>街做咩？ (Guangzhou topolect)
Nǐ <u>chū</u>jiē zuò miē?
you– <u>go</u> –street do what
'Why are you going to the street?'

<GO OUT, OUTSIDE>: *Chū* can denote the concept of OUTSIDE, which can be found in Guangzhou topolect, Xi'an topolect, and Urumqi topolect. Examples are as follows:

(17) 行<u>出</u>啲。 (Guangzhou Chinese)
Xíng <u>chū</u> dī.
go– <u>outside</u>
'I will go outside.'

(18) 把你底好酒往<u>出</u>拿。 (Urumqi topolect)
Bǎ nǐ dǐ hǎo jiǔ wǎng <u>chū</u> ná.
take you wine– <u>outside</u>
'Take your wine outside.'

This present paper exhaustively explored the above 9 colexified pairs of *chū* in Mandarin Chinese and Chinese topolects from pertinent data. Following the Chinese topolectal families delineated in the Language Atlas of China (2nd edition) (Xi & Zhang, 1990), a compilation of 21 exemplary topolects has been chosen to illustrate the outcomes, as depicted in Table 1.

Table 1: The colexfied pairs of *chū* in Mandarin Chinese and topolects

	Mandarin	Chongming	Chengdu	Pingxiang	Meixian	Yangzhou	Xining	Xi'an	Nanchang	Taiyuan	Fuzhou	Guangzhou	Haikou	Jinan	Shanghai	Loudi	Jixi	Danyang	Xiamen	Harerbin	Urumqi
GO OUT	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
HAPPEN	+	+	+	+	+	+	+	+	+	+	+		+	+	+	+					+
ARRIVE	+		+		+		+	+		+		+			+						+
EXCEED	+				+			+		+				+	+		+				+
TAKE OUT	+				+			+		+		+			+	+	+				+
PRODUCE	+	+			+			+		+	+	+			+			+			+
PAY	+	+																+	+		
COME TO AN END			+	+	+	+			+	+	+	+	+			+		+			
GO		+											+								
OUTSIDE								+					+								+

3.2   The colexificational network of *chū*

This part resorts to CLICS[3] to justify the colexified pairs of *chū* in Modern Chinese. As illustrated earlier, given its fundamental meaning and top frequency, "GO OUT" is deemed to be the pivot node in the colexificational network. When "GO OUT" is selected as the target concept in CLICS[3], the result can be exhibited in Figure 1.
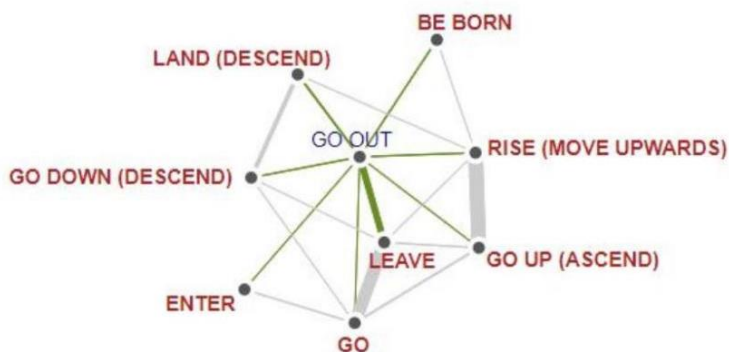
Figure 1: The Colexificational Network of "GO OUT" in CLICS[3]

As sketched in Figure 1, 8 colexified pairs of "GO OUT" can be identified, including <GO OUT, LAND (DESCEND)>, <GO OUT, GO DOWN (DESCEND)>, <GO OUT, ENTER>, <GO OUT, GO>, <GO OUT, LEAVE>, <GO OUT, GO UP (ASCEND)>, <GO OUT, RISE (MOVE UPWARDS)>, <GO OUT, BE BORN>.

Moreover, the colexficational network of *chū* in modern Chinese and "GO OUT" in different languages are combined to showcase their overlap and complementarity. It is important to mention here that both the threshold of the nodes and lines in CLICS[3] are three, indicating that more than three kinds of language have such a concept and more than three sorts of forms hold such colexified pairs respectively. This present paper identifies the colexified pair of *chū* in Modern Chinese in tandem with the same threshold of nodes and lines to make the two colexficational networks well-matched.

The conjoined colexficational network of *chū* "GO OUT" is demonstrated in Figure 2, and only one overlapping colexified pair is existent, that is <GO OUT, GO> (see the shadow part).
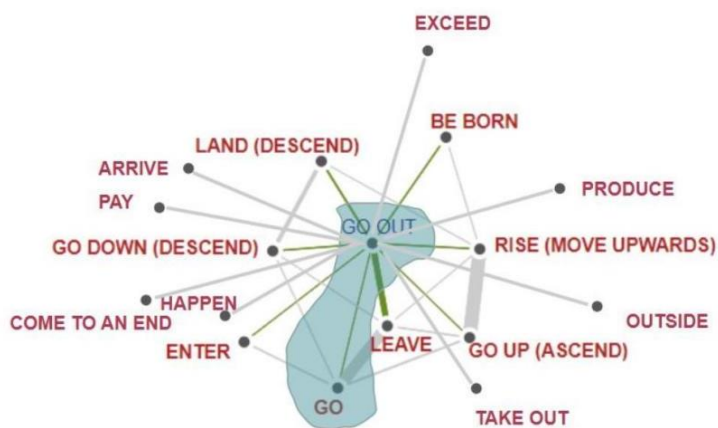
Figure 2: The colexificational network of *chū* ("GO OUT")

4   The colexificational mechanisms of *chū*

The colexified extension of *chū* is not arbitrary but follows certain rules. As is explained by Gibbs and Matlock (1997, p.215), "the lexical organization of polysemous words is not a repository of random, idiosyncratic information, but is structured by general cognitive principles that are systematic and recurrent throughout the lexicon". Conceptual metaphor and metonymy are long advertised as two fundamental cognitive mechanisms we live by. Specifically, they are viewed as the internal force of semantic expansion of the lexicon (Lakoff & Johnson, 2003) as well as the underlying reason for colexification (Gast & Koptjevskaja-Tamm, 2022). This present paper contends that conceptual metaphor and metonymy are the core mechanisms during the colexified extended process of *chū*, among which image schema, particularly for container schema, comes into play.

Additionally, Lakoff and Johnson (1980) pointed out that the space domain is the most fundamental one in human cognition, which can be further projected into some abstract domains, such as the time domain. And for Heine et al. (1991), abstract concepts, derived from other space concepts, are conventionally counted as the most fundamental concepts in human language as such. As *chū* is patently relative to space in modern Chinese, its colexification has to do with mapping onto three domains, including the space domain, the time domain, and the state domain. In what follows, the core mechanisms of *chū* will be elaborated in terms of diversified domains.

4.1   Metaphoric mapping in the space domain

The first colexified chain of *chū* extended from metaphoric mappings in the space domain can be illustrated as follows: GO OUT > TAKE OUT. It is significant to mention here that the meaning or concept of TAKE OUT is extended on the basis of the fundamental concept of GO OUT, including not only the real move but also some mental ones.

When referring to the spatial concept, container schemata lend themselves to modeling in terms of pertinent conceptual metaphors. Specifically, container schemata divide the space into three parts, including the interior, the exterior, and the boundary, which can be further projected onto one entity with a boundary. To take one example, a room has a boundary (i.e. wall), separating the space into interior and exterior, so it is precisely for this reason that some physical motions can be described as "moving from one room to another", which can be schematized as A ROOM IS A CONTAINER. More illustrative examples are as follows:

(19) 地震了，赶快跑出这栋楼。 (Mandarin Chinese)
       Dìzhèn le, gǎnkuài pǎo <u>chū</u> zhè dòng lóu.
       an earthquake happen go– <u>out</u> –this building
       'There's an earthquake. Run out of the building quickly.'

(20) 她把书拿出了教室。 (Mandarin Chinese)
       Tā bǎ shū ná <u>chū</u> le jiàoshì.
       she take book– <u>out</u> –of classroom
       'She took the book out of the classroom.'

As for physical activity, the correspondences exist likewise. Many aspects in the domain of CONTAINER, such as the interior, exterior, and boundary,

can be selectively highlighted in the BUILDING domain, as demonstrated in Table 2.

Table 2: The container mapping of *chū* in the space domain

Source: CONTAINER domain	Target: BUILDING domain
whole container	whole building
boundary	wall/ceiling/door/floor
interior of container	interior of building
exterior of container	exterior of building
departure of container	go out of the building

When the mapping of *chū* is projected onto abstract spatial concepts without clear boundaries, such as hospitals and governments, the meaning of *chū* can be extended to "going out of institution" or "removing affiliation of a space", exemplified in (21) and (22). Such physical motion from the interior to the exterior can be sketched as a process including the source point, the target point, the moving direction, and the figure, corresponding to the CONTAINER metaphoric elements. Specifically, the source point corresponds to the interior, the target point to the exterior, the moving direction to the motion direction, and the figure to the entity. Additionally, the mapping of *chū* can be further projected onto the more abstract mental spatial concept, such as emotions, which can be schematized as EMOTION IS CONTAINER, exemplified in (23).

(21) 要遵守秩序，不然会被赶出医院。 (Mandarin Chinese)
    Yào zūnshǒu zhìxù, bùrán huì bèi gǎn <u>chū</u> yīyuàn.
    should obey order or will be expelled– <u>out</u> –the hospital
    'You must obey the order, otherwise you will be expelled from the hospital.'

(22) 那个犯人被驱逐出了这个国家。 (Mandarin Chinese)
    Nàgè fànrén bèi qūzhú <u>chū</u> le zhège guójiā.
    the prisoner be expelled– <u>out</u> –this country
    'The prisoner was deported from the country.'

(23) 不要难过，你要尽快走出悲伤才行。 (Mandarin Chinese)
    Bùyào nánguò, nǐ yào jǐnkuài zǒu<u>chū</u> bēishāng cái xíng.
    do not sad, you need to take yourself– <u>out</u> –of it as soon as possible
    'Don't be sad, you have to get over the sadness as soon as possible.'

There is no denying that the motion of *chū* is not limited to the real physical move, but also the virtual move in the mind based on spatial experience. Metaphorically speaking, the figure of *chū* can be conceptualized as an event, idea, activity, etc., which are the objects of TAKE OUT.

4.2   Metaphoric mapping in the time domain

The second colexified chain of *chū* extended from metaphoric mappings in the time domain can be illustrated as follows: GO OUT > COME TO AN END.

Space and time are two core elements of human cognition (Langacker, 1987; Evans, 2004; Croft & Cruse, 2004), and time is in the long run more complex since it could be understood in terms of another concrete concept. For some reason, time can be compared to an entity, within which time frame is compared to a container. So when the entity moves from the container inside (within an episode of time) to the container outside (outside an episode of time), the basic concept of GO OUT can be extended to the concept of COME TO AN END. Pertinent mappings are shown in Table 3.

Table 3: The container mapping of *chū* in the time domain

Source: CONTAINER domain	Target: TIME domain
whole container	full episode of time
boundary	time point
interior of container	within an episode of time
exterior of container	outside an episode of time
departure of container	come to an end

Illustrative examples are given as follows:

(24) 上海刚刚出梅，即连日大热。 (Fuzhou topolect)
Shànghǎi gānggāng <u>chū</u>méi, jí liánrì dà rè.
Shanghai just– out –mould rain season begin hot
'Mould rain season just came to an end, it is starting to be very hot.'

(25) 出年我就要学识听粤语。 (Guangzhou topolect)
<u>Chū</u>nián wǒ jiù yào xuéshí tīng yuèyǔ.
<u>out</u> –this year I learn listsen Cantonese
'When this year comes to an end, I will learn to understand Cantonese.'

4.3   Metaphoric mapping in the state domain

The third colexified chain of *chū* extended from metaphoric mappings in the state domain can be illustrated as follows: GO OUT > HAPPEN.

When the form *chū* expresses the concept of HAPPEN, it can be viewed as a change from one state to another. Specifically, when an entity is inside the container, it is confidential and well-protected and often counted as non-existent. Meanwhile, it can apparently be perceived without protection when it leaves the container. So the concept of HAPPEN as in *Shēntǐ chū le máobìng* 身体出了毛病 "Something wrong happened to the body", can be interpreted as a change of the container (i.e., "body") from the healthy state to its ill state. Pertinent mappings are shown in Table 4.

Table 4: The container mapping of *chū* in the state domain

Source: CONTAINER domain	Target: STATE domain
whole container	body
boundary	the surface of the body
interior of container	inside the body
exterior of container	outside the body
departure of container	state change

It is important to mention that the image schema, particularly for container schema, is also extended in the course of the colexified extensions of *chū* from the spatial domain to the state domain. To put it in another way, it changes from a concretely spatial entity to an abstract state, which reconciles with human embodied experience (Lakoff, 1987). More illustrative examples are as follows:

(26) 停车，发动机<u>出</u>故障了，不能再往前开了。 (Mandarin Chinese)
　　 Tíngchē, fādòngjī <u>chū</u> gùzhàngle, bùnéng zài wǎng qián kāile.
　　 stop car and engine go– <u>out</u>
　　 'Stop the car, something wrong happened with the engine.'

(27) 手上<u>出</u>咧个疙瘩。 (Xi'an topolect)
　　 Shǒushàng <u>chū</u> liě gè gēdá.
　　 on the hand – <u>happen</u> – a blister
　　 'There is a blister on the hand.'

Chū extends its colexified chain through mapping onto the space domain, time domain, and state domain. Broadly speaking, the colexified extensions of *chū* are on a par with the metaphoric hierarchy put forward by Heine et al. (1991, p.48), that is, person > object > activity > space > time > quality. As is explained by Wu (2003), the essence of such hierarchy per anthropocentricity as well as the concept of EMBODIMENT that people understand things from their most familiar empirical category to the empirical category which is away from them.

4.4   Metonymic extensions of *chū*

The fourth colexified chain of *chū* extended from conceptual metonymy can be illustrated as follows: GO OUT > PAY, GO OUT > PRODUCE, GO OUT > GO, GO OUT > ARRIVE, GO OUT > OUTSIDE, GO OUT > EXCEED.

Traditionally, conceptual metonymy can be defined as an apparatus in which the understanding of one conceptual entity provides mental access to another. Meanwhile, Langacker (2013, p.69) uses the concept of "profile" in Cognitive Grammar, that is, the specific focus of attention within its immediate scope, to characterize metonymy as a "shift in profile". The colexified extensions of *chū* can be profiled into a diversified substructure based on its motion event.

Additionally, Talmy (2000) points out that two core elements should be taken into account when it comes to a basic translational motion, that is, figure and ground. The former refers to the moving entity against the reference point, and the latter is related to the physical environment of the moving event, including the source, goal, and position of the entity. Zeng (2009) incorporates the view and further points out that there are different degrees of salience in a translational motion, resulting in different semantic extensions. The motion even of *chū* can be depicted as the scene where an entity (figure) moves from a starting point in a space (ground) outside to an ending point. Its image schema can be sketched in Figure 3. The circle represents the figure, the rectangle represents the ground, and the arrow represents the moving direction and path. The ground builds a space in which the figure is located, and then the figure moves from the starting point A toward the ending point C in stages. It is important to mention that point B expresses the figure is moving to the boundary of the ground and away from it, forming the middle part or stage of the motion event.
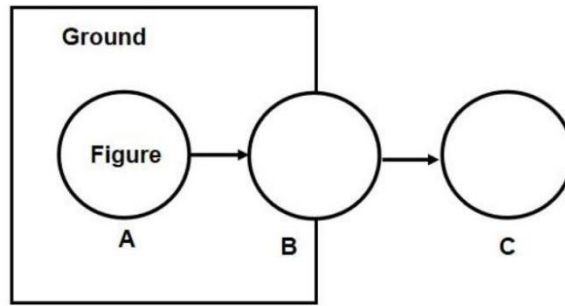
Figure 3: The Image Schema of *Chū* ("GO OUT")

According to human experience, their sight can follow a moving object when it moves towards an ending point along a certain path. When it stops, their focus may fall on the ending point. In the motion event of *chū*, when human follows the moving figure and situate their focus on the ending point, *chū* extends to the concept of ARRIVE and OUTSIDE, as in the expressions of *chūchǎng* 出场 "arrive and come on the stage" and *wǎng chū ná* 往出拿 "take something outside". When human locates their focus on the middle part B, that is, the figure overlaps the boundary of the ground and away from it, *chū* extends to the concept of PAY and PRODUCE, as in the expressions of *chū qián* 出钱 "pay the bill" and *chū réncái* 出人才 "produce talents". When human ignores the starting point and end point of the moving event, and only focus on the path of the moving event, *chū* extends to the concept of EXCEED, as in the expression of *bù chū bàn gè yuè* 不出半个月 "in a half month". And when human locates their focus both on the path and the ending point of the motion event, *chū* extends to the concept of GO, as in the expression of *chū jiē* 出街 "go to the street".

The concepts of GO OUT, ARRIVE, OUTSIDE, PAY, PRODUCE, EXCEED, and GO can evoke the same motion event of *chū*, but they profile different facets of it. GO OUT profiles the whole motion event, ARRIVE profiles the ending point of the event. PAY and PRODUCE profile the middle part of the motion event. EXCEED profiles the path of the event and GO profiles both the path and ending point of the event. As these extended concepts of *chū* profile different elements within the same moving event and shift their profiles in the same conceptual base, the concepts of ARRIVE, OUTSIDE, PAY, PRODUCE, EXCEED, and GO are metonymically associated with GO OUT. Different profiling of *chū* can be represented in Table 5.

Table 5: Different profiling of *chū*

Focus	Extended Concept
ending point	ARRIVE
middle part	PAY
middle part	PRODUCE
path	EXCEED
path and ending point	GO

5 Conclusion

Previous studies of *chū* mainly focus on its polysemy, syntactic restriction, and diachronic evolution in Mandarin Chinese. Less attention has been paid to its colexification and colexificational mechanism. Meanwhile, even similar exploration of *chū* in Chinese topolects has received no attention. Based on the Mandarin Chinese data from BCC Chinese Corpus and the Modern Chinese Dictionary (7th edition) (Institute of Linguistics, 2016), and Chinese topolectal data from a large dictionary of modern Chinese topolects (43 volumes) (Li, 2012). Designating the basic concept of GO OUT as the pivot concept, this present paper has identified 9 colexified pairs in Modern Chinese. Using the apparatus of CLICS³, this present paper also acquires some cross-linguistic evidence to some degree, to justify the colexifications for *chū* (GO OUT) across topolects. According to the cross-topolectal evidence, the colexificational network of *chū* (GO OUT) in Modern Chinese is constructed.

This present paper has also explored the mechanisms for the colexified extensions of *chū*, that is, conceptual metaphor and metonymy. The metaphoric colexified extensions of *chū* can map onto the space domain, time domain, and state domain. Specifically, the extended concept of TAKE OUT maps onto the space domain, the extended concept of COME TO AN END maps from the space domain to the time domain, and the extended concept of HAPPEN maps from the space domain to the state domain. All the above metaphoric colexified extensions follow the order of "space > time > state". As for the metonymic colexified extensions, the extended concepts of ARRIVE, PAY, PRODUCE, EXCEED, and GO can profile different substructures (i.e. ending point, middle part, path) of the image schema in the motion event of chū. This paper elaborates upon the types of metaphoric and metonymic colexified extensions in this manner that will give important implications for studies of other directional verbs.

Acknowledgement

References

Chu, Z. X., Xu, C. H., He, F. L., Huang, C. P., & Yin, D. Z. (1999). The "V *chū* +N *wài*" pattern of modern Chinese-and why is this pattern rare in modern times. *Ancient Chinese Studies*, 4, 26-29.

Croft, W., & Cruse, D. A. (2004). *Cognitive linguistics*. Cambridge: Cambridge University Press.

Evans, V. (2004). *The structure of time: Language, meaning and temporal cognition.* Amsterdam & Philadelphia: John Benjamins.

François, A. (2008). Semantic maps and the typology of colexification. In V. Martine (Ed.), *From polysemy to semantic change: Towards a typology of lexical semantic associations* (pp. 163-215). Amsterdam and Philadelphia: John Benjamins Publishing Company.

Gast, V., & Koptjevskaja-Tamm, M. (2022). Patterns of persistence and diffusibility in the European lexicon. *Linguistic Typology*, 26(2), 403-438.

Gibbs, R., & Matlock, T. (1997). Psycholinguistic perspectives on polysemy. In H., Cuyckens. & B. Zawada (eds.). *Polysemy in Cognitive Linguistics* (pp. 213-237). Amsterdam and Philadelphia: John Benjamins Publishing Company.

Heine, B., Ulrike, C., & Friederike, H. (1991). *Grammaticalization: A conceptual framework.* Chicago: Universal of Chicago Press.

Hu, W., Gan, Y. W., & Chen, Y. F. (2019). Directional verbs of Guangzhou Nansha Cantonese "bury, open, rise, fall, and leave". *Journal of Jiaying College*, 5, 85-88.

Hu, X. Y. (2012). *A Study of Grammaticalization of Chinese directional verbs*. Guangxi: Guangxi Normal University.

Institute of Linguistics, CASS. (2016). *Modern Chinese Dictionary* (7th Edition) (Chinese Edition). The Commercial Press.

Joshua, J., Jeseph, W., & Henry, T. (2019). Emotion semantics show both cultural variation and universal structure. *Science*, 366(6472): 1517-1522.

Lakoff, G. (1987). *Women, fire and dangerous things: what category reveal about the mind*. Chicago and London: University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: The University of Chicago Press.

Lakoff, G., & Johnson, M. (2003). Why cognitive linguistics require embodied realism. *Cognitive Linguistics*, *13*(3), 245-264.

Langacker, R. W. (1987). *Foundations of cognitive grammar*. Stanford: Stanford University Press.

Langacker, R. W. (2013). *Essentials of cognitive grammar*. Oxford: Oxford University Press.

Lehrer, A. (1992). A theory of vocabulary structure: Retrospectives and prospectives. *Thirty Years of Linguistic Evolution*, 243-256.

Li, R. (Ed.). (2012). Great dictionary of modern Chinese dialects (43 Volumes). Jiangsu Education.

Liu, F. (2015). *The evolution of directional verbs*. Chengdu: Southwest Jiaotong University Press.

Liu, Y. H. (1998). *Directional complement interpretation*. Beijing: Beijing Language and Culture University Press.

Rzymski, C., & Tresoldi, T. et al. (2019). *The Database of Cross-Linguistic Colexifications, reproducible analysis of cross-linguistic polysemies*. DOI: 10.1038/s41597-019-0341-x. https://clics.clld.org/

Shi, X. Y. (2001). The morpheme meaning analysis of *chū* and *jìn:* A re-exploration on the basis of the interpretation of Modern Chinese Dictionary. *Lexicography*, 3, 60-66.

Talmy, L. (2000). *Toward a cognitive semantics*. Massachusetts: MIT Press.

Wang, G. Q. (2005). *Research on directional verbs*. Beijing: Huaxia Press.

Wu, C. X. (1997). The use of the word *chū* on oracle bone inscriptions. *Ancient Chinese Studies*, 1, 29-31.

Wu, F. X. (2003). A typological study of grammaticalization of the comitative preposition in Chinese language. *Studies of the Chinese Language,* 1, 43-58.

Xi, Z. H., & Zhang, Z. X. (1990). *Language Atlas of China* (2nd Edition). Shanghai: The Commercial Press.

Zeng, C. L. (2009). The type and property of motion event in Chinese. *Journal of Kashgar Teachers College*, 4, 48-51.

Zhou, S. Y. (2018). *History of Chinese words*. Beijing: China Renmin University Press.

The Structure of Chinese Compounds: The Perspective of Predicative Implicitness

YANG Yongzhong
Yunnan University of Finance and Economics, China
wmyoung@sina.com

Abstract

This paper examines Chinese nominal compounds with respect to their internal structure, thematic relations, generation process, and constraint mechanism from the perspective of predicate implicitness. Findings reveal that constituent functions in these compounds vary based on their structural positions, closely aligning morphological and syntactic structures. Predicate implicitness necessitates hierarchical adjunction, disallowing cross-layered adjunction. Corresponding relations exist between theta-roles, semantic relations, syntactic, and morphological structures. The study delineates differences between agentive and non-agentive compounds and explores how syntactic structure influences morphological structure. It also delves into theta-roles, argument structure, and linear order, arguing that constituent word order adheres to the Prominence and Locality Principles, dictated by their syntactic hierarchy positions.

Keywords: compound, predicate implicitness, morphological structure, syntactic structure, theta-role

Povzetek

Članek obravnava kitajske pridevniške tvorjenke glede na njihovo notranjo strukturo, tematske odnose, proces nastajanja ter mehanizem omejitev z vidika povedkovne implicitnosti. Ugotovitve razkrivajo, da se funkcije komponent v teh tvorjenkah razlikujejo glede na njihove strukturne položaje, tesno se prilagajajo morfološkim in skladenjskim strukturam. Implicitnost povedka zahteva hierarhično urejen primik (adunkcijo), križni primiki niso dovoljeni. Obstajajo ujemajoči se odnosi med theta-vlogami, semantičnimi odnosi, skladenjsko in morfološko strukturo. Študija razmejuje razlike med udeleženskimi in neudeleženskimi tvorjenkami ter raziskuje, kako skladenjska struktura vpliva na morfološko. Prav tako se poglobi v theta-vloge, strukturo argumenta in linearni red ter ugotavlja, da se vrstni red komponent prilagaja načelom izpostavljenosti in lokalnosti, ki jih določajo njihovi položaji v skladenjski hierarhiji.

Ključne besede: tvorjenka, povedkovna implicitnost, morfološka struktura, skladenjska struktura, theta-vloga

1   Introduction

The argument structure of Chinese nominal compounds and their way of generation, the constraints involved in their internal structure have recently been hot topics in the circle of Chinese linguistics and especially in the circle of generative grammar (Gu & Shen, 2001; Shi, 2003; He, 2004; Cheng, 2005; He & Wang, 2005; Yang, 2006, 2015, 2016a; Zhou, 2006; Tang, 2008, 2014; Zhuang & Liu, 2011, among others). It is generally believed that the morphological structure of Chinese is closely associated with its syntactic structure and hence both the morphological structure and the syntactic structure of Chinese are basically consistent. In this case, the structure of Chinese nominal compounds can be regarded as a micro syntactic structure (Chao, 1968, p. 189-243; Lu, 1964, p. 2; Ren, 1981, p. 134-135; Zhu, 1982, p. 33; Huang & Liao, 2007, p. 8; Tang, 2008). This approach has a strong power of explanation, for it can account for many language facts. It, however, meets with some issues. When the predicate occurs, the SV compound is often ungrammatical, whereas the VO compound is grammatical. When the predicate does not occur, the compound in which the agent precedes the patient is ungrammatical, whereas the compound in which the patient precedes the agent is grammatical, as illustrated in (1).

(1) 工人　　　开采　　石油　　　→ a. 石油　　　开采
　　gongren　kaicai　Shiyou　　　 shiyou　　kaicai
　　worker　 extract　Petroleum　 petroleum　extract
　　'The worker extracts petroleum.'　'petroleum extraction'

　　　　　　　　　　　　　　　　 → b.* 工人　　　石油
　　　　　　　　　　　　　　　　　　　gongren　shiyou
　　　　　　　　　　　　　　　　　　　worker petroleum

　　　　　　　　　　　　　　　　 → c. 石油　　　工人
　　　　　　　　　　　　　　　　　　 shiyou　　gongren
　　　　　　　　　　　　　　　　　　 petroleum　worker
　　　　　　　　　　　　　　　　　　 'petroleum worker'

If non-agentive theta-roles[1], such as instrumental, temporal, and local, function as the predicate[2], two circumstances may arise. When the predicate occurs, the SV compound is ungrammatical, but only the VO compound is grammatical. When the predicate does not occur, the compound in which the patient precedes the non-agent is ungrammatical, however, the compound in which the non-agent precedes the patient is grammatical[3], as illustrated in (2).

(2) 红木　　　　制造　　　家具　　→ a.　家具　　　　制造
　　　hongmu　　zhizao　　Jiaju　　　　　jiaju　　　　zhizao
　　　rosewood　make　　　Furniture　　　furniture　make
　　　'The furniture is made of rosewood.'　　　'furniture making'

　　　　　　　　　　　　　　　　　→ b.　红木　　　　家具
　　　　　　　　　　　　　　　　　　　　hongmu　　jiaju
　　　　　　　　　　　　　　　　　　　　rosewood　furniture
　　　　　　　　　　　　　　　　　　　　'rosewood furniture'

　　　　　　　　　　　　　　　　　→ c.*　家具　　　　红木
　　　　　　　　　　　　　　　　　　　　jiaju　　　　hongmu
　　　　　　　　　　　　　　　　　　　　furniture　rosewood

Obviously, the data in (1) and (2) constitute a striking contrast. It seems that there are differences between (1) and (2) in terms of structure and

---

[1] The terminology used in this article is as follows: *Agent:* the person or thing carrying out the action; *Patient:* the person or thing affected by the action; *Theta-roles:* semantic roles; *Agentive theta-roles:* the theta-roles related to the agent; *Non-agentive theta-roles:* the theta-roles unrelated to the agent; *Instrumental:* a semantic role indicating an instrument used for some purpose; *Local:* a semantic role indicating positions or movements in space; *Temporal:* a semantic role indicating or involving time; *Oblique:* a syntactic element accompanying a verb which is not a subject or object, or the equivalent.

[2] Larson (1988), Baker (1988), and Grimshaw (1990) argue that the theta-roles are arranged hierarchically. Therefore, the assignment of the theta-roles follows the top-to-bottom order.

[3] An anonymous reviewer claims that the two SVs 红木雕刻艺术品 *hongmu diaoke yishupin* or 中国制造艺术品 *zhongguo zhizao yishupin* in an SVO sentence with the O 艺术品 *yishupin* are perfectly acceptable. However, we argue that the two constructions are not counterexamples to our analysis. 红木雕刻艺术品 *hongmu diaoke yishupin* means *yong hongmu diaoke de yishupin*, i.e. artwork that is carved of rosewood. 红木 *hongmu* is an instrumental instead of an agent. Nor is it the agentive subject of the construction. In the same, 中国制造艺术品 *zhongguo zhizao yishupin* means *zai zhongguo zhizao de yishupin*, artwork which is made in China. In this case, 中国 *zhongguo* is a local instead of a patient. Nor is it the agentive subject of the construction. As a matter of fact, neither 红木雕刻艺术品 *hongmu diaoke yishupin* nor 中国制造艺术品 *zhongguo zhizao yishupin* is an SVO sentence.

grammaticality. Moreover, the analysis of the syntactic structure seems to influence the morphological structure. How to account for the theta-roles of the morphological structure of nominal compounds, the argument structure, and the linear ordering of the surface structure in the framework of generative grammar is worthy of further research.[4]

The goal of the paper is to conduct research into the internal structure of nominal compounds and the thematic relations between various constituents in terms of predicate implicitness, the way to generate nominal compounds, and the constraints that they are subject to.

The paper is organized as follows. Section 2 addresses the concept of predicate implicitness and its syntactic-semantic representations. Section 3 addresses the interrelation between predicate implicitness, the way of generation of nominal compounds and their types of structure. Section 4 concludes the paper.

2   The concept of predicate implicitness and its syntactic-semantic representations

Predicate implicitness means that in the transformation of a sentence into a compound, the predicate must be structurally implicit, thereby giving rise to a covert form of the compound. To put it differently, the predicate is involved in the lexical-semantic representations of nominal compounds but it is covert phonologically, and its occurrence in nominal compounds can be attested or recovered by grammatical evidence. In this case, the predication relation between the nouns is completely preserved through the semantic linking.

Predicate implicitness is a necessary condition on which two nouns in the syntactic structure co-occur in the morphological structure. If the predicate occurs in the morphological structure, then, only VO compounds can be generated. In this case, only the noun which bears the role of an object can occur. Only when the predicate does not occur in the morphological structure, can the compound containing two nouns (viz. agent and patient) be generated. In this case, the semantic relationship between the nouns is characterized by a modifier head. To understand their semantic relationship, we must activate the implicit predicate following the semantic linking. Take the compound *hongmu jiaju* 红木家具 (lit. rosewood furniture, i.e. rosewood

---

[4] Yuan (1995) addresses the referential rules of *de* 的 constructions in terms of such semantic roles as agent, experiencer, patient, resultative, instrumental, temporal, local, etc. However, it fails to address Chinese nominal compounds in terms of these semantic roles.

furniture) for example. To understand it, we must activate the implicit predicate *zhizao* 制造 (lit. make), thereby obtaining the semantics of *hongmu zhizao jiaju* 红木制造家具 (lit. rosewood make furniture, i.e. The furniture is made of rosewood). When more than one predicate occurs implicitly, ambiguity may arise. Take the compound *shiyou gongsi* 石油公司 (lit. petroleum company, i.e. petroleum company) for example. It may activate such implicit predicates as *kantan* 勘探 (lit. explore, i.e. explore), *jinglian* 精炼 (lit. refine, i.e. refine), *xiaoshou* 销售 (lit. sell, i.e. sell), etc. How to determine the predicate depends on the specific context. It follows that under this circumstance the specific context and the knowledge background of the addresser and the addressee are vital to the elimination of semantic ambiguity.[5] The implicit predicate functions in the deep semantic structure and the surface syntactic structure. Hence it is crucial to the semantic interpretation and the structural building of compounds (cf. Yuan, 1995).

Predicate implicitness causes the agent to be merged with the patient directly. Then, how to arrange them is an issue that is worth probing. According to the syntactic structure, the agent is supposed to precede the patient, thereby giving rise to a compound in which the agent precedes the patient. Such a compound, however, cannot be found in language reality. In contrast, only a compound in which the patient precedes the agent is allowed in the morphological structure. In this case, the patient functions as the modifier of the agent, as illustrated in (3)-(5).

(3) 教师　　　教　　　英语　　→　英语　　教师
　　jiaoshi　　jiao　　yingyu　　　yingyu　jiaoshi
　　teacher　teach　English　　　English　teacher
　　'The teacher teaches English.'　　'English teacher'

　　　　　　　　　　　　　　　　→* 教师　　英语
　　　　　　　　　　　　　　　　　　jiaoshi　yingyu
　　　　　　　　　　　　　　　　　　teacher　English

---

[5] The cognitive principle of relevance accounts for the licensing and interpretation mechanisms of implicit predicates and manners of co-composition. A predicate can occur lexically unrealized only if the same contextual effects can be reached using and interpreting them as in the case of overt arguments or predicates but with less processing effort. The cognitive principle of relevance also motivates manners of co-composition in such a way that the meanings of arguments influence the activation of the meaning relevant from the potential meanings of predicates. The interpretation involves lexical-semantic representations and contexts. Therefore, a predicate can be left implicit or an implicit predicate can be recovered in the immediate contexts of utterances containing this predicate according to its grammatical characterization (Németh T. & Bibok, 2010).

(4) 故事　　描写　　　战争　　　→　战争　　　故事
　　gushi　miaoxie　zhanzheng　　zhanzheng gushi
　　story　describe　war　　　　　War　　　story
　　'The teacher teaches English.'　　'war story'

　　　　　　　　　　　　　　　→* 故事　　　战争
　　　　　　　　　　　　　　　　 gushi　　zhanzheng
　　　　　　　　　　　　　　　　 story　　war

(5) 工人　　　开采　　　石油　　→　石油　　　工人[6]
　　gongren　kaicai　shiyou　　　shiyou　　gongren
　　worker　　extract　petroleum　petroleum worker
　　'The worker extracts petroleum.'　'petroleum worker'

　　　　　　　　　　　　　　　→* 工人　　　石油
　　　　　　　　　　　　　　　　 gongren　shiyou
　　　　　　　　　　　　　　　　 worker　　petroleum

If the noun preceding the predicate is a non-agentive constituent, only the compound in which the non-agentive constituent precedes the patient is allowed in the morphological structure. In this case, the patient serves the function of the noun head, as illustrated in (6)-(8).

(6) 红木　　　　制造　　家具　　→　红木　　　家具
　　hongmu　　zhizao　Jiaju　　　hongmu　jiaju
　　rosewood　make　Furniture　　rosewood furniture
　　'The furniture is made of rosewood.'　'rosewood furniture'

　　　　　　　　　　　　　　　→* 家具　　　红木
　　　　　　　　　　　　　　　　 jiaju　　　hongmu
　　　　　　　　　　　　　　　　 furniture　rosewood

(7) 绍兴　　　　出产　　黄酒　　→　绍兴　　　黄酒
　　Shaoxing　chuchan　Huangjiu　Shaoxing huangjiu
　　Shaoxing　produce　rice-wine　Shaoxing rice-wine
　　'Rice wine is produced in Shaoxing.'　'Shaoxing rice-wine'

　　　　　　　　　　　　　　　→* 黄酒　　　绍兴
　　　　　　　　　　　　　　　　 huangjiu　Shaoxing
　　　　　　　　　　　　　　　　 rice-wine　Shaoxing

---

[6] For the sake of consistency of numbering and convenience of reading, several examples are repeated (e.g. (1) and (5); (2) and (6), etc.) by introducing a new number instead of the previous one.

(8) 今日　　　出版　　　　报纸　　　　　→　今日　　　　报纸
　　 jinri　　　chuban　　 Baozhi　　　　　 jinri　　　　baozhi
　　 today　　 publish　　 Newspaper　　　　today　　　newspaper
　　 'The newspaper was published today.'　　　'today's newspaper'

　　　　　　　　　　　　　　　　　　　　　→* 报纸　　　　今日
　　　　　　　　　　　　　　　　　　　　　　 baozhi　　　jinri
　　　　　　　　　　　　　　　　　　　　　　 newspaper today

Comparing the nouns that occupy the subject position in (6)-(8), we find that the agentive theta-roles in (6) serve the function of the subject of the sentence, whereas the non-agentive theta-roles in (7) and (8) function as the adverbial of the sentence though they occur in the position preceding the predicate. Hence the non-agentive theta-roles can be understood as the noun functioning as the adverbial alone, or it, together with the preposition which occurs covertly, functioning as the adverbial.[7] Then, we have to account for why agentive theta-roles can function only as the noun head instead of the modifier in the morphological structure while non-agentive theta-roles can function only as the modifier instead of the noun head in the morphological structure. We argue that given constituent extraction, all types of theta-roles, including agent, patient, instrumental, temporal, and local, can be extracted. However, if the structure contains the agent, only the noun phrase out of which the agentive theta-role is extracted can be operated further. To put it differently, both the predicate and the phrase marker *de* 的[8] can be implicit, thereby giving rise to a modifier-head compound. In contrast, neither the predicate nor the phrase marker *de* 的 can be implicit concerning the noun phrase out of which a patient is extracted. As a consequence, no

---

[7] At present, there are two analyses of the preposition implicitness. One analysis argues that it is due to the fact the preposition is compressed by the verb and the subsequent noun and hence it weakens phonologically, thereby giving rise to its disappearance. In this case, it is similar to an empty category in terms of syntax. Another analysis claims that the disappearance of the preposition is not due to deletion or ellipsis. Instead, it results from the incorporation of the preposition and its object (Yang, 2007a, 2007b, 2009).

[8] *De* 的 is a linking marker. It is used to link the preceding constituent with the following constituent. The preceding constituent bears the feature [+N] and functions as the agent, patient, or possessor while the following constituent bears the feature [-V] or [+N]. *De* 的 is used to show the difference between the phrase and the sentence and to mark the symmetry between them. It can be adjoined to the Spec or complement. Given syntactic distribution, it often functions as the adjoined constituent of the Spec. It occurs behind the personal pronoun, proper name, or verb, but it cannot follow the demonstrative pronoun, because the demonstrative pronoun usually functions as the syntactic head and the genitive property of *de* 的 determines its usage as a clitic that follows a personal pronoun or proper name. It cannot occur behind the demonstrative pronoun to denote specificity (Yang, 2014a).

modifier-head compound can be generated. Rather, compounds deriving from such theta-roles are ungrammatical, as shown in (3)-(8). If there is no agent in the structure, then, only the noun phrase out of which the agentive theta-roles are extracted allows the occurrence of the implicit predicate and the non-occurrence of the phrase marker *de* 的, thereby giving rise to a modifier-head compound. In contrast, the noun phrase out of which the adjuncts, including instrumental, temporal, and local, are extracted does not allow the occurrence of the implicit predicate and the non-occurrence of the phrase marker *de* 的, as shown in (9) and (10).

(9) 教师　　　教　　　英语
　　 jiaoshi　　Jiao　　yingyu
　　 teacher　　Teach　　English
　　 'The teacher teaches English.'

　　 → a. 教师　　　教　　　的　　　英语　　　→* 教师　　　英语
　　　　 jiaoshi　　jiao　　de　　　Yingyu　　　 jiaoshi　yingyu
　　　　 teacher　　teach　　AUX　　English　　　 teacher　English
　　　　 'English which the teacher teaches'

　　 → b. 教　　　英语　　　的　　　教师　　　→ 英语　　　教师
　　　　 jiao　　yingyu　　de　　　Jiaoshi　　　 yingyu　jiaoshi
　　　　 teach　　English　　AUX　　Teacher　　　 English　teacher
　　　　 'the teacher who teaches English'　　　 'English teacher'

(10) 红木　　　　制造　　　家具
　　　 hongmu　　zhizao　　Jiaju
　　　 rosewood　　make　　Furniture
　　　 'The furniture is made of rosewood.'

　　 → a. 红木　　　　制造　　的　　家具　　　→ 红木　　　　家具
　　　　 hongmu　　zhizao　de　　Jiaju　　　　 hongmu　　jiaju
　　　　 rosewood　make　　AUX　Furniture　　　 rosewood　furniture
　　　　 'furniture made of rosewood'　　　　 'rosewood furniture'

　　 → b. 制造　　家具　　　的　　红木　　　→* 家具　　　红木
　　　　 zhizao　Jiaju　　　de　　hongmu　　　 jiaju　　hongmu
　　　　 make　　furniture　AUX　rosewood　　　 furniture　rosewood
　　　　 'rosewood for making furniture'

According to (9) and (10), the way to generate nominal compounds can be summarized as follows.

(11) a. NP1 V NP2 →V NP2 *de* NP1 →NP2 NP1 (NP1=agent)
    b. NP1 V NP2 →NP1 V *de* NP2 →NP1 NP2 (NP1= instrumental, temporal, and local)

Predicate implicitness must cause the two nominal constituents to be merged directly. However, only the morphological structure of the modifier-head compound containing the non-agentive theta-roles conforms with the syntactic structure with respect to the surface order, whereas the morphological structure of the modifier-head compound containing the agentive theta-roles is in contrast to the syntactic structure of the surface order. Then, whether the generation pattern of nominal compounds in (11) can be accounted for in the framework of generative grammar is an issue that we shall address in the following section.

3   Predicate implicitness, the generation of nominal compounds and the types of their structure

If we make a further observation of the internal structure of the modifier-head compound, we may find that both the structure of the compound containing the agentive theta-role and the structure of the compound containing the non-agentive theta-role are head-final. The former is headed by the agent, whereas the latter is headed by the patient. The noun that occurs in the predicate position is actually not the head of the structure, but a peripheral constituent, and hence it can be regarded as an adjunct. To put it differently, the argument structure of the compound is quite in conformity with its original syntactic structure.

(12)

Syntactic structure	Morphological structure	
N1 + N2	N1 + N2	
head + non-head	*head + non-head	compound with agentive theta-roles
non-head + head	non-head + head	compound with non-agentive theta-roles
subject + object	adverbial + object	

As (12) shows, the syntactic structure is quite consistent with the morphological structure. In the compound with agentive theta-roles, the head precedes the non-head. N1 bears semantic weight, whereas N2 functions as a complement. Therefore, N1 + N2 amounts to N + Complement, viz. NC. In this case, the morphological structure is completely in conformity with the syntactic structure. In the same vein, in the compound with non-agentive theta-roles, the non-head precedes the head. N1 represents

possession, property, state, material, purpose, time, location, etc. N2 bears semantic weight and hence N1 + N2 amounts to Adjective + N, viz. A + N. The semantic weight and the syntactic weight of the two types of compounds are completely consistent. Therefore, the syntactic structure is consistent with the morphological structure. The above conclusions can be summarized as follows.

(13) a. Syntactic / Morphological structure: $[_{NP} N1[_{N'} N2]]$
　　　　 Semantic relation:　　　　　　　　 N + C

　　 b. Syntactic / Morphological structure: $[_{NP}[_{N'} N1]N2]$
　　　　 Semantic relation:　　　　　　　　 A + N

As shown in (13), N1 asymmetrically c-commands N2 but N2 cannot c-command N1.[9] Therefore, the determiner that is characterized by definiteness is supposed to occur in the highest position of DP. It precedes either the combination NP1 + NP2 or the combination NP2 + NP1. It cannot occur between NP1 and NP2. Nor can it occur between NP2 and NP1. Otherwise, the Lexical Integrity Hypothesis would be violated.[10]

At the syntactic level, NP1 occurs in the subject position and hence it is a dominating constituent. In contrast, NP2 occurs in the object position and hence it is a subordinate constituent. There is a predicate intervening between them. At the phrase level, NP1 which is extracted out of the construction is still a dominating constituent. NP2 is still a subordinate constituent though it precedes NP1. In effect, there still exists a predicate intervening between them. In this case, the predicate cannot be implicit. Otherwise, the syntactic-semantic relation between them would not be clearly expressed. At the morphological level, the predicate is implicit. NP2, which functions as the modifier of NP1, occurs in the position preceding NP1. NP2, however, is still a subordinate constituent, whereas NP1 is still a dominating constituent. It is suggested that if the agent precedes the patient, the predicate must occur at all levels, including the syntactic level, the phrase level, and the morphological level. Otherwise, the construction would be ungrammatical. In contrast, if the patient precedes the agent, the predicate

---

[9] According to the Linear Correspondence Axiom (Kayne, 1994, p. 33) and the Principle of Category Order (Dai, 2003), if an arbitrary constituent X c-commands another arbitrary constituent Y, then, Y cannot c-command X.

[10] According to the Lexical Integrity Hypothesis, syntactic operations and semantic interpretations cannot influence the subcomponent of a word (Jackendoff, 1972; Selkirk, 1984).

must occur only at the syntactic level and the phrase level, whereas at the morphological level, it must be implicit.

(14) a. Syntactic level: *agent + patient   (without a predicate)
       e.g.* 去年   工人   石油
             qunian   gongren   shiyou
             last year Worker   petroleum

   b. Phrase level: *agent + patient   (without a predicate)
       e.g.* 工人   的   石油
             gongren   De   shiyou
             worker   AUX   petroleum

   c. Morphological level: *agent + patient (without a predicate)
       e.g.* 工人   石油
             gongren   shiyou
             worker   petroleum

(15) a. Syntactic level: *patient + agent   (without a predicate)
       e.g.* 去年   石油   工人
             qunian   shiyou   gongren
             last year petroleum worker

   b. Phrase level: *patient + agent   (without a predicate)
       e.g.* 石油   的   工人
             shiyou   De   gongren
             petroleum AUX   worker

   c. Morphological level: patient + agent (without a predicate)
       e.g.* 石油   工人
             shiyou   gongren
             petroleum worker

   The representations of the theta-roles such as instrumental, temporal, and local are similar to the theta-role of an agent at the syntactic level, the phrase level, and the morphological level. Hence it is unnecessary to go into details. The predicate must occur at the syntactic level and the phrase level, whereas it must be implicit at the morphological level.

(16) Syntactic level   Morphological level
    a. agent + predicate + *de* + patient   →*agent + patient
    b. non-agent + predicate + *de* + patient → non-agent + patient
    c. predicate + patient + *de* + agent   → patient + agent
    d. predicate + patient + *de* + non-agent →*patient + non-agent

At the phrase level, the construction is grammatical regardless of whether the extracted constituent is a patient, agent, or non-agent. At the morphological level, however, the agent can only occur behind the patient and the non-agent must precede the patient. Moreover, linear order cannot be reversed, as illustrated in (17) and (18).

(17) a. 工人   开采   的   石油   →* 工人   石油
        gongren   kaicai   de   shiyou   gongren shiyou
        worker   extract AUX petroleum   worker   petroleum
        'petroleum extracted by the worker'
    b. 红木   制造   的   家具   → 红木   家具
        hongmu   zhizao   de   jiaju   hongmu jiaju
        rosewood make   AUX furniture   rosewood furniture
        'furniture made of rosewood'   'rosewood furniture'

(18) a. 开采   石油   的   工人   → 石油   工人
        kaici   shiyou   de   gongren   Shiyou   gongren
        extract petroleum AUX worker   petroleum worker
        'worker who extracts petroleum'   'petroleum worker'
    b. 制造   家具   的   红木   →* 家具   红木
        zhizao jiaju   de   furniture   Jiaju   hongmu
        make   furniture   AUX rosewood   furniture rosewood
        'rosewood for making furniture'

The patient can function as the modifier of the agent only, but it cannot function as the modifier of the non-agent. The agent cannot function as the modifier of the patient, but the non-agent can function as the modifier of the patient. However, there remain some questions. We have to account for why the agent can function as a noun head instead of a modifier at the morphological level while the non-agent can function as a modifier instead of a noun head. We argue that the reason lies in the morphological structure of compounds. According to Sportiche (1988), Kuroda (1988), Larson (1988, 1990), Cheng (1999, p. 239-245), and Yang (2007a, 2011, 2012, 2014b, 2016a, 2016b), the assignment of theta-roles is locally constrained. The predicate must assign theta-roles to the arguments within its projection. Therefore, the argument must occur within the maximal projection of the predicate. The

theta-role of the external argument is not assigned by the predicate but by the maximal projection of the predicate. According to the X-bar theory, a maximal projection can have only one specifier and one complement. In order to satisfy these conditions, the lexical representation of compounds with the external argument adopts the form, as shown below.

(19) [$_{VP1}$ NP1[$_{V'}$ V1[$_{VP2}$ NP2[$_{V'}$ V2 NP3]]]]

As (19) shows, when an external argument occurs, there will be a null predicate in the representation. Its complement is the maximal projection of the subject, namely, the structural representation with various internal arguments. NP1 is an external argument and NP3 is an internal argument. They are assigned theta-roles by VP and V, respectively. V1 is a null predicate while V2 is a major predicate. The representation of the external argument entails an empty argument position (i.e. NP2) and an empty predicate position (i.e. V1), for there is some asymmetry between the conceptual system and the syntactic system. The conceptual system cannot correspond to the syntactic structure until it has been conceptualized.[11] It follows that the null predicate is set to satisfy the requirement of the conceptual system and the syntactic system simultaneously.[12] Therefore, the structure of the constructions *gongren kaicai shiyou* 工人开采石油 and *hongmu zhizao jiaju* 红木制造家具 should be analyzed as follows.

(20) [$_{VP1}$工人[$_{V'}$ V1 [$_{VP2}$ NP2[$_{V'}$ 开采   石油]]]]
　　　　gongren　　　　　　　kaicai shiyou
　　　　worker　　　　　　　　extract petroleum

　→[$_{VP1}$工人[$_{V'}$ V1[$_{VP2}$石油 $_i$[$_{V'}$ φ t$_i$]]]]
　　　　gongren　　shiyou
　　　　worker　　　petroleum

---

[11] Jackendoff (1990) argues that the conceptual structure corresponds to the syntactic structure. Based on this argument, Tai (2002) claims that semantics which the syntactic structure can express is abstract and simplified after having been conceptualized instead of rich semantics containing the conceptual system.

[12] When an external argument occurs, there will be a null predicate in the representation. Its complement is the maximal projection of the subject, namely, the structural representation with various internal arguments. Since there are theta-roles in the conceptual system, the syntactic system must provide a null predicate to realize symmetry between the conceptual system and the syntactic system.

(21) [$_{VP1}$ NP1 [$_{V}$ V1[$_{VP2}$ 红木 [$_{V}$ 制造   家具]]]]
                 hongmu   zhizao Jiaju
                 rosewood make furniture

   →[$_{VP1}$ NP1 [$_{V}$ V1[$_{VP2}$ 红木   家具 $_i$[$_V$ ϕ t$_i$]]]]
                 hongmu   Jiaju
                 rosewood Furniture

As (20) and (21) show, there are so many nodes between *gongren* 工人 and *shiyou* 石油 that they cannot form a close combination. In contrast, between *hongmu* 红木 and *jiaju* 家具 there is only one implicit node, which gives rise to a syntactic empty category as a result of the implicitness of the predicate. Therefore, *hongmu* 红木 and *jiaju* 家具 can be merged directly because a modifier must be externally merged with the category which it modifies. It is noteworthy that the agent and the patient cannot be merged directly to form a compound, whereas the non-agent and the patient can be merged directly to form a compound. The patient *jiaju* 家具 moves leftward to the position [NP2 VP2] and is merged with the non-agent *hongmu* 红木 which occupies the position NP2, thereby giving rise to the compound *hongmu jiaju* 红木家具. Though the patient *shiyou* 石油 can move leftward to occupy the empty position, it cannot be merged with the agent *gongren* 工人, thereby giving no rise to the compound *gongren shiyou* 工人石油. Following Gruber (2001), we argue that *shiyou gongren* 石油工人 can be regarded as the result of leftward movement of the patient object *shiyou* 石油, as shown in (22). *Shiyou* 石油 moves leftward from the position [O N`] and crosses over *gongren* 工人 that occupies the position [S NP], thereby giving rise to the modification of *gongren* 工人. If the construction contains a non-agent, the circumstances will be different. Since the non-agent occupies the position [N N`], the surface order can be generated without resorting to movement, as shown in (23).

(22) [$_{DP}$ D[$_{NP}$ 工人[$_{N}$ 石油]]]
          gongren shiyou
          worker   petroleum

   →[$_{DP}$ 石油 $_i$[$_{NP}$ 工人 [$_{N}$ N t$_i$]]]
          shiyou   gongren
          petroleum worker

(23) [$_{DP}$ D[$_{NP}$ S[$_{N}$ 红木   家具]]]
              hongmu   jiaju
              rosewood furniture

If there is a demonstrative in the construction and it occupies position D, then, *shiyou* 石油 moves from its base-generation position to the position [A AP], as illustrated in (24). In the same vein, the non-agent *hongmu* 红木 in the non-agentive compound moves from its base-generation position to the position [A AP], as illustrated in (25). Such movement operations are N-to-A movement, which is a category of head movement. The motivation of movement is that the noun head is modified by the demonstrative. The noun head moves to assign the genitive case to the agent noun that functions as the subject. (cf. Bernstein, 2001)

(24) [$_{DP}$ 那些[$_{AP}$ A[$_{NP}$ 工人[$_{N'}$ 石油]]]
          naxie   gongren shiyou
          those   worker   petroleum

    →[$_{DP}$ 那些[$_{AP}$ 石油 $_i$[$_{NP}$ 工人 [$_{N'}$ N t$_i$]]]
        naxie   shiyou   gongren
        those   petroleum worker

(25) [$_{DP}$ 那件[$_{AP}$ A[$_{NP}$ S[$_{N'}$ 红木   家具]]]
          najian   hongmu jiaju
          that-Cl   rosewood furniture

    →[$_{DP}$ 那件[$_{AP}$ 红木 $_i$[$_{NP}$ S[$_{N'}$ t$_i$ 家具]]]
        najian hongmu   jiaju
        that-Cl rosewood   furniture

As (24) and (25) show, the movement of the agent and the non-agent takes place to save the morphological structure so that it may not crash. It is suggested that the construction that is grammatical at the syntactic level and the phrase level can be ungrammatical at the morphological level. To put it differently, a grammatical syntactic structure can generate a grammatical phrase structure, but it is not sure to generate a grammatical morphological structure. Then, we have to account for what causes the situation. Though the constituents preceding and following the predicate can be extracted out of the syntactic structure containing the agentive theta-roles and generate a grammatical phrase structure, only at the morphological structure the phrase structure out of which the agent is extracted is grammatical, as shown in (26).

(26) Syntactic level Phrase level   Morphological level
    NP1 + V + NP2 →NP1 + V + *de* + NP2 →*NP1 + NP2 (NP1=agent, NP2=patient)
                          →V + NP2 + *de* + NP1 → NP2 + NP1   (NP1=agent, NP2=patient)

In contrast, both the non-agent and the patient can be extracted out of the syntactic structure containing the agentive theta-roles to generate a grammatical phrase structure. However, only the phrase structure out of which the patient is extracted can generate a grammatical morphological structure, as shown in (27).

(27) Syntactic level   Phrase level   Morphological level

NP1 + V + NP2   →NP1 + V + de + NP2 →NP1 + NP2 (NP1= instrumental, temporal, local, etc., NP2=patient)

→V + NP2 + de + NP1 →*NP2 + NP1 (NP1= instrumental, temporal, local, etc., NP2=patient)

If we simplify (26) and (27) further and ignore the differences between the theta-roles, we will have the structure, as shown in (28) below.

(28) Syntactic level   Morphological level

NP1 + V +NP2 →*NP2 + V + NP1 (NP1=agent)

NP1 + V + NP2 →*NP2 + V+ NP1 (NP1= instrumental, temporal, local, etc.)

Based on the observation of (28), we argue that though the predicate V can dominate the patient NP2, it cannot modify the agent NP1 and the non-agent NP1, thereby giving rise to an ungrammatical construction. Interdependent semantic features and high conventionality are reflected in the domain of government of the verb, i.e. the verb's command over the preceding noun and its modification of the subsequent noun. According to the Verb-Governing Rule (VGR), V governs N1 and modifies N2. If and only if V governs N1 and modifies N2, an OV compound can be grammatical (Yang, 2006). According to Fu (2004), the pattern V + N refers to an event, which is formed by extracting the constituent N out of the framework of V in order to qualify the connotation of V and to make the event V abstract. The constituent N, which has been extracted, is mainly an object. Rather, when people need to make a certain verb abstract and to reclassify it, they tend to qualify its connotation by means of an object and choose the pattern N + V to express it, the premise of which is that V can govern the object. Only those that have a strong power of government over their objects may give rise to the pattern N + V, i.e. NP1 + V + NP2 in the present paper. The construction process of the meaning of N + V is associated with the structure of the narration concept in people's minds, i.e. the general framework of the event V. The framework includes various roles related to event V in the outside world and people's knowledge of event V and its effect upon it. The verb governs its own event framework, in which the theta-role of V and the non-

argument role are occupied by nominal constituents. And the theta-role and the non-argument role are both part of the framework. N + V is the result of reference by means of the event framework. To put it differently, in the event framework of V, some constituents are extracted or projected and merged with V in accordance with a certain pattern so as to express a certain meaning and to refer to a certain object. It shows that N + V is often used to denote the argument relation, especially the object relation. Different semantic relations have their own optimal options. Object relations tend to choose N + V, but seldom choose V + N. Non-argument relations are just the opposite. It is suggested that different argument relations often select different structural forms (Yang, 2006). Consider the following data.

(29)   工人   开采   石油
        gongren   kaicai   shiyou
        worker   extract   petroleum
        N1 + V + N2

   → 工人   石油   开采
        gongren   shiyou   kaicai
        worker   petroleum   extract
        N1 + N2 + V

   → 石油   工人   开采
        shiyou   gongren   kaicai
        petroleum worker   extract
        N2 + N1 + V

   → 石油   工人
        shiyou   gongren
        petroleum worker
        N2 + N1

(30)   故事   描写   战争
        gushi   miaoxie   zhanzheng
        story   describe   war
        N1 + V + N2

   → 故事   战争   描写
        gushi   zhanzheng   miaoxie
        story   war   describe
        N1 + N2 + V

→ 战争　　　故事　　　描写
zhanzheng gushi　　　miaoxie
war　　　　story　　　describe
N2 + N1 + V

→ 战争　　　故事
zhanzheng gushi
war　　　　story
N2 + N1

(31)　　红木　　　制造　　　家具
　　　　hongmu　 zhizao　　Jiaju
　　　　rosewood make　　 furniture
　　　　N1 + V + N2

　　　→ 红木　　　家具
　　　　hongmu　 jiaju
　　　　rosewood furniture
　　　　N1 + N2

(32)　　今日　出版　　　报纸
　　　　jinri　 chuban　　 Baozhi
　　　　today publish　　 newspaper
　　　　N1 + V + N2

　　　→ 今日　报纸
　　　　jinri　 baozhi
　　　　today newspaper
　　　　N1 + N2

As the data in (29)-(32) show, the generation of agentive compounds undergoes the process of the object being preposed or shifted. It first moves to the left edge of VP, viz. the specifier position of VP. Then it continues to move to the left edge of NP to give rise to the surface order NP2 + NP1 + V. V drops off to give rise to the compound NP2 + NP1. In contrast, the generation of non-agentive compounds does not undergo the process of the object being preposed. It can give rise to the compound NP1 + NP2 by means of predicate implicitness. The transformation of NP1 + V+ NP2 into NP2 + NP1 suggests that the semantic relation between NP1 and NP2 has changed from predication into modification. Due to predicate implicitness, NP2 which functions as the object cannot be assigned an accusative case, and hence it must move. Since the object NP2 cannot be assigned an accusative case by the predicate V, it has to move from the object position to the specifier

position preceding the subject NP1 and functions as the modifier of NP1 in order to avoid violation of the syntactic constraints. The patient noun functions as the object in the syntactic structure, but it functions as the modifier in the morphological structure. It follows that the function of every constituent is determined by its position in the structure. Furthermore, the linear positions of the constituents in the surface structure are determined by the positions of the constituents in the underlying structure. Rather, they are determined by the relationship of the c-command arising from the merger. The asymmetrical c-command relationship deriving from merger determines the order of the constituents in compounds (cf. Dai, 2003, p. 93-94).

In terms of linear order, the order of every constituent in the morphological structure is just opposite to its order in the syntactic structure, as shown below.

(33) Syntactic order: agent > patient > instrumental/local/temporal, etc.
     Morphological order: patient > instrumental/local/temporal, etc. > patient > agent

(34) | 工人 | 开采 | 石油 |
     |------|------|------|
     | gongren | kaicai | Shiyou |
     | worker | extract | petroleum |
     | (agent) | | (patient) |

  → | 石油 | | 工人 |
    |------|--|------|
    | shiyou | | gongren |
    | petroleum | | Worker |
    | (patient) | | (agent) |

(35) | 制造 | 家具 | 用 | 红木 |
     |------|------|-----|------|
     | zhizao | Jiaju | yong | hongmu |
     | make | furniture | with | rosewood |
     | | (patient) | | (oblique) |

  → | 用 | 红木 | 制造 | 家具 |
    |-----|------|------|------|
    | yong | hongmu | zhizao | jiaju |
    | with | rosewood | make | furniture |
    | | (oblique) | | (patient) |
    | | 红木 | | 家具 |
    | | hongmu | | jiaju |
    | | rosewood | | furniture |
    | | (oblique) | | (patient) |

The oblique case is base-generated in the position behind the patient. It moves to the position preceding the predicate V to give rise to the surface syntactic structure. The patient and the oblique case move to the position preceding the agent, respectively. The marker of the oblique case drops off or incorporates with the object of the preposition, thereby giving rise to a compound. The patient and the oblique case move because predicate implicitness results in semantic indefiniteness. Another explanation is that the marker of the oblique case restrains the oblique constituent and obstructs the patient from moving to the position preceding the oblique case. As a result, it has to stay in situ, thereby giving no rise to such an ungrammatical construction as *jiaju hongmu zhizao* 家具红木制造. The marker of the oblique case (i.e. preposition) incorporates with the oblique constituent to form NP. NP obstructs the patient from moving leftward alone. But NP can move with the patient to the position preceding the agent in a pied-piping way.

(36) 红木　　　家具　　　制造　　　厂
　　　hongmu　Jiaju　　zhizao　　chang
　　　rosewood　furniture　make　　factory
　　　(oblique)　(patient)　(predicate)　(agent)
　　　'rosewood furniture making factory'

When the agent or the oblique case does not occur, the predicate can be merged with the patient to form VO compounds, such as *jiaju zhizao* 家具制造 and *shiyou kaicai* 石油开采. If the agent occurs, the predicate must be implicit. In this case, the agent and the patient can enter into morphology to give rise to nominal compounds. Since the agent is base-generated in the specifier position of the light verb projection, it is suggested that only when the structure contains a light verb, can the agent occur and be licensed (cf. Chomsky, 1995, pp. 219-394). Neither the structure of VO compounds nor the structure of SV compounds contains the light verb.[13] Therefore, the agent cannot occur. This is why the compounds, such as *gongren shiyou* 工人石油 and *gushi zhanzheng* 故事战争, are ungrammatical. Then, we have to account for the grammaticality of such compounds as *shiyou kaicai* 石油开采, *zhanzheng miaoxie* 战争描写 and *jiaju zhizao* 家具制造. In effect, these compounds are identical regarding their morphological structure. Specifically, they are all modifier-head compounds instead of SV compounds. The noun modifies the verb which has gerundized. To put it differently, the

---

[13] The light verb can occur only at the syntactic level and cannot enter into the internal structure of a word. Otherwise, the Lexical Integrity Hypothesis would be violated (Lin, 2001; Huang, 2005; Tang, 2008).

verb has undergone nominalization. As a consequence, it is characterized by nouns.

(37)   [ₙ N   V]]
              (object)   (verb)
     a. [ₙ 石油   开采]]
          Shiyou   kaicai
          petroleum extract
          'petroleum extraction'

     b. .[ₙ 战争   描写]]
          zhanzheng miaoxie
          War   describe
          'war description'

     c. [ₙ 家具   制造]]
          Jiaju   zhizao
          furniture   make
          'furniture making'

As (37) shows, N is adjoined to V to give rise to adjunction. N is an object. It is noteworthy that both *zhizao* 制造 and *miaoxie* 描写 have the feature [-V] because it has gerundized, and hence it can be modified by a noun or pronoun. This shows that the head determines the form of the complement. If the head is D, the complement can only be NP or VP with the feature [-V]. If the VP occupies the Spec position, the construction is generally ungrammatical unless the VP has the feature [-V] (cf. Yang, 2010). The whole construction bears the nominal feature because the verb has lost some verbal features. In the same vein, if N1 is merged with N2, a nominal construction will be generated, as shown below.

(38) a. [ₙ N1   N2]]   b. [ₙ N1   N2]]
            (object) (subject)   (oblique) (object)

In terms of (38a), the object N1 is adjoined to the subject N2 to give rise to adjunction. In terms of (38b), the oblique N1 is adjoined to the object N2 to give rise to adjunction. Such a morphological structure is quite consistent with its original syntactic structure, as illustrated below.

(39) a. [_{VP} 教师　　　[_{v'} 教　　英语]]
　　　jiaoshi　　　jiao　Yingyu
　　　Teacher　　　teach　English
　　　(subject)　　　(object)
　　　'teacher who teaches English'

　　b. [_N 英语　　教师]]
　　　yingyu　jiaoshi
　　　English　teacher
　　　(object)　(subject)
　　　'English teacher'

(40) a. [_{VP} 教师　　　[_{v'} 教　　[_{VP} 英语　　[_{v'} 在 大学]]
　　　jiaoshi　　jiao　　yingyu　　zai Daxue
　　　teacher　　teach　　English　　at University
　　　(subject)　　　(object)　　　(oblique)
　　　'teacher teaches English at university'

　　b. [_N 大学　　　[_N 英语　　教师]]
　　　Daxue　　　yingyu　jiaoshi
　　　university　　English　teacher
　　　(oblique)　　(object)　(subject)
　　　'university English teacher'

(41) a. [_{VP} 工厂　　　[_{v'} 制造　[_{VP} 家具　　[_{v'} 用　　红木]]
　　　gongchang　　zhizao　jiaju　　yong hongmu
　　　Factory　　　make　furniture　with rosewood
　　　(subject)　　　(object)　　　(oblique)
　　　'factory makes furniture with rosewood'

　　b. [_N 红木　　　[_N 家具　　工厂]]
　　　Hongmu　　jiaju　　gongchang
　　　rosewood　　furniture factory
　　　(oblique)　　(object)　(subject)
　　　'rosewood furniture factory'

(39)-(41) show the process of derivation from the syntactic structure to the morphological structure. They quite explicitly account for why the object can be adjacent to the subject and modify it, whereas the oblique can only modify the morphological object composed of the object and the subject. It follows that the morphological structure is a mirror image of the syntactic structure. The syntactic consequences of predicate implicitness are that it triggers the occurrence of adjunction. The constituent which is at the lower level of the thematic hierarchy is adjoined to the constituent which is at the

higher level of the thematic hierarchy and functions as the modifier of the latter. Constituent adjunction must take place following the order of hierarchy and hence no cross-level adjunction is allowed. To put it differently, the constituent which is at the lower level of the thematic hierarchy cannot cross the constituent at the intermediate level to be adjoined to the constituent at the highest level of the thematic hierarchy. Accordingly, the order of adjunction of the constituents in the morphological structure can be summarized as follows.

(42) oblique < object < subject

According to (42), the object is first adjoined to the subject, for they are core constituents, whereas the oblique case is a peripheral constituent. Therefore, the adjunction of the object and the subject proceeds the adjunction of the oblique and the morphological object composed of the object and the subject, thereby giving rise to a new morphological complex, viz. [oblique[object subject]].

There is corresponding relation between the thematic hierarchy and linear order. As far as head-initial languages are concerned, the constituent at the highest level of the thematic hierarchy occurs in the left position of the syntactic structure, that is, the subject position that dominates, whereas it occurs in the right position of the morphological structure, that is, the noun head position. In contrast, the constituent that is at the lower level of the thematic hierarchy often occurs in the right position of the syntactic structure, that is, the position that is dominated, whereas it occurs in the left position of the morphological structure, that is, the modifier position. In this case, the position of a constituent in the syntactic structure and the morphological structure is determined by the position of the thematic hierarchy. The constituent which is at the higher level of the thematic hierarchy functions as the syntactic head or the morphological head, whereas the constituent which is at the lower level of the thematic hierarchy functions as the syntactic complement or the morphological modifier. It follows that the thematic hierarchy is consistent with linear order. In the same vein, the theta-roles, the syntactic constituents, and the morphological constituents are symmetrically distributed. The head constituent is an agent in the syntactic structure, while in the morphological structure, it is a noun head characterized by subjectivity. The oblique constituent functions as the adjunct of the verb in the syntactic structure, while in the morphological structure it functions as the adjunct of the noun head. The subject/agent is a head, the object/patient is an internal

modifier, and the oblique constituent is an external modifier.[14] According to property, the object/patient functions as a connotative attributive, whereas the oblique constituent functions as an extensional attributive.[15] Hence the oblique constituent, which functions as the external modifier, precedes the object/patient, which functions as the internal modifier, which precedes the subject/agent, which functions as the head.

Since the structure of Chinese nominal compounds is head-final (i.e. the noun on the right determines the basic semantics of the word), the noun which is characterized by subjectivity functions as the noun head. In this case, it corresponds to the agentive subject of the syntactic structure and the theta-role at the highest level of the thematic hierarchy. The oblique and the patient function as the modifiers of the agentive noun which functions as the head, respectively. The oblique is at the most external layer of the noun construction, whereas the patient is adjacent to the noun head. Such a linear order corresponds to the object and the adverbial of the syntactic structure as well as the lower and the lowest theta-roles in the thematic hierarchy. Based on the above argument, we came to the following conclusion.

(43) Theta-roles: agent >   patient   > instrumental/temporal/local, etc.
              ⇕   ⇕   ⇕   ⇕
     Semantic relations: subject >   object   >   oblique
              ⇕   ⇕   ⇕   ⇕
     Syntactic structure: subject >   object   >   adverbial
              ⇕   ⇕   ⇕   ⇕
     Morphological   head   > connotative >   extensional attributive
     structure: attributive

Constituents are generally arranged hierarchically and only those constituents that are adjacent in the thematic hierarchy can be merged. If the constituents that are not adjacent to the thematic hierarchy are merged, there will be ungrammatical constructions. It seems that hierarchical subjacency is a constraint on the merger of constituents. On the other hand, there is an asymmetrical c-command between constituents. A constituent cannot be merged with another constituent unless the former c-commands the latter. Rather, merged constituents must be close mates in the thematic

---

[14] When two modifiers precede the noun, the first takes scope over the second and is attached higher in terms of hierarchy.

[15] The connotative attributive is an attributive that adds lexical-semantic elements to the noun. It often occurs in the form of a substantive or open category. The extensional attributive is used to assign the property of reference or quantity to the noun. It often occurs in the form of a reference or quantifier constituent (Liu, 2008).

hierarchy. In a word, no merger can cross over more than one bounding node in a step. Based on the above discussion, we propose the constraints on the merger of constituents.

(44) The constraints on constituent merger
     a. α can be merged with β if and only if α c-commands β.
     b. Merger takes place leftward.
     c. Merger takes place between close mates.

In terms of linear order, constituents cannot be merged unless there is an asymmetrical c-command between them. Hence the order of constituents cannot be reversed. In other words, various types of constituents follow the linearity principle. If constituents are merged rightward, the combinations will be illicit, as shown in (45) and (46).

(45) a.*subject + object
     b.*subject + oblique
     c.*object + oblique

(46) a.* 工人　　石油
       gongren shiyou
       worker　petroleum

     b.* 教师　　大学
       jiaoshi　daxue
       teacher　university

     c.* 家具　　红木
       jiaju　　hongmu
       furniture rosewood

Based on (45) and (46), we argue that modification is a relation of asymmetrical c-command. In effect, the c-command which is base-generated determines the relation between the modifier and the modified. Furthermore, c-command has different representations at the syntactic level and the morphological level. In terms of the syntactic level, if X asymmetrically c-commands Y, then, X precedes Y. In terms of the morphological level, if X asymmetrically c-commands Y, then, X functions as the head while Y functions as the modifier. Furthermore, an argument of a predicate must c-command the predicate and the determiner must c-command the theta-bearer. Nominal modification is always mediated by syntactic sisterhood. The relationship between a modifier and a modified is thematic and hence it is subject to the same restrictions as theta-marking. It

follows that the two elements involved in a thematic relation are in a local relation at LF. To put it differently, nominal modification involves a kind of thematic relation (cf. Reeve & Hicks, 2017).

However, there is a crucial question we have not answered yet. We have to account for what syntactic constraints the word order of the various constituents in compounds is subject to. We argue that it is subject to the double constraints of the Locality Principle and the Prominence Principle, which coordinate and constrain the operation of the grammatical system. According to Hu (2002, 2010) and Yang (2013), the grammatical system prefers to select and process or compute the constituent which is most local (i.e. Locality Principle), whereas it prefers to select and process or compute the constituent which is most prominent (i.e. Prominence Principle). What is the most optimal is that the most local constituent corresponds to the most prominent constituent. However, the most local constituent is not necessarily the most prominent constituent. Similarly, the most prominent constituent is not necessarily the most local constituent. Nevertheless, prominence corresponds to headedness, and locality corresponds to modifiability. The more prominent a constituent is, the more likely it is to occur as a head. Accordingly, the more local a constituent is, the more likely it is to occur as a modifier. Following Bresnan (2001), we argue that prominence is determined by the following factors: 1) the linear order determined by the constituent structure (c-structure); 2) the syntactic hierarchy determined by the function structure (f-structure); 3) the thematic hierarchy; 4) the grammatical function. It is noteworthy that the thematic hierarchy is represented with the feature [±agent], whereas the grammatical function is represented with the feature [±subject]. The value of the feature [±subject] and the feature [±agent] is determined by the thematic hierarchy and the grammatical function, respectively.

(47) The thematic hierarchy
      [+agent]>[-agent]

(48) The hierarchy of the grammatical function
      [+subject]>[-subject]

The interaction between the feature [±subject] and the feature [±agent] gives rise to (49).

(49) a. [+subject, +agent]>[-subject, +agent]
      b. [+subject, -agent]>[-subject, -agent]

Locality is determined by the complexity of the structure. As far as the various constituents in compounds are concerned, we argue that the more peripheral they are, the more local they are, and the more likely they are to function as modifiers. Conversely, the less peripheral they are, the more prominent they are, and the more likely they are to function as heads. Obviously, there is an asymmetry between locality and prominence. The locality and prominence of the various constituents in compounds are determined by their positions in the syntactic hierarchy. The higher they are in the syntactic hierarchy, the more likely they are to function as heads. In contrast, the lower they are in the syntactic hierarchy, the more likely they are to function as modifiers. The positions in which the various constituents in compounds enter the syntactic structure effect the interpretation.

4   Conclusion

The argument structure, generation, and constraints of Chinese nominal compounds have been topics in the circle of Chinese linguistics, especially in the circle of generative grammar. This paper conducts research into Chinese nominal compounds with regard to the internal structure and the thematic relations between various constituents as well as the way of generation and constraints from the perspective of predicate implicitness. It is found that in the case of predicate implicitness, SV compounds can hardly be grammatical, whereas only VO compounds can be grammatical. Furthermore, agent-patient compounds cannot be grammatical, whereas only patient-agent compounds can be grammatical. If the preserved predicate is preceded by constituents, such as instrumental, temporal, and local, SV compounds cannot be grammatical, whereas only VO compounds can be grammatical. If the predicate does not occur, patient-agent compounds cannot be grammatical, whereas non-agent-patient compounds can be grammatical. The patient can function as the modifier of the agent only, but it cannot function as the modifier of the non-agentive constituent. The agent cannot function as the modifier of the patient, but the non-agentive constituent can function as the modifier of the modifier of the patient. The reason lies in the morphological structure of compounds. The function of every constituent is determined by the position of the structure. The positions in which it enters the syntactic structure have an effect on the interpretation. There is no light verb in both VO compounds and SV compounds. Therefore, the agent cannot occur. This gives a reasonable account of the grammaticality of such compounds as *shiyou gongren* 石油工人 and *zhanzheng gushi* 战争故事 and the ungrammaticality of such compounds as *gongren shiyou* 工人石油 and *gushi zhanzheng* 故事战争. As for *shiyou kaicai* 石油开采, *zhanzheng miaoxie* 战争描

写, and *jiaju zhizao* 家具制造, they are all modifier-head compounds instead of SV compounds. The morphological structure is quite consistent with its original syntactic structure. It follows that the morphological structure is a mirror image of the syntactic structure. As a consequence, predicate implicitness syntactically triggers the occurrence of an adjunction. Constituent adjunction must take place in accordance with the hierarchy and hence no cross-level adjunction is allowed. There is a corresponding relation with thematic relations, semantic relations, syntactic structure, and morphological structure. We argue that the word order of the various constituents in compounds is subject to the double constraints of the Prominence Principle and the Locality Principle. Prominence corresponds to headedness, and locality corresponds to modifiability. Obviously, there is certain asymmetry between locality and prominence. The prominence and locality of the various constituents in compounds are determined by their positions in the syntactic hierarchy.

Abbreviations

ϕ	null
A	adjective
AP	Adjective phrase
AUX	auxiliary
Cl	classifier
D	determiner
DP	determiner phrase
LF	Logical form
N	noun
N`	the intermediate projection of the noun
NP	noun
O	object
OV	object-verb
S	subject
Spec	specifier
SV	subject-verb
t	trace
V	verb
V`	the intermediate projection of the verb
VO	verb-object
VP	verb phrase

References

Baker, M. C. (1988). *Incorporation: A Theory of Grammatical Function Changing*. Chicago and London: The University of Chicago Press.

Bernstein, J. B. (2001). The DP hypothesis: Identifying clausal properties in the nominal domain. In M. Baltin and C. Collins (Eds.), *The Handbook of Contemporary Syntactic Theory*. Oxford: Blackwell Publishers Ltd., (pp. 536-561).

Bresnan, J. (2001). *Lexical-Functional Syntax*. Oxford: Blackwell.

Chao, Y. (1968). *A Grammar of Spoken Chinese*. Berkeley and Los Angeles: University of California Press.

Cheng, G. (1999). *Linguistic Universalism*. Shanghai: Shanghai Foreign Language Education Press.

Cheng, G. (2005). "*Zhe*" synthetic compounds in Chinese and their implications for UG. *Modern Foreign Languages* 3, 232-238.

Chomsky, N. (1995). The Minimalist Program. Cambridge, MA: The MIT Press.

Dai, M. (2003). *Generalized Leftward Merger Theory in the Minimalist Framework*. Beijing: Foreign Language Teaching and Research Press.

Fu, A. (2004). A quantitative survey of noun phrase constructions in Chinese. *Studies of the Chinese Language* 6, 508-520.

Grimshaw, J. (1990). *Argument Structure*. Cambridge, MA.: The MIT Press.

Gruber, J. S. (2001). Thematic relations in syntax. In M. Baltin & C. Collins (Eds.), *The Handbook of Contemporary Syntactic Theory*. Oxford: Blackwell Publishers Ltd., (pp. 257-298).

Gu, Y., & Shen, Y. (2001). The derivation of synthetic compounds in Chinese. *Studies of the Chinese Language* 2, 122-133.

He, Y., & Wang, L. (2005). Real and pseudo synthetic compounds in Chinese. *Language Teaching and Linguistic Studies* 5, 11-21.

He, Y. (2004). The loop theory in Chinese morphology. *Contemporary Linguistics* 3, 223-235.

Hu, J. (2002). *Prominence and Locality in Grammar: The Syntax and Semantics of Wh-questions and Reflexives*. PhD. dissertation, the City University of Hong Kong.

Hu, J. (2010). The distribution and selection of arguments: Prominence and Locality in grammar. *Studies in the Chinese Language* 1, 3-20.

Huang, B., & Liao, X. (2007). *Modern Chinese*. Beijing: Higher Education Press.

Huang, C.-T. J. (2005). Syntactic analyticity and the other end of the parameter. Lecture notes, LSA 2005 Summer Institute, MIT and Harvard University.

Jackendoff, R. S. (1972). *Semantic Interpretation in Generative Grammar*. Cambridge, MA: The MIT Press.

Jackendoff, R. S. (1990). *Semantic Structures*. Cambridge, MA: The MIT Press.

Kayne, R. (1994). *The Antisymmetry of Syntax*. Cambridge, MA: The MIT Press.

Kuroda, S. Y. (1988). Whether we agree or not: A comparative syntax of English and Japanese. *Lingvisticae Investigatione, 12*(1) , 1-47.

Larson, R. K. (1988). On the double object construction. *Linguistic Inquiry* 19, 335-391.

Larson, R. K. (1990). Double object revisited: A reply to Jackendoff. *Linguistic Inquiry* 21, 589-632.

Lin, T. (2001). *Light verb syntax and the theory of phrase structure.* Ph.D. diss., University of California, Irvine.

Liu, D. (2008). Typological features of nominal phrases in Chinese. *Studies of the Chinese Language* 1, 3-20.

Lu, Z. (1964). *The Morphology of Chinese*. Beijing: Science Press.

Németh T., E., & Bibok, K. (2010). Interaction between grammar and pragmatics: The case of implicit arguments, implicit predicates and co-composition in Hungarian. *Journal of Pragmatics* 42, 501-524.

Ren, X. (1981). *The Morphology of Chinese*. Beijing: China Social Sciences Press.

Reeve, M. & Hicks, G. (2017). Adjunct extraposition: Base generation or movement? *Syntax* 20(3) , 215-248.

Selkirk, E. (1984). *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, MA: The MIT Press.

Shi, D. (2003). Chinese attributive V-N compounds. *Studies of the Chinese Language* 6, 483-495.

Sportiche, D. (1988). A theory of floating quantifiers and its corollaries for constituent structure. *Linguistic Inquiry* 19, 425-449.

Tai, J. H. (2002). Conceptual structures and non-autonomous syntax: Some conceptualization principles in Chinese grammar. *Contemporary Linguistics* 1, 1-12.

Tang, S. (2008). On the argument structure of Chinese compounds. *Language Teaching and Linguistic Studies* 4, 10-17.

Tang, S. (2014). Asymmetry in Chinese compound formation. *Chinese Linguistics* 1, 69-77.

Yang, Y. (2006). Generation of OV inversion. *Linguistic Sciences* 3, 39-48.

Yang, Y. (2007a). The syntactic status of NP in Vi+NP. *Studies in Language and Linguistics* 2, 59-64.

Yang, Y. (2007b). The pragmatic motivation of Vi+NP as anomalous syntax. *Chinese Linguistics* 1, 58-63.

Yang, Y. (2009). A parametric approach to the typology of non-patient objects. *Modern Foreign Languages* 1, 33-41.

Yang, Y. (2010). A reanalysis of *de* and *de* construction. *Journal of Foreign Languages* 5, 34-44.

Yang, Y. (2011). Non-prototypical patient object sentences in Chinese. *Philippine Journal of Linguistics* 42, 105-127.

Yang, Y. (2012). Derivation of serial verb constructions. *Language Research* 3, 665-690.

Yang, Y. (2013). Modal particles and their projections: A minimalist approach. *Language Research* 1, 95-136.

Yang, Y. (2014a). The syntactic status of *de* and the internal structure of *de* constructions in Chinese. *Language Research* 1, 121-157.

Yang, Y. (2014b). A parametric approach to the typology of serial verb constructions. *Contemporary Foreign Languages Studies* 3, 1-8.

Yang, Y. (2015). VO inversion and the direction of transference of object in compounds. *Journal of Zhejiang International Studies University* 5, 54-60.

Yang, Y. (2016a). The argument structure of the serial verb construction and its derivation. *Journal of Zhejiang International Studies University* 3, 1-8.

Yang, Y. (2016b). Derivation of sentences with object-oriented adverbials. *Journal of Guangdong Foreign Studies University* 5, 45-53.

Yuan, Y. (1995). Predicate implicitness and its syntactic consequences. *Studies of the Chinese Language* 4, 241-255.

Zhou, R. (2006). A study on Chinese /OVN compounds under typological evidence: University vs. individuality. *Studies of the Chinese Language* 4, 301-312.

Zhu, D. (1982). *Lectures of Grammar*. Beijing: The Commercial Press.

Zhuang, H., & Liu, Z. (2011). Morphology of Chinese synthetic compounds and their prosodic constraint. *Chinese Teaching in the World* 4, 497-506.

Wh-questions in Japanese: Challenges for Vietnamese L2 learners

Trang PHAN
Ca' Foscari University of Venice, Italy
thihuyentrang.phan@unive.it

Giang VU
Vietnam National University Hanoi, Vietnam
giangvt@ulis.vnu.edu.vn

Abstract

The study explored how Vietnamese learners of Japanese handle structural differences in *wh*-questions. Despite both being *wh-in-situ* languages, the divergent word order poses challenges for Vietnamese learners. The research involved 65 high school students with 2.5 to 13 years of Japanese learning. Survey results showed a positive correlation between Japanese proficiency and performance. Students were more accurate when Japanese *wh*-questions mirrored Vietnamese word order, especially in 'why' and 'when' questions. Difficulties arose when structures differed. Advanced learners could acquire L2 *wh*-questions with parametric values distinct from L1, but native-like interpretations in Japanese remained challenging. The study advocates for explicit classroom instruction on Japanese *wh*-question word order, particularly with scrambled sentences, to enhance accuracy and improve overall teaching effectiveness.

Keywords: *wh*-movement, *wh-in-situ*, scrambling, Vietnamese, Japanese

Povzetek

Raziskava obravnava razumevanje in uporabo *wh*-vprašanj med vietnamskimi učenci japonščine. Kljub temu, da sta oba jezika jezika *wh-in situ*, se različen besedni red izkaže kot izziv za vietnamske učence. Raziskava vključuje 65 učencev, ki so se japonščino učili v razponu med dvema letoma in pol ter trinajstimi leti. Rezultati ankete kažejo pozitivno povezavo med znanjem japonščine in uspešnostjo. Učenci so bolj natančni, ko so japonska *wh*-vprašanja podobna vietnamskemu vrstnemu redu, še posebej pri vprašanjih 'zakaj' in 'kdaj'. Težave nastanejo pri različnih strukturah med jezikoma. Nekateri napredni učenci lahko obvladajo japonska *wh*-vprašanja, vendar doseči naravno tolmačenje ostaja izziv. Študija zagovarja učenje vrstnega reda japonskih *wh*-vprašanj, še posebej pri premešanih stavkih, za povečanje natančnosti in učinkovitosti poučevanja japonskega jezika.

Ključne besede: premik *wh*-vprašalnice, *wh*-vprašalnice *in situ*, preurejanje, vietnamščina, japonščina

1   Objectives

In Vietnam, the burgeoning diplomatic, economic, and cultural ties with Japan have sparked a growing interest in Japanese language acquisition. Nevertheless, language acquisition remains a challenging task. In light of this, the present study seeks to investigate the extent to which the syntactic typological differences between Vietnamese and Japanese impact Vietnamese students' acquisition of Japanese *wh*-question word order. Ultimately, the findings of this study may inform pedagogical strategies to improve the efficacy of teaching and learning Japanese for Vietnamese language learners.

2   Why Vietnamese learners of Japanese?

The Japanese and Vietnamese languages were chosen for this study due to their interesting typological similarities and differences concerning question formation and word order.

The first typological distinction which highlights the similarities between Japanese and Vietnamese is the one between *wh*-movement and *wh-in-situ*, which refers to whether a language employs the movement of interrogative words (such as 'who', 'what', 'where', 'how', 'why', 'when', etc.) to form a question. In *wh*-movement languages, they utilize a syntactic operation where the interrogative word moves to the beginning of the sentence, leaving behind a trace in its original position. In contrast, *wh-in-situ* languages keep the interrogative word in its original position and form the question without movement.

In many languages such as English and German, *wh*-movement is employed to transform declarative sentences into *wh*-questions. Consider the example in English:

(1) a. He went to the store.

 b. Where did he go?

In English, crafting a *wh*-interrogative sentence requires both *wh*-movement and *do*-insertion. The *wh*-word, such as 'where' in (1b), is shifted to the beginning of the sentence, and a tense auxiliary verb is inserted before the subject, leading to a change in word order between the *wh*-interrogative and its declarative counterpart.

In contrast, *wh-in-situ* languages include Chinese and Korean (Huang, 1982; Cheng & Rooryck, 2000; Beck & Kim, 1997). Let's take a look at an example in Mandarin Chinese:

(2) a. 他 去 了   商店。   Mandarin Chinese

     tā qù le   shāngdiàn

     he go PERF store

     'He went to the store.'

   b. 他 去 了   哪里? Mandarin Chinese

     tā qù le   nǎlǐ?

     he go PERF where

     'Where did he go to?'

In Mandarin Chinese, as demonstrated in example (2b), a notable distinction from English interrogative sentences emerges. Unlike English, where the interrogative word is relocated to the beginning of a sentence, the Mandarin Chinese interrogative word, *nǎlǐ* 'where' remains fixed within the sentence's original position. Remarkably, this characteristic feature is accompanied by the absence of verb movement or inversion in Mandarin Chinese questions. The sentence's word order and structure remain unchanged when compared to the corresponding declarative sentence (2a).

Mandarin Chinese exemplifies a *wh-in-situ* language by retaining the *wh*-word in its original position, showcasing its distinct approach to forming interrogative sentences compared to those languages that employ *wh*-movement and other syntactic transformations (Pesetsky, 1987).

The distinction between *wh*-movement and *wh-in-situ* has been a topic of interest in typological and generative grammar studies. According to Dryer (1991), *wh-in-situ* is more common among the world's languages. The reasons for this typological variation are still a subject of debate among linguists on whether these parameters are innately determined by the human mind or whether they are the result of functional pressures on the language system (Hawkins, 2004).

As it will become clear in the following sections, both Japanese and Vietnamese are languages that have *wh-in-situ* (Saito, 1992; Tomioka, 2007; Tran, 2009; Tsai, 2009), meaning question words can appear in their base positions in a sentence. Studies in foreign language acquisition have consistently shown that learning is more accessible when the target language and the native language share typological similarities (Krashen, 1982; Ringbom, 2006). This is because the use of cross-linguistic similarities, known as transfer,

plays a crucial role in language learning. Therefore, it is expected that the shared typological features between Vietnamese and Japanese could greatly improve Vietnamese learners' understanding of Japanese.

However, despite this apparent similarity, a pivotal typological contrast between Japanese and Vietnamese complicates the language learning process. This distinction lies in the contrast between head-final and head-initial languages. Japanese adopts a head-final structure (Subject-Object-Verb), where the usual sentence order places the subject, followed by the object, and concludes with the verb. Conversely, Vietnamese follows a head-initial pattern (Subject-Verb-Object), positioning the verb before the object. This fundamental disparity in word order poses a challenge for Vietnamese speakers attempting to construct *wh*-questions in Japanese accurately.

To sum up, while both languages employ *wh-in-situ* structures, the difference in word order between Japanese and Vietnamese presents a significant obstacle for Vietnamese learners. This contrast is expected to create difficulties in mastering the formation of *wh*-questions in Japanese. Therefore, our study specifically focuses on the challenges of Vietnamese learners acquiring Japanese *wh*-interrogatives.

While the past decade has seen a surge in research on Japanese as a Second/Foreign Language acquisition, there remains a significant gap in understanding how Vietnamese learners specifically grapple with the complexities of Japanese. Existing studies predominantly feature English and Chinese L1 learners of Japanese, leaving the experiences of Vietnamese learners in Japanese language acquisition largely unexplored, especially concerning word order acquisition (Lieberman et al., 2006; Umeda, 2008). In this context, the term 'transfer' (Ringbom, 2006) is particularly relevant, encompassing the impact of the learners' L1 (Vietnamese) on their Japanese learning process. Positive transfer occurs when the influence of Vietnamese leads to accurate acquisition, whereas negative transfer results in errors or difficulties.

To comprehensively understand these transfer patterns and their implications, our study emphasizes a thorough contrastive analysis, focusing specifically on the distinctions between Vietnamese (learners' L1) and Japanese (the target language). By unraveling the complexities of linguistic transfer within this specific learner group, we aim to provide valuable insights into the challenges Vietnamese learners had to face when mastering Japanese *wh*-interrogatives, offering nuanced pedagogical suggestions for more effective language teaching strategies.

3   *Wh*-questions in Japanese and Vietnamese

3.1   The list and positions of *wh*-words

This research primarily focuses on the position of question words in *wh*-questions and compares the two languages, Japanese and Vietnamese. In particular, the basic *wh*-words examined in the two languages are listed below.

Table 1: List of *wh*-words investigated in Japanese and Vietnamese

Language						
	だれ dare	なに/何 nani/nan	どこ doko	どのように donoyouni	なぜ naze	いつ itsu
Japanese						
Vietnamese	ai	cái gì	ở đâu	như thế nào	vì sao	khi nào
English	who	what	where	how	why	when

Noam Chomsky emphasized the creative nature of language, stating that language is a process of free creation where the principles of generation are fixed, yet the way they are employed is diverse and unrestricted (Chomsky, 1970). This insight underscores the significance of comprehending the distinct principles and parameters inherent in various languages, including the formation of *wh*-questions in Japanese and Vietnamese. Understanding these linguistic nuances is essential for effective language learning and usage. Our study is based on the assumption that all languages in the world follow certain rules when constructing sentences. Although languages around the world may have diverse external manifestations, all sentences can be divided into three different domains when it comes to structure: CP (complementizer phrase), the domain of the pragmatics (usually located before the subject); IP (inflection phrase), the domain of functional elements (that are normally located between the subject and the predicate); and VP (verb phrase), the domain of lexical elements (located inside the predicate).

Applying this assumption to our discussion, we take into consideration three positions of *wh*-words in a sentence: a) before the subject, b) between the subject and the verb, and c) after the verb, as illustrated in Figure 1 below.
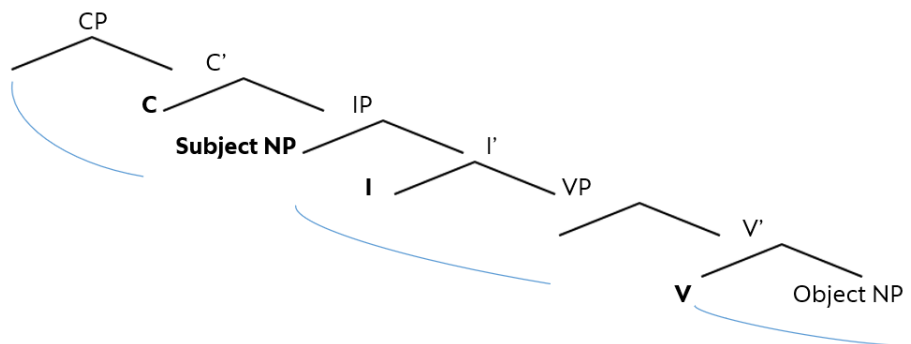
Figure 1: Three positions of *wh*-words in a sentence

We will examine each of the *wh*-words from Table 1 in all three positions. In the following section, we provide a brief overview of the different positions of *wh*-words in Japanese and Vietnamese, which sets out the theoretical foundation for our survey.

3.2   The position of *wh*-words in Japanese

The principal reservoir of Japanese language data utilized in this study comprises the works of Zhang and Hayashi (2000), Gunji (2011), and Minna no Nihongo Shokyuu 1 Honsatsu [Everyone's Japanese Elementary 1 Textbook] (1998).

Let's first have a look at the different positions of *dare* だれ ('who') in a sentence.

(3) *dare* だれ ('who')

    a. だれ に　あなた は　会う　の　ですか。 Japanese
       dare ni　anata　wa　au　　no　desu ka
       who DAT you　　TOP meet ACC POL Q
       'Who are you meeting with?'

    b. あなた は　だれ に　会う　の　ですか。 Japanese
       anata　wa　**dare** ni　au　　no　desu ka
       you　　TOP **who** DAT meet ACC POL Q
       'Who are you meeting with?'

c. あなた が　　会う　の　　は　　だれ ですか。　Japanese

anata　ga　　au　　no　wa　**dare** desu ka

you　　NOM meet ACC TOP **who** POL Q

'Who are you meeting with?'

In example (3a), *dare* appears at the beginning of the sentence, preceding the subject *anata*. In example (3b), it appears in the middle of the sentence after the subject *anata* and before the verb *au*. In example (3c), *dare* appears at the end of the sentence, after the object *kare* and the verb *au*. The examples emphasize that the use of *dare* ('who') as a question word can be used in all three different positions of a *wh*-word in a sentence.[1] A similar pattern is observed for other *wh*-words in Japanese.

(4) *nani/nan* なに ('what')

a. 何　　を　　あなた は　　食べたい　　ですか。　Japanese

nani　wo　anata　wa　tabetai　　desu ka

what DAT you　　TOP want to eat POL Q

'What do you want to eat?'

b. あなた は　　何　　を　　食べたい　　ですか。　Japanese

anata　wa　**nani** wo　tabetai　　desu ka

you　　TOP **what** DAT want to eat POL Q

'What do you want to eat?'

c. あなた が　　食べたい　の　　は　　何　　ですか。　Japanese

anata　ga　　tabetai　　no　wa　**nani** desu ka

you　　NOM want to eat ACC TOP **what** POL Q

'What do you want to eat?'

---

[1] Several factors, including word order syntax, pragmatics, and prosody influence the formation of *wh*-questions (Kitagawa, 2005; Tomioka, 2007). However, this study focuses solely on word order syntax and prosody, excluding discourse level and pragmatics. The exploration of discourse-level aspects, particularly topicalization, is deferred to future research endeavors.

(5) *doko* どこ('where')

  a. どこ　で　私　　は　両替する　　　の　ですか。 Japanese

    doko　de　watashi wa　ryougae suru　　no　desu ka

    where DAT I　　　TOP　money exchange ACC POL Q

    'Where do I exchange money?'

  b. 私　　は　どこ　で　両替する　　　の　ですか。 Japanese

    watashi wa　**doko**　de　ryougae suru　　no　desu ka

    I　　　TOP **where** DAT money exchange ACC POL Q

    'Where do I exchange money?'

  c. 私　　が　両替する　　　の　は　どこ　ですか。 Japanese

    watashi ga　ryougae suru　　no　wa　**doko** desu ka

    I　　　NOM money exchange ACC TOP **where** POL Q

    'Where do I exchange money?'

(6) *donoyouni* どのように ('how')

  a. どのように あなた は　休み　を　過ごします　か。 Japanese

    donoyouni anata　wa　yasumi wo　sugoshimasu ka

    how　　you　TOP holiday DAT spend　　Q

    'How do you spend your holiday?'

  b. あなた は　どのように 休み　を　過ごします　か。 Japanese

    anata　wa　**donoyouni** yasumi wo　sugoshimasu ka

    you　TOP **how**　　holiday DAT spend　　Q

    'How do you spend your holiday?'

  c. あなた は　休み　を　どのように 過ごします　か。 Japanese

    anata　wa　yasumi wo　**donoyouni** sugoshimasu ka

    you　TOP holiday DAT **how**　　spend　　Q

    'How do you spend your holiday?'

(7) *naze* なぜ ('why')

  a. なぜ あなた は　彼　に　会う　の　ですか。 Japanese

    naze anata　wa　kare ni　au　no　desu ka

    why you　TOP him　DAT meet ACC POL Q

    'Why are you meeting him?'

b. あなた は　なぜ 彼　に　会う の　ですか。 Japanese

anata　wa　**naze** kare ni　au　no　desu ka

you　　TOP **why** him DAT meet ACC POL Q

'Why are you meeting him?'

c. あなた が　　彼　に　会う の　は　なぜ ですか。 Japanese

anata　ga　　kare ni　au　no　wa　**naze** desu ka

you　　NOM him DAT meet ACC TOP **why** POL Q

'Why are you meeting him?'

(8) *itsu* いつ ('when')

a. いつ あなた は　北京　に　着く の　ですか。 Japanese

itsu　anata　wa　Pekin　ni　tsuku no　desu ka

when you　　TOP Beijing DAT arrive ACC POL Q

'When will you arrive in Beijing?'

b. あなた は　いつ　北京　に　着く　の　ですか。 Japanese

anata　wa　**itsu**　Pekin　ni　tsuku no　desu ka

you　　TOP **when** Beijing DAT arrive ACC POL Q

'When will you arrive in Beijing?'

c. あなた が　　北京　に　着く の　は　いつ ですか。 Japanese

anata　ga　　Pekin　ni　tsuku no　wa　**itsu**　desu ka

you　　NOM Beijing DAT arrive ACC TOP **when** POL Q

'When will you arrive in Beijing?'

To summarize, there are two types of *wh*-words in a sentence, the former includes *nani/nan* なに ('what') or *dare* だれ ('who'), which function as arguments, whereas the latter type includes *doko* どこ('where'), *donoyouni* どのように ('how'), *naze* なぜ('why'), or *itsu* いつ ('when'), which function as sentence adjuncts. It can be observed that in the above examples, regardless of their function within the sentence, all of these *wh*-words can appear in all three positions: before the subject, between the subject and the predicate, and after the verb. These examples illustrate that, unlike English where the word order of questions is typically fixed, Japanese allows for greater flexibility in the placement of *wh*-words, which in turn enables greater expressivity in sentence formation and helps convey nuances in meaning. As such, understanding the various positions in which *wh*-words can appear is an important aspect of learning and using the Japanese language effectively.

3.3   The position of *wh*-words in Vietnamese

Initially, we shall examine the various syntactical positions of *ai* ('who') and *cái gì* ('what') in the Vietnamese language within a sentence.

(9) *ai* ('who')

  a. ? **Ai**   bạn sẽ   gặp?   Vietnamese
     what you will meet
     Intended: 'Who will you meet with?'

  b. *Bạn **ai**   sẽ   gặp?   Vietnamese
      you   **who** will meet
      Intended: 'Who will you meet with?'

  c. Bạn sẽ   gặp   **ai**?   Vietnamese
     you will meet **who**
     'Who will you meet with?'

In *wh*-questions that function as objects, *ai* ('who') can naturally appear in the base position, following the verb. Placing it before the verb results in an ungrammatical sentence, and positioning it at the beginning of the sentence diminishes the sentence's grammaticality. In the latter context, 'who' can only be interpreted contrastively, such as in the sentence, 'Who is it that you will meet with?' Consequently, we labeled the sentence with a question mark.

(10) *cái gì* ('what')

  a. ? **Cái gì** bạn muốn   ăn? Vietnamese
     what   you want   eat
     Intended: 'What do you want to eat?'

  b. *Bạn **cái gì** muốn   ăn? Vietnamese
      you   **what** want   eat
      Intended: 'What do you want to eat?'

  c. Bạn muốn ăn   **cái gì**? Vietnamese
     you want eat **what**
     'What do you want to eat?'

Similar to *ai* ('who'), *cái gì* ('what') can be used in two positions of the sentence; and when it is fronted to the beginning of the sentence, it is interpreted contrastively, cf. 'What is this that you want to eat?'.

Hereinbelow is presented a comprehensive listing of the feasible placements of other *wh*-words in Vietnamese.

(11) *ở đâu* ('where')

    a. ? **Ở đâu** tôi sẽ đổi   tiền?   Vietnamese

       where   I   will exchange money

       Intended: 'Where will I exchange money?'

    b. *Tôi sẽ **ở đâu** đổi   tiền?   Vietnamese

       I   will **where** exchange money

       Intended: 'Where will I exchange money?'

    c. Tôi sẽ đổi   tiền   **ở đâu**? Vietnamese

       I   will arrive Beijing **where**

       'Where will I exchange money?'

In the examples given, the interrogative word *ở đâu* ('where') appears most naturally in the sentence-final position only. Placing it in the sentence-initial position results in a decrease in grammaticality, and the mid-sentence position results in ungrammaticality.

(12) *như thế nào* ('how')

    a. ***Như thế nào** bạn sẽ trải   qua   kì   nghỉ?   Vietnamese

       how   you will experience through CLF vacation

       Intended: 'How will you spend your vacation?'

    b. *Bạn sẽ **như thế nào** trải   qua   kì   nghỉ?   Vietnamese

       you   will **how**   experience through CLF   vacation

       Intended: 'How will you spend your vacation?'

    c. Bạn sẽ trải   qua   kì   nghỉ   **như thế nào**? Vietnamese

       you will experience through CLF   vacation **how**

       Intended: 'How will you spend your vacation?'

The *wh*-word *như thế nào* ('how') in the above examples only appears in one position, after the verb.

(13) *vì sao* ('why')

    a. **Vì sao** bạn gặp   anh ta? Vietnamese

      why   you meet him

      'Why are you meeting with him?'

    b. Bạn **vì sao** gặp   anh ta? Vietnamese

      you **why**   meet him

      'Why are you meeting with him?'

    c. Bạn gặp   anh ta **vì sao**? Vietnamese

      you meet him   **why**

      'Why are you meeting with him?'

In the given examples, *vì sao* ('why') in principle can appear in all the three positions.

(14) *khi nào* ('when')

    a. **Khi nào** bạn đến   Bắc Kinh? Vietnamese

      when   you arrive Beijing

      'When will you arrive in Beijing?'

    b. Bạn **khi nào** đến   Bắc Kinh? Vietnamese

      you **when**   arrive Beijing

      'When will you arrive in Beijing?'

    c. Bạn đến   Bắc Kinh **khi nào** Vietnamese

      you arrive Beijing   **when**

      'When did you arrive in Beijing?'

In these examples, the question word *khi nào* ('when') appears in all three positions. However, there is a difference in terms of the tense used in the sentences. Specifically, *khi nào* ('when') is interpreted in the future tense when it precedes the subject and the verb whereas it is interpreted in the past tense when it follows the verb. This has been noted by researchers as a feature of the Vietnamese language (Cao, 2004; Duffield, 2007; Tsai, 2009; Tran, 2009, Tran, 2021). That is, the placement of *khi nào* ('when') in Vietnamese appears to be exceptionally versatile; nevertheless, its positioning is, in fact, contingent upon the tense utilized within the given sentence.

That is, the inflexibility of the position of *wh*-words in Vietnamese is due to at least two factors, namely, the tense of the verb and the contrastiveness of information.

Based on the data presented above, we can summarize the positions of question words in Japanese and Vietnamese in Table 2.

Table 2: Position of *wh*-words in Japanese and Vietnamese[2]

		why	when	where	how	what	who
Japanese	before the subject	v	v	v	v	v	v
	between the subject and the verb	v	v	v	v	v	v
	after the verb	v	v	v	v	v	v
Vietnamese	before the subject	v	v	?	*	?	?
	between the subject and the verb	v	v	*	*	*	*
	after the verb	v	v	v	v	v	v

The empirical findings discussed in Section 3 imply that despite both Japanese and Vietnamese being characterized by a *wh-in-situ* structure, whereby question words are positioned in their base positions within a sentence, Japanese exhibits greater flexibility in the placement of wh-words than Vietnamese. Among the six investigated *wh*-words, 'why' and 'when' demonstrate the highest degree of commonality between Vietnamese and Japanese, while 'where', 'how', 'what', and 'who' display the most notable differences. The present study, therefore, aims to examine the level of proficiency among Vietnamese students in comprehending and utilizing the word order about Japanese interrogative sentences containing *wh*-words.

4   The survey

4.1   Participants

The target population for the survey comprises 65 individuals learning Japanese at the Foreign Language Specialized School, an affiliate of the VNU University of Languages & International Studies, Vietnam National University

---

[2] This study presents an initial exploration of *wh*-words in relation to the fundamental sentence components in Vietnamese and Japanese, focusing on subject arguments and verbs. Moving forward, it is essential to expand our analysis to include topics as well. We express gratitude to an anonymous reviewer for this valuable suggestion, which will be a significant aspect of our future investigations.

Hanoi. The participants are distributed across different grades, with 19 students from grade 10, 11 students from grade 11, and 35 students from grade 12. The duration of Japanese language instruction for the surveyed students ranges from 2.5 to 13 years, with 11 individuals among the cohort having lived in Japan. The Japanese language curriculum employed in the school features the following textbooks: Japanese for grade 10, Japanese for grade 11, and Japanese for grade 12, all of which are published by the Vietnam National Education Publishing House.

Regarding Japanese language competence, the students have been categorized according to their proficiency level as per the Japanese Language Proficiency Test (JLPT), with three students classified as N5, twenty-five as N4, twenty-two as N3, thirteen as N2, and two as N1, as illustrated in Figure 2.
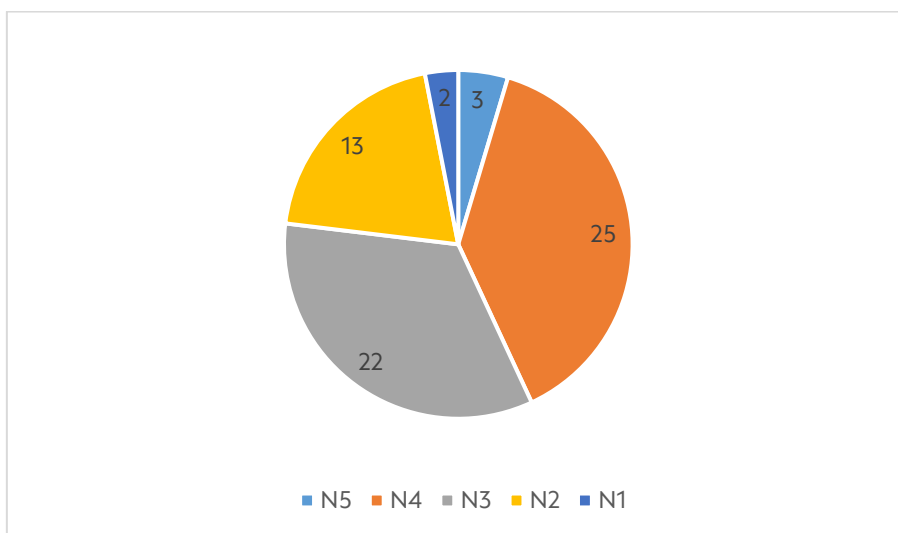
Figure 2: Participants by Japanese language proficiency

In our study, N4 and N5 represent advanced proficiency levels, N3 indicates intermediate proficiency, and N2 and N1 are considered beginner levels.

4.2 The stimuli

The survey methodology entails an online test with a duration of 20-30 minutes comprising 18 stimuli questions (6 *wh*-words, each in 3 positions) in the multiple-choice format, requiring responses indicating correctness or incorrectness by selecting either the affirmative or negative symbol (i.e., ⭕ or ✕). It should be emphasized that communication or collaboration with other participants during the test is strictly prohibited.

4.3   Predictions

The primary objective is to investigate potential correlations between differences in the student's native language (L1) and the target language (L2), their proficiency levels, and the percentage of the related word order errors in L2 comprehension attributed to L1 transfer.

The analysis of the collected data is expected to reveal the following outcomes:

(i) Word Order Similarity: Students are likely to provide accurate responses when the word order of Japanese *wh*-questions aligns with that of Vietnamese. In other words, when the word order structures in both languages are similar, students are anticipated to demonstrate higher accuracy in their responses.

(ii) Proficiency Levels: We anticipate that students with higher levels of proficiency in Japanese, especially those classified as intermediate or advanced, will exhibit greater accuracy when answering survey questions compared to their counterparts with lower proficiency levels. This expectation aligns with the notion that proficiency levels often impact the ability to grasp complex grammatical structures.

By investigating these aspects, this research aims to provide insights into the relationship between L1-L2 differences, proficiency levels, and the acquisition of specific grammatical structures, particularly in the context of Japanese *wh*-questions.

4.4   Results

The findings of the survey on Japanese *wh*-words proficiency among Vietnamese students are as follows. The column chart in Figure 3 below provides a visual representation of the accuracy rate of Vietnamese students' responses with respect to their Japanese language proficiency.
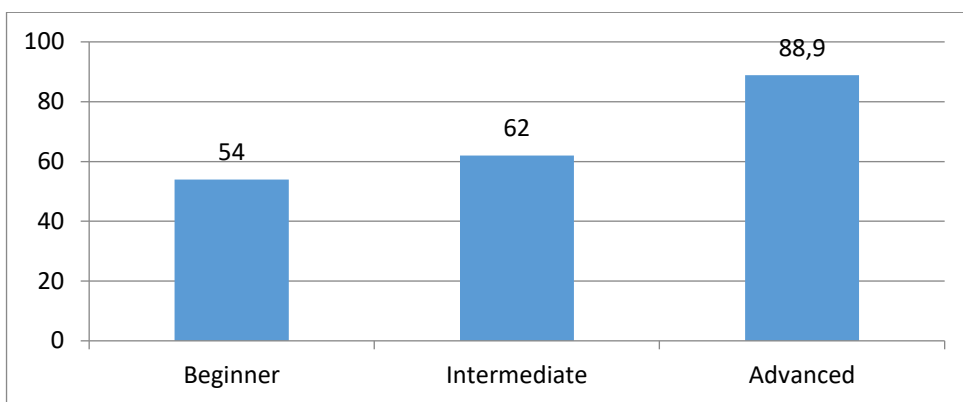
Figure 3: Accuracy rate of students by their Japanese language proficiency

As depicted in Figure 3, students exhibiting advanced Japanese language proficiency demonstrate the highest level of accuracy, followed by individuals at intermediate proficiency levels and, lastly, those at lower levels of proficiency. Students who display better command of the Japanese language are more likely to provide correct responses, thereby aligning with the earlier predictions formulated based on the survey outcomes.

The evident pattern displayed in Figure 3 implies a direct correlation between proficiency and accuracy. These results corroborate our initial hypotheses, highlighting that students possessing a higher level of Japanese language proficiency tend to offer more precise responses. While these conclusions offer a fundamental understanding of proficiency and accuracy, further research is essential to unveil the intricate nuances and potential non-linear patterns in language acquisition processes.

Furthermore, the study showed that certain advanced learners could grasp L2 *wh*-constructions with different parametric values from their native language. Nevertheless, attaining native-like interpretations of *wh*-phrases in Japanese posed ongoing challenges, underscoring the complexity of language acquisition in this context.

Now let's see how Vietnamese students respond to different types of *wh*-words. To this end, the subsequent chart provides a detailed account of the accuracy rate of their responses about different categories of *wh*-words in Japanese.
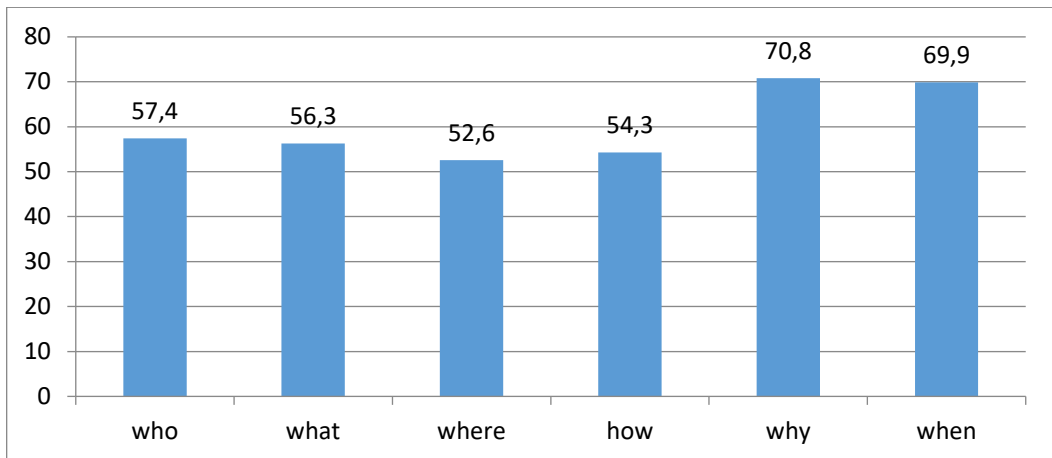
Figure 4: Accuracy rate of students with respect to different *wh*-words in Japanese

Figure 4 presents a compelling demonstration that the accuracy of responses to the Japanese *wh*-words 'why' and 'when' is notably high, amounting to approximately 70% of the surveyed population. This high accuracy can be attributed to the substantial shared characteristics between Vietnamese and Japanese in the context of these specific *wh*-words. Among the six *wh*-words investigated, 'why' and 'when' exhibit the most pronounced similarities between the two languages, as shown in the previous section.

Notably, this observation raises intriguing questions for future research. It prompts an exploration into whether there exists a deeper theoretical basis explaining the exceptional accuracy of *why*-questions. Previous literature highlights 'why' as a distinctive element (Lin, 1992; Tsai, 1999; Ko, 2005), underscoring the need for further investigation into the theoretical underpinnings contributing to the heightened accuracy in this particular category.

In the next step, we consider how Vietnamese students respond to different positions of *wh*-words.
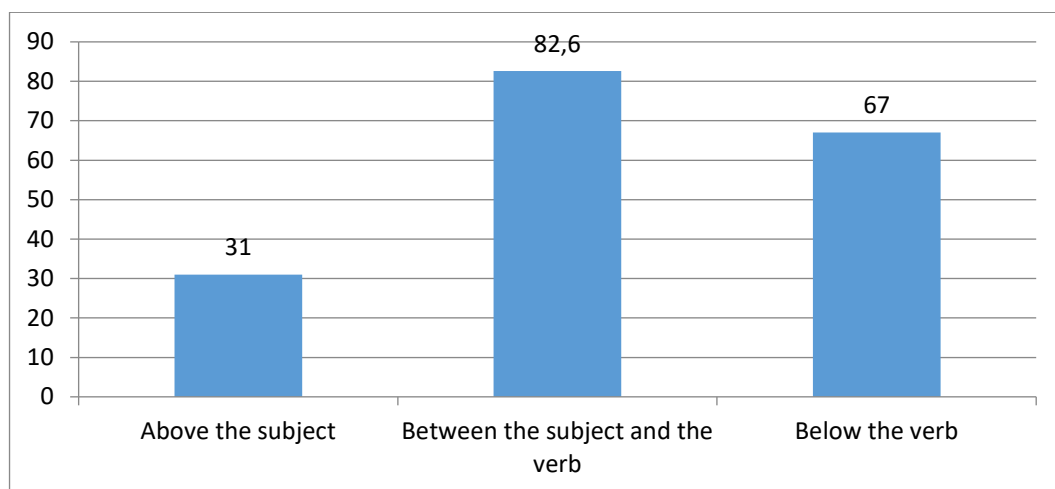
Figure 5: Accuracy rates of students according to different positions of Japanese *wh*-words

Based on the data presented in Figure 5, it is evident that the most accurate responses to questions containing *wh*-words are found when these words are positioned between the subject and the verb (or mid-sentence position) with an accuracy rate of 82.6%. The accuracy rate drops to 31% when the *wh*-word is placed before the subject. Conversely, the accuracy rate increases when the *wh*-word is positioned after the verb. The rationale behind the superior accuracy rate of *wh*-words in the mid-sentence position can be attributed to the subject-object-verb (SOV) word order, which is typical in the Japanese language. Hence, the question word is most likely to appear in the middle of the sentence. This knowledge is commonly documented in Japanese language textbooks that Vietnamese students currently use in the curriculum. For instance, the beginner's level textbook Mina no Nihongo 1 includes questions that feature question words in the mid-sentence position.

(15) a. それ は　　何　ですか。 (Unit 2) Japanese
　　　Sore wa　 nan　desu ka
　　　DEM TOP what POL Q
　　　'What is that?'

　　……名刺　　　　です。 (Unit 2) Japanese
　　　Meishi　　　 desu
　　　business card POL
　　　'It's a business card.'

b. トイレ は　どこ　ですか。　(Unit 3) Japanese

 Toire wa **doko** desu ka

 Toilet TOP **where** POL Q

 'Where is the toilet?'

…… あそこ　です。　(Unit 2) Japanese

  asoko desu

  there POL

  'It's over there.'

The accuracy rate of questions placed at the beginning of a sentence is notably lower, primarily due to the subject-object-verb (SOV) basic sentence order that is prevalent in the Japanese language. The structure with questions at the beginning (also known as 'scrambling') using the OSV formula is not extensively addressed in Japanese language textbooks for foreign learners, making it challenging for students to comprehend the intricacies of scrambled sentences and their function within the language. Jorden (1962) emphasizes the importance of studying and understanding scrambled sentences in her book *Japanese: The Spoken Language*, highlighting that it is a crucial aspect of the contemporary Japanese language, particularly in spoken language where it is commonly used for emphasis. See also Saito (1985, 1992), Miyagawa (2003). However, this crucial aspect of the language is often overlooked in language instruction, leading to difficulties for learners. Research conducted by scholars such as Mazuka et al. (2002) and Tamaoka et al. (2005, 2014) has revealed that even native Japanese children encounter challenges in understanding this type of sentence. Therefore, it is understandable why foreign learners of Japanese tend to make mistakes regarding this type of sentence. Therefore, both teachers and students need to acknowledge the significance of scrambled sentences in Japanese and allocate sufficient time and effort to studying and mastering their use. By doing so, learners can develop a comprehensive understanding of the language and achieve a higher level of proficiency in communicating in Japanese.

5   Conclusion

The proficient utilization of interrogative words constitutes a vital aspect of the Japanese grammatical system. The current research endeavor scrutinizes the proficiency of Vietnamese students in employing Japanese *wh*-words within question sentences. Although both Japanese and Vietnamese are considered

wh-in-situ languages, the syntactical differences between these languages create a significant hurdle for Vietnamese students when attempting to grasp the proper order of *wh*-words in Japanese questions.

The study recruited 65 high school students who have studied Japanese for varying durations (ranging from 2.5 years to 13 years) at the Foreign Language Specialized School, which is affiliated with the VNU University of Languages & International Studies, Vietnam National University, Hanoi.

The findings of this study corroborate the underlying hypotheses. Firstly, students with higher levels of Japanese proficiency (i.e., intermediate or advanced levels) demonstrated a better ability to answer survey questions in comparison to those at lower proficiency levels. Secondly, students displayed a greater aptitude for accurately responding to Japanese *wh*-questions that share a similar word order with Vietnamese, especially 'why' and 'when' questions.

The research showed that learners could acquire L2 *wh*-constructions with different parametric values from their L1, as demonstrated by some advanced learners. However, achieving target-like interpretations of *wh*-phrases in Japanese proved challenging.

Effective instruction on the specific word order of Japanese *wh*-questions plays a pivotal role in enhancing students' accuracy rates. However, the existing introductory Japanese language curriculum primarily focuses on providing basic information about the meaning and usage of interrogative words, neglecting to take into account their syntactic role within a sentence or the specific sentence structures in which they are utilized. Therefore, teachers must convey this information clearly and comprehensively. Even for novice learners of Japanese, introducing fundamental concepts related to question words would be beneficial. Moving beyond the basics, intermediate or advanced-level students stand to benefit significantly from learning scrambled sentences. This approach facilitates a deeper and more nuanced understanding of the usage of Japanese *wh*-words and their appropriate placement within sentences.

It is essential to create opportunities for the development of both explicit and implicit knowledge. However, it is crucial not to assume that explicit knowledge will automatically translate into implicit knowledge. Learning a language is a gradual process that requires time. While instruction is valuable for drawing learners' attention to grammatical forms, the key to developing implicit knowledge lies in consistent exposure to meaningful input and regular practice. Therefore, it is vital to provide opportunities for learners to

repeatedly use target grammatical forms in meaningful communicative contexts, fostering a gradual and comprehensive language acquisition process.

Abbreviations

ACC   accusative
CLF   classifier
DAT   dative
DEM   demonstrative
NOM   nominative
PERF   perfect
POL   politeness
Q   question
TOP   topic

Contributions

Authorship credit for this paper is as follows: Trang Phan authored Chapters 1 and 2, Sections 3.1 and 3.3, as well as Chapter 5. Giang Vu authored Section 3.2 and Chapter 4.

References

Beck, S., & Kim, S-S. (1997). On Wh- and Operator Scope in Korean. *Journal of East Asian Linguistics, 6*, 339–384.

Cao, X. H. (2004). *Tiếng Việt sơ thảo ngữ pháp chức năng* [A preliminary functional grammar of Vietnamese]. Hanoi: Education Publishing House.

Cheng, L., & Rooryck, J. (2000). Licensing Wh-in-situ. *Syntax, 3*, 1-19.

Chomsky, N. (1970). Language and freedom. Lecture presented at the University Freedom and Humane Sciences Symposium, Loyola University, Chicago, 8 Jan. 1970. In A. Arnove (Ed.), *The Essential Chomsky* (2008, pp. 75-91). New York: New Press.

Dryer, M. S. (1991). SVO languages and the OV:VO typology. *Journal of Linguistics, 27*, 443–482.

Duffield, N. (2007). Aspects of Vietnamese Clausal Structure: Separating tense from assertion. *Linguistics, 45*(4), 765-814.

Gunji, T. (2011). *Nihongo wa donna gengo ka? - Ruikeironteki kanten kara no Nihongo* [What kind of language is Japanese?—From a typological point of view]. *Theoretical and Applied Linguistics at Kobe Shoin: Tookusu, 14.*

Hawkins, J. (2004). *Efficiency and complexity in grammars.* Oxford: Oxford University Press.

Huang, J. (1982). Move WH in a Language without Wh-movement. *The Linguistic Review, 1*(4), 369-416.

Jorden, E. H. (1962). *Japanese: The Spoken Language. Part 1.* Yale University Press.

Kitagawa, Y. (2005). Prosody, Syntax, and Pragmatics of Wh-questions in Japanese. *English Linguistics, 22*(2), 302-346.

Ko, H. (2005). Syntax of Why-in-situ: Merge Into [SPEC,CP] in the Overt Syntax. *Natural Language & Linguistic Theory, 23*, 867–916.

Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon Press.

Lieberman, M., Aoshima, S., & Phillips, C. (2006). Nativelike biases in generation of wh-questions by nonnative speakers of Japanese. *Studies in Second Language Acquisition, 28*, 423-448.

Lin, J.-w. (1992). Syntax of *Zenmeyang* 'How' and *Weishenme* 'Why' in Mandarin Chinese. *Journal of East Asian Linguistics, 1*, 293–331.

Mazuka, R., Itoh, K., & Kondo, T. (2002). Costs of scrambling in Japanese: An Eye-tracking Investigation of Pre-head and Head-driven Processing for Scrambled Japanese sentence processing. In M. Nakayama (Ed.), *East Asian Language Processing* (pp. 131–166). Stanford, CA: CSLI Publications.

Minna no Nihongo Shokyuu 1 Honsatsu [Everyone's Japanese Elementary 1 Textbook] (March 1, 1998). Suriiee nettowaaku Publishing House.

Miyagawa, S. (2003). A-movement scrambling and options without optionality. In S. Karimi (Ed.), *Word Order and Scrambling* (pp. 177–200). Oxford, UK: Blackwell.

Pesetsky, D. (1987). Wh-in-situ: Movement and Unselective Binding. In E. Reuland & A. ter Meulen (Eds.), *The Representation of (In)definiteness* (pp. 98-129). Cambridge, MA: MIT Press.

Ringbom, H. (2006). *Cross-linguistic Similarity in Foreign Language Learning*. Bristol, Blue Ridge Summit: Multilingual Matters.

Saito, M. (1992). Long distance scrambling in Japanese. *Journal of East Asian Linguistics, 1*(1), 69–118.

Saito, M. (1985). *Some Asymmetries in Japanese and Their Theoretical Implications*. Cambridge, MA: MIT Press.

Tamaoka, K., Sakai, H., Kawahara, J.-i., Lim, Y., Miyaoka, H., & Koizumi, M. (2005). Priority information used for the processing of Japanese sentences: Thematic roles, case particles, or grammatical functions? *Journal of Psycholinguistic Research, 34*(3), 281–332.

Tamaoka, K., Asano, M., Miyaoka, Y., & Yokosawa, K. (2014). Pre- and post-head processing for single-and double-scrambled sentences of a head-final language as measured by the eye tracking method. *Journal of Psycholinguistic Research, 43*, 167–185.

Tomioka, S. (2007). Pragmatics of LF intervention effects: Japanese and Korean Wh-interrogatives. *Journal of Pragmatics, 39*(9), 1570-1590.

Tsai, E. (2009). *Wh-dependency in Vietnamese and the Syntax of Wh-in-situ* (Master's thesis). National Tsing Hua University, Taiwan.

Tsai, W.-t. D. (1999). The Hows of Why and The Whys of How. In F. Del Gobbo & H. Hoshi (Eds.), *UCI Working Papers in Linguistics, 5* (pp. 155–184). UC Irvine.

Tran, T. (2009). *Wh-quantification in Vietnamese*. (Doctoral dissertation). University of Delaware.

Tran, T. (2021). Non-canonical word order and temporal reference in Vietnamese. *Linguistics, 59*(1), 1-34.

Umeda, M. (2008). *Second language acquisition of Japanese Wh-constructions* (Doctoral dissertation). McGill University.

Zhang, Y., & Hayashi, Y. (2000). *Chinese/Japanese/English Contrast Practical Chinese Conversation*. Tokyo: Hakusuisha Publishing House.

Sino-Vietnamese Bilingual Dictionaries from the Middle Ages to the Present

Duc Du HUYNH
School of International Studies, Zhengzhou University, China
dr_huynhducdu@zzu.edu.cn

Abstract

Sino-Vietnamese bilingual dictionaries, spanning the Middle and Modern Ages, are crucial sources for studying East Asian Chinese character lexicography evolution. Despite their significance, this article addresses the scholarly community's limited attention to this area. Focusing on General Survey, Developmental Stages, and Intrinsic Value, we explore Sino-Vietnamese bilingual dictionaries' historical trajectory, categorizing it into three key epochs. These epochs are further divided based on regional disparities and different compilers, offering a nuanced exploration of major historical events in Vietnam. The goal is to present a comprehensive panorama, laying a solid groundwork for future scholarly inquiry.

Keywords: Sino-Vietnamese bilingual dictionaries, comprehensive Exploration, staging, Middle Ages, Modern Ages

Povzetek

Sino-vietnamski dvojezični slovarji od srednjega veka do moderne dobe služijo kot ključni viri za preučevanje razvoja dvojezične leksikografije v okviru kitajskih pismenk v vzhodni Aziji. Kljub njihovi pomembnosti je znanstvena skupnost temu področju namenila le omejeno pozornost, pri čemer je pustila številne vidike neizkoriščene. Ta članek ponuja celovit pregled trenutnega stanja, pri čemer se osredotoča na tri osnovne perspektive: splošni pregled, razvojne faze in osnovno vrednost le-teh. Zgodovinski razvoj dvojezičnih kitajsko-vietnamskih slovarjev razdelimo na tri ključna obdobja, ki jih dalje delimo na različne podkategorije glede na regionalne razlike ter variacije, ki so posledica različnih sestavljavcev. Glavni cilj članka je predstaviti celovit pogled na dvojezične sino-vietnamske slovarje od srednjega veka do moderne dobe ter položiti trdne temelje za nadaljnje znanstveno raziskovanje in napredek.

Ključne besede: sino-vietnamski dvojezični slovarji, celovit pregled, razvrstitev, srednji vek, moderna doba

1   Introduction

East Asian dictionaries trace their origins to China, serving as repositories of knowledge and vital sources of clarification. With the diffusion of Chinese characters and culture to neighboring countries, diverse Chinese dictionaries found their way to Japan, Korea, and Vietnam, facilitating enhanced interpretation and mastery of the Chinese language. During this period, Chinese characters assumed the role of the official script in each country. As linguistic and writing environments matured, individual nations developed distinct writing systems grounded in Chinese characters, tailored to the idiosyncrasies of their respective languages.

Subsequently, to enhance the comprehension of Chinese characters and texts, successive generations of scholars engaged in the interpretation, translation, and exegesis of Chinese characters within their linguistic frameworks, giving rise to bilingual dictionaries elucidating Chinese characters. Notably, Vietnam, profoundly influenced by Chinese culture, assimilated Chinese characters and cultural elements into its local ethos. However, the compilation of dictionaries in Vietnam commenced later than in Japan and Korea and has received limited scholarly attention.

Vietnam's tropical climate and tumultuous historical backdrop, marked by conflicts, have contributed to the scarcity of surviving ancient texts and a paucity of scholarly articles on Sino-Vietnamese characters in bilingual dictionaries from the Middle Ages to the Modern Ages.

2   Research overview and general survey of dictionaries

Sino-Vietnamese characters featured in bilingual dictionaries from the Middle Ages to the Modern Ages constitute a vital and indispensable resource for delving into the realm of East Asian Chinese dictionaries and the historical trajectory of dictionary compilation. An examination of existing research outcomes pertaining to Sino-Vietnamese characters in bilingual dictionaries of the Middle Ages and the Modern Ages underscores the limited scholarly attention devoted to this subject. The number of scholars engaging in such study remains notably modest, particularly in China, where the broader Vietnamese aspect of East Asian Chinese literature is comparably underexplored.

This article endeavors to present the collective findings of prior generations and esteemed scholars, offering a macroscopic perspective that encompasses both domestic and international scholarly endeavors.

2.1   Overview of research in China

Studies on extraterritorial Chinese dictionaries in China, particularly analytical investigations into the historical evolution of Sino-Vietnamese bilingual dictionaries from the Middle Ages to the Modern Ages and their distinctive characteristics, still need to be explored. Despite the limited scholarly output, certain scholars have endeavored to explore and examine this subject, contributing valuable information and research insights.

Liang Maohua (2017) provides a succinct overview of the compilation style and characteristics of four notable Sino-Vietnamese characters in bilingual dictionaries, namely *Chi Nam Ngoc Am Giai Nghia* 指南玉音解义, *Tam Thien Tu Giai Am* 三千字解音, *Tu Duc Thanh Che Tu Hoc Giai Nghia Ca* 嗣德圣制字学解义歌, and *Dai Nam Quoc Ngu* 大南国语. The analysis reveals significant divergences in the layout of Sino-Vietnamese characters when compared to ancient Chinese dictionaries, notably their organization into alternating six and eight rhymes for enhanced recitation and memorization. While this article provides a brief introduction to the four dictionaries, it refrains from in-depth thematic examination, thereby offering a foundation for future research. Additionally, Liang Maohua's doctoral thesis, entitled *A Study of the History of Vietnamese Script Development*, incorporates a subsection exploring the characteristics of the dictionaries as mentioned above. In conclusion, he posits that the distinctive layout features of Sino-Vietnamese bilingual dictionaries diverge significantly from their Chinese counterparts, rendering them a valuable resource for understanding the trends and usage characteristics of Nom characters' development.

Furthermore, by centering the investigation on *Chi Nam Ngoc Am Giai Nghia*, Wen Min (2017) discerned that this dictionary meticulously preserves a multitude of lexical items dating back to the Tang, Song, Yuan, and Ming dynasties, rendering it a valuable resource for historical research. Numerous words within the dictionary have undergone a process of Vietnamization and are identified as Nam-origin words. The study extends its focus to an examination of the vocabulary contained in this dictionary, revealing discernible traces of cultural exchange between the two nations, thus serving as a crucial reference for the compilation of Chinese dictionaries. Within the realm of philology, the author expounds upon the preservation of Vietnamese Chinese character patterns and conducts comparative analyses between Chinese characters and Nom characters, laying the groundwork for future research endeavors.

2.2   Overview of research abroad

The outcomes within this category are scant in number and predominantly concentrate on traditional Chinese dictionaries from the Middle Ages, thus needing more focused results.

Phan Van Cac (1994) provides a concise overview of *Chi Nam Ngoc Am Giai Nghia*, Dao Duy Anh's *Chinese-Vietnamese Dictionary* (1932), Van Tan's *Chinese-Vietnamese Dictionary* (1957), and Phan Van Cac's *Chinese-Vietnamese Dictionary* (1994). The author delineates the fundamental strengths and weaknesses of each dictionary. In conclusion, the article briefly addresses lexical issues within Chinese lexicography, encompassing Chinese characters and typography, serving as a valuable reference for future compilations of similar dictionaries.

Le Anh Tuan (2002) conducts a quantitative and categorical analysis of Sino-Vietnamese bilingual books, with a specific focus on dictionaries. Utilizing *Chi Nam Ngoc Am Giai Nghia* and *Ngu Thien Tu* 五千字 as exemplars; the author classifies them into various sub-categories. While serving as a concise introduction to Sino-Vietnamese bilingual books, the article serves as a catalyst for future research in the domain of dictionaries.

In a study aligning with the subject and methodology similar to this research, Tran Trong Duong (2017) categorizes Sino-Vietnamese dictionaries into four distinct types: monolingual dictionaries, Chinese-Nom dictionaries, Chinese-Nom-Thai multilingual dictionaries, and Chinese-Nom-Latin/French multilingual dictionaries. The article underscores the Chinese-Nom bilingual dictionary as a distinctive document originating in Vietnam during the Middle Ages, emerging within a bicultural, bilingual, and bi-literate milieu. Notably, it functions as a unique hybrid of dictionaries and textbooks, aimed at instructing Chinese characters and basic Confucian Chinese. This insight underscores the distinct functional disparity of Sino-Vietnamese characters in bilingual dictionaries compared to their counterparts in China, Japan, and Korea.

Conducting a fundamental investigation, both quantitative and qualitative, of the lexical components in Sino-Vietnamese characters within bilingual dictionaries through the lens of lexicographical theory, Ha Dang Viet (2014) provides a comprehensive summary of the common attributes inherent in this series. The research material employed encompasses *Nhat Dung Thuong Dam, Tu Duc Thanh Che Tu Hoc Giai Nghia Ca, Nam Phuong Danh Vat Bi Khao*, and *Dai Nam Quoc Ngu*. Predominantly, the noteworthy features of nineteenth-century Vietnamese dictionaries lie in their character as 'political' texts, imparting instruction on fundamental Chinese characters and catering to imperial examination needs. The scope of this article encompasses a rudimentary exploration of the lexicon exclusively, leaving untouched the core aspects of

characters, exegesis, and phonology. Nonetheless, it represents a foundational reference material for examining Vietnamese Chinese dictionaries.

Moreover, La Minh Hang (2017) offers an observation grounded in Chinese lexicography theory, concurrently analyzing *Dai Nam Quoc Ngu*, *Nhat Dung Thuong Dam*, and *Nam Phuong Danh Vat Bi Khao*. The study identifies three predominant characteristics inherent in Vietnamese dictionaries, of which the first is that Vietnamese dictionaries do not align with the attributes of Chinese encyclopedias. The second characteristic is that Vietnamese lexicographic tools are categorized into various types, including dictionaries, language dictionaries, language thesauruses, encyclopedic dictionaries, and reference books. Finally, Vietnamese dictionaries from the Middle Ages predominantly adhere to the three talents, namely heaven, earth, and man as the prevailing editorial paradigm, featuring a macro structure organized hierarchically from larger to smaller entities.

Hoang Thu Thuy (2019) delineates the present status of Sino-Vietnamese bilingual dictionaries from the Middle Ages, briefly shedding light on features such as rhyming in the carrier form and the multifunctionality embodied in the microstructure. However, the exploration remains confined to an introductory level and does not delve into ontological research.

Additionally, Trinh Khac Manh (2011) conducts a chronological study that categorizes the development of Sino-Vietnamese bilingual dictionaries into distinct stages. The argument posits the initiation of Sino-Vietnamese characters in the fifteenth century, extending until the nineteenth century. This period is principally divided into two parts, the first being a tool for interpreting the Chinese language in Nom script and the second a tool for compiling Vietnamese history, geography, culture, and education in Chinese characters.

Moving into the twentieth century, both Chinese characters and Nom characters gradually waned, with Quoc Ngu characters assuming dominance. Consequently, all Sino-Vietnamese language tools were compiled in Quoc Ngu characters, and were further categorized into four groups; 1. tool for interpreting Chinese characters, 2. Chinese character font cross-reference toolkit, 3. specialist tool, and 4. tool for professional study and research in Chinese and Nom characters.

The author holds an optimistic perspective on the tools for Chinese and Nom characters, intending to develop an extensive toolkit for the Chinese and Nom characters corpus. Despite recognizing the potential for advancement in Chinese and Nom script tools, the author acknowledges that limited human resources, inadequate funding, and a scope that does not cater to a broad audience remain challenges.

The only comprehensive chronological division of Sino-Vietnamese bilingual dictionaries is presented by Tran Van Chanh (2018), who categorizes them into three distinct periods. Early dictionaries belong to the period before 1930 and are exemplified by works such as *An Nam Dich Ngu* 安南译语, *Chi Nam Ngoc Am Giai Nghia*, *Nhat Dung Thuong Dam*, etc. Following is the transitional period from 1930 to 1950, which is represented by Dao Duy Anh's *Chinese-Vietnamese Dictionary* (1932), Nguyen Tran Mo's *Nam Hoa Dictionary* (1942), Thieu Chuu's *Chinese-Vietnamese Dictionary* (1942), etc. Finally, works like Van Tan's *Chinese-Vietnamese Dictionary* (1957), Nguyen Van Khon's *Chinese-Vietnamese Dictionary* (1960), Ly Van Hung's *New Chinese-Vietnamese Dictionary* (1960), etc. feature the last period, the so-called period of modern dictionaries from 1950 to the present.

The author appears to designate Dao Duy Anh's *Chinese-Vietnamese Dictionary* (1932) as a pivotal point, demarcating ancient dictionaries that precede it and modern dictionaries that succeed it. A concise overview of representative works from each period is also provided. The study extends its scope to encompass various specialized dictionaries, including Buddhist dictionaries, medical dictionaries, scientific dictionaries, and specialized tools, contributing some reference value to the historical understanding of the compilation of Chinese dictionaries in Vietnam.

Essentially, these investigations represent an initial foray into the fundamental dimensions of Sino-Vietnamese bilingual dictionaries from the Middle Ages to the Modern Ages. They establish a foundation and offer conceptual frameworks for further in-depth inquiry, providing a preliminary understanding of the developmental history of Vietnamese dictionaries. However, it is crucial to acknowledge that the current research is foundational, needing a detailed reflection of the distinctive characteristics and core values inherent in each chronological period. This lacuna offers a promising avenue for subsequent research to delve into and elucidate the intricacies of the subject matter.

3   Sino-Vietnamese bilingual dictionaries: Stages of compilation through time

The complete abolition of the Vietnamese imperial examination system in 1919 marked a significant turning point. This era also witnessed the initial invasion by French colonists from 1858 to 1945, followed by the establishment of the Democratic Republic of Vietnam in 1945 and the resolution of the war in the North. Subsequently, from 1945 to 1954, the South experienced a re-invasion by French colonists, leading to the establishment of the Republic of Vietnam

from 1954 to 1975. The culmination of these historical events unfolded with the complete reunification of Vietnam in 1975. This article explores the compilation of Sino-Vietnamese characters in bilingual dictionaries in the Middle Ages and the Modern Ages, situated amidst the nexus of these pivotal events in Vietnamese history.

3.1 Classical bilingual dictionaries of Sino-Vietnamese characters (approx. 1600-1950)

Since the introduction of Chinese characters to Vietnam and their adoption as the official script, they served as the primary means of recording Vietnamese politics, history, and culture until 1919. The ancient Vietnamese held a deep admiration for Chinese culture, often referring to Chinese characters as 'Confucian characters' or 'sage characters', among other designations. These characters were utilized for transcribing Chinese classics, crafting Vietnamese literature, and even forming Nom characters based on components of Chinese characters. In the pursuit of understanding the classical Chinese language, engaging with classical Chinese literature, and participating in imperial examinations for career prospects, Vietnamese scholars compiled an array of tools spanning different eras. Unfortunately, due to Vietnam's humid climate and the numerous wars, the preservation of ancient dictionaries has been scarce.

The seventeenth-century work, *Chi Nam Ngoc Am Giai Nghia*, is currently acknowledged by scholars as the oldest extant Sino-Vietnamese bilingual dictionary. However, the exact origin of traditional Sino-Vietnamese bilingual dictionaries remains to be determined, even though their zenith likely occurred between the eighteenth and early twentieth centuries.

At a macroscopic level, the period under consideration for Sino-Vietnamese bilingual dictionaries predominantly encompasses resources aimed at children's literacy. However, it also extends to encyclopedias, rhyming books, Buddhist dictionaries, and script dictionaries.

Children's literacy dictionaries, such as *Tam Thien Tu Giai Am* 三千字解音 (1831), *Tu Hoc Huan Mong* 字学训蒙 (1877), *Tu Hoc Cau Tinh Ca* 字学求精歌 (1879), *Thien Tu Van Giai Am* 千字文解音 (1890), *Tu Hoc Tu Ngon Thi* 字学四言诗 (1882), *Tam Thien Tu Giai Dich Quoc Ngu* 三千字解译国语 (1915), etc., are primarily composed in the Vietnamese *lục bát* (six-eighths) poetic style for facile memorization. These dictionaries are both captivating and informative, covering a wide spectrum of subjects, including astronomy, geography, human ethics, flora, fauna, and rituals. Noteworthy examples like *Tam Thien Tu*

Giai Am and *Thien Tu Van Giai Am* utilize a gloss reading approach, where a Chinese character corresponds to Nom characters or Quoc Ngu characters.

Encyclopedic dictionaries from this period encompass notable works such as *Chi Nam Ngoc Am Giai Nghia* (17th century), *Nhat Dung Thuong Dam* (1827), *Tu Duc Thanh Che Tu Hoc Giai Nghia Ca* (1897), *Dai Nam Quoc Ngu* (1899), *Nam Phuong Danh Vat Bi Khao* (1901), *Chi Nam Bi Loai* 指南备类 (unknown), *Tu Loai Dien Nghia* 字类演义 (unknown), and others.

Buddhism dictionaries include *Van Phap Chi Nam* 万法指南 (1894) and *Dao Giao Nguyen Luu* 道教源流 (1845).

During this period, bilingual dictionaries of Sino-Vietnamese characters exhibit distinctive characteristics:

1. Prevalence of children's literacy and encyclopedic dictionaries: The era is marked by a notable abundance of children's literacy dictionaries and encyclopedic dictionaries while rhyming dictionaries and script dictionaries are comparatively fewer. Children's literacy dictionaries predominantly incorporate commonly used Chinese characters, and most encyclopedic dictionaries are of medium size.

2. Nationalized nature of children's literacy and encyclopedic dictionaries: Children's literacy dictionaries and encyclopedic dictionaries often exhibit a nationalized character. They employ unique Vietnamese poetic styles, such as the alternating six and eight, or double seven, six, and eight rhymes, serving as carriers. The content is interpreted either in Nom characters or Quoc Ngu characters. While a minority of dictionaries may employ a blend of Chinese characters and Nom characters, the predominant usage remains with Nom characters.

3. Predominant compilation by private Confucian scholars: The majority of dictionaries during this period were compiled by private Confucian scholars, with a relatively low percentage attributed to official dictionaries.

4. Influence of Chinese style and content: Dictionaries from this period bear evident traces of Chinese style and content. The style mirrors that of contemporary Chinese texts, and the content closely resembles the selection found in Chinese texts. However, as this period progresses, dictionary compilers begin to consider the distinctive characteristics of the Vietnamese language and the language habits of the people. Consequently, the selection of phonetics, vocabulary, and script aligns more closely with the practical needs and preferences of the Vietnamese populace. The content and style progressively localize, emphasizing practicality, modernity, and simplicity.

5. Association with other texts: Dictionaries from this period did not hold high status independently but were often appended to scriptures and other texts. For instance, *Tang Dinh Au Hoc Tu Tri Tap Tu Thai Tran Dai Toan* 增订幼学须知杂字采珍大全 is attached to the end of *Nhat Dung Thuong Dam* 日用常谈, and *Tam Thien Tu Toan Yeu* 三千字纂要 is appended to the end of *Dao Giao Nguyen Luu* 道教源流.

3.2   Emerging paradigms in bilingual dictionaries of Sino-Vietnamese characters (approx. 1950-)

The abolition of the imperial examination system in 1919, the diminishing influence of Chinese characters in Vietnamese society, and the pervasive influx of Western culture collectively contributed to the complete departure of Chinese characters and Nom characters from the historical narrative. Concurrently, as the movement advocating for the Quoc Ngu character gained prominence, bilingual dictionaries of Sino-Vietnamese characters underwent a transformative phase during the early twentieth century, signifying a pivotal shift from traditional to modern (Western) codification.

During the initial decades of the twentieth century, only a handful of Confucian scholars persisted in upholding Confucian education, while Western missionaries actively propagated Western education on a grand scale. This catalyzed the emergence of numerous Confucian scholars in Vietnam who possessed proficiency in both Chinese and Western languages. Truong Vinh Ky stands as a prominent representative of this era, contributing significantly to the compilation of numerous Chinese-Western, Chinese-Vietnamese, and Western-Vietnamese bilingual dictionaries.

Simultaneously, Western missionaries engaged in compiling a considerable number of Chinese-French, Chinese-Vietnamese, and French-Vietnamese dictionaries to promote Catholicism and disseminate Western knowledge. The early twentieth century thus stands as a critical juncture in the evolution of Sino-Vietnamese bilingual dictionaries, marking the transition from traditional to modern frameworks influenced by Western linguistic paradigms.

3.2.1   A transitional dictionary for missionaries (approx. 1900-1950)

Alexandre de Rhodes's *Dictionarium Annamiticum Lusitanum et Latinum* (1651) among foreign scholars stands as the pioneering effort in the compilation of dictionaries related to the Vietnamese language. Subsequently, with the onset and progression of Western colonial rule, a wave of Western missionaries arrived in Vietnam. While engaged in preaching, these missionaries undertook

the compilation of dictionaries, primarily intended to facilitate the dissemination of their teachings.

During this missionary period, there was a deliberate acknowledgment of Traditional Sino-Vietnamese in bilingual dictionaries. Missionaries often fused traditional Chinese bilingual dictionaries with new Western dictionaries, extracting the strengths of each, resulting in the creation of several new bilingual dictionaries tailored for Vietnamese readers.

Depending on the thematic focus of the compilation, these dictionaries are broadly categorized into four main types. Firstly, radical arrangement following the Kangxi dictionary can be found in G. Pauthier's *Dictionnaire étymologique Chinois-Annamite Latin-Français* (1867) and Charles Edouard Bailly's *Dictionnaire Chinois - Français* (1889). Following are the dictionaries arranged in Latin alphabetical order, such as Jean-Louis Taberd's *Dictionarium Anamitico-Latinum* (1867) and F. M. Génibrel's *Dictionnaire Annamite-Français* (1898). The third are children's literacy dictionaries, of which Edmond Nordemann's *Petit vocabulaire Chinois-Annamite-Français* (1895) is the best representative. Finally, the fourth type are Nom-foreign dictionaries, such as the one by Jean Bonet's *Dictionnaire Annamite-Français* (1899-1900).

During this era, bilingual dictionaries of Sino-Vietnamese characters exhibit the following characteristics:

1. Dominance of Western missionary compilation: The majority of the new type of bilingual dictionaries of Sino-Vietnamese characters from this period were crafted by Western missionaries who arrived with colonial influence.

2. Rich and diverse content, varied styles: These dictionaries boast abundant and diverse content, often adopting either Kangxi dictionary headings or Latin alphabetical phonetic order for organization and accessibility.

3. Locale-centric characteristics: Chinese characters within these dictionaries exclusively represent native Vietnamese characters, and the pronunciations are rooted in southern Vietnamese dialects. The meanings employed in these dictionaries predominantly align with southern Vietnamese vocabulary, establishing a more localized essence compared to traditional Sino-Vietnamese bilingual dictionaries.

4. Shift from traditional compilation style: The new type of bilingual dictionaries of Sino-Vietnamese characters marks a decisive departure from the traditional style of dictionary compilation, signaling a transition towards contemporary Vietnamese dictionary methodologies.

5. Initiative by Vietnamese scholars in the Enlightenment period: Concurrently, Vietnamese scholars undertook the independent compilation of

new dictionaries during the Enlightenment and Development period, spanning from the early twentieth century to 1975.

3.2.2 Independent compilations: New dictionaries by Vietnamese scholars during the Era of enlightenment and development (1900-1975)

During the period from the early to mid-twentieth century, a dynamic interplay unfolded between Western missionaries compiling the new type of Sino-Vietnamese bilingual dictionaries alongside their traditional counterparts, and Vietnamese Confucians, influenced by these developments, began contemplating the compilation of this innovative genre. Possessing a Confucian education, these scholars found themselves at the intersection of tradition and modernity as Vietnam underwent increased colonization. Fluent in both Western languages and Quoc Ngu characters, they diverged from the traditional Confucian stance and exhibited receptiveness to the new type of Sino-Vietnamese bilingual dictionaries crafted by Western missionaries.

From the early twentieth century until 1975, these intellectuals, often referred to as the new Confucians charted a course aligned with the trajectory of the evolving dictionaries. Grounded in the practicalities of Vietnam's evolving landscape, they spearheaded the compilation of a new type of Sino-Vietnamese bilingual dictionaries.

Between 1945 and 1975, cultural development in the Democratic Republic of Vietnam (Northern Vietnam) experienced a temporary slowdown as post-war economic recovery took precedence, with the entire population engaged in social and productive construction. Consequently, the compilation of dictionaries was briefly suspended, although individual scholars authored noteworthy dictionaries. In contrast, the Republic of Vietnam (Southern Vietnam) witnessed stable social development, a relatively robust economy, and a conducive academic research environment. Confucianism found widespread acceptance in the South during this period.

Compilation of new-style dictionaries in Northern Vietnam

The primary entity involved in the compilation of new dictionaries in northern Vietnam was the Khai Tri Tien Duc Society (1919-1945). This society played a pivotal role in fostering the cultural advancement of Northern Vietnam, comprising key members who were influential ministers of the Nguyen dynasty and officials of the Indochina political office in Northern Vietnam. Among their influential contributions, the *Vietnamese Dictionary* (1931), edited by Pham Quynh, Nguyen Van Vinh, Tran Cong Kim, and Duong Bac Khao,

stands out prominently. This dictionary has significantly shaped the subsequent development of Sino-Vietnamese bilingual dictionaries, serving as a foundational reference.

Organized in Latin phonetic order and spanning 663 pages with 24,500 entries, the Vietnamese Dictionary marked the inception of self-compiled Sino-Vietnamese bilingual dictionaries in Vietnam. Noteworthy for not only recording general vocabulary but also incorporating specialized terms encompassing Western thought and mechanical technology, the dictionary further includes vocabulary from the diverse dialects of North, South, and Central Vietnam. Its launch sparked extensive scholarly commentary at the time and underwent multiple reprints by Tan Van Publishing House and Mac Lam Publishing House upon reaching the southern regions of Vietnam. This enduring reprinting attests to the significant value of the Vietnamese Dictionary in the study of the history of the Sino-Vietnamese bilingual dictionary compilation.

Another influential work during this period in northern Vietnam was the *Chinese-Vietnamese Dictionary* (1932), also known as *The Chinese-Vietnamese Dictionary in Brief*, authored by Dao Duy Anh. This dictionary, a collaborative effort with his wife, compiled generic Chinese and Vietnamese words from various sources, including *Nam Phong Tap Chi* 南风杂志, classical texts like *Cung Oan Ngam Khuc* 宫怨吟曲 and *Truyen Kieu* 传翘, and Chinese dictionaries such as *Ci Yuan* 辞源, *Kangxi zidian* 康熙字典, and *The Dictionary of the Chinese State Language* 中国国语大辞典. Beyond compilation, Dao Duy Anh provided interpretations for each collected Chinese and Vietnamese word.

The *Chinese-Vietnamese Dictionary* garnered significant attention from contemporary scholars and has maintained enduring relevance, evidenced by its continuous reprinting. Compiled with the explicit goal of revitalizing Confucianism, the dictionary provides straightforward explanations for over 5,000 individual characters. Its focus lies predominantly on meanings applicable in contemporary society, avoiding in-depth explorations of the original character meanings. Dao Duy Anh's work also introduces some challenges regarding the pronunciation of numerous Chinese characters, reflecting either the author's pronunciation or the prevailing vernacular of the time, diverging from the original Chinese pronunciation. Examples include 吃 (pronounced ngật, originally cật) and 讧 (pronounced công, originally hồng)

Another notable bilingual dictionary from this period in northern Vietnam is Thieu Chuu's *Chinese-Vietnamese Dictionary* (1942). Compiled to promote Buddhist culture and facilitate scripture reading, this dictionary exhibits distinctive features compared to its predecessors. One of such features is that it predominantly comprises "explanatory single characters," 8,318 characters in

total. It also utilizes the traditional 214 radical ordering instead of the prevalent Latin phonetic ordering and focuses the interpretation section on the ancient meanings of words rather than their contemporary meanings in Vietnam. It further retains Fanqie pronunciation for words, although an allophone and a vernacular pronunciation accompany some characters, incorporates examples from Buddhist scriptures or other classical texts, and employs simple and comprehensible illustrations to convey concepts that may be challenging to articulate verbally.

Despite its merits, Thieu Chuu's dictionary does have certain drawbacks, notably the absence of a phonetic checklist at the end of the book. This omission hinders reader searches and complicates usability.

Additionally, there is the noteworthy *Tam Nguyen Tu Dien* (1941) by the northern Vietnamese scholar Le Van Hoe. This publication represents an etymological exploration of Vietnamese words derived from Chinese origins, presented in Latin alphabetical order. Delving into the original meanings of these words, the work also investigates their etymological roots, the principles behind their construction, and the evolution of their meanings. As such, it stands as a valuable resource for comprehending words of Chinese origin.

During this era, the characteristics of Sino-Vietnamese characters in bilingual dictionaries are delineated by:

1. A relatively limited number of dictionaries were produced in northern Vietnam, primarily concentrated before 1945, with a notable scarcity in production between 1945 and 1975.

2. With the exceptions of Thieu Chuu's *Chinese-Vietnamese Dictionary* and Nguyen Tran Mo's *Nam Hoa Dictionary*, all other dictionaries adopt a phonetic arrangement. This pattern suggests that the compilation of Sino-Vietnamese bilingual dictionaries during this period predominantly followed the Western missionary approach.

Compilation of new-style dictionaries in Southern Vietnam

In contrast to the 844-year history of Confucianism in northern Vietnam, the southern region's connection with this philosophical tradition spans a shorter duration of 52 years. The prominence of Confucianism is evidently more deeply entrenched in northern Vietnam than in the South. However, the widespread dissemination of Confucianism in the southern region, despite its relatively shallow historical roots, holds significant implications for understanding the characteristics of dictionaries in this period. Therefore, it is imperative to analyze the factors contributing to the extensive reach of Confucianism in the South, where its historical foundations are less profound.

Several factors elucidate the widespread influence of Confucianism in southern Vietnam. Following the establishment of the socialist system in northern Vietnam, Confucianism was viewed as conservative, antiquated, and feudal, prompting continuous suppression. Many Confucian scholars relocated to the South and played a role in the establishment of the Republic of Vietnam. Consequently, Confucianism regained importance in the early years of the state. The renewed study of Confucian classics and the founding of Confucian-related associations, such as the Vietnam Confucian Association (VAC) and Vietnam Association of Ancient Studies (VATS), further fueled the development of Confucianism in the southern region.

The stable social development, relatively advanced economy, and conducive academic research environment in the Republic of Vietnam (South) contributed to the resurgence of Confucianism. Additionally, southern Vietnam's global connections, frequent exchanges with Hong Kong, Macao, and Taiwan, along with an increasing population of Chinese expatriates, played a role in fostering the development of Confucianism during this period.

In the context of southern Vietnam's robust promotion of Confucianism, the compilation of Sino-Vietnamese bilingual dictionaries during this period served multiple purposes, such as Confucianism promotion, where the dictionaries aimed to promote Confucianism, aligning with the prevalent influence of this philosophical tradition in southern Vietnam. Another objective was the preservation of ancient cultural heritage, safeguarding traditional values and linguistic elements. The third objective was the enhancement of Vietnamese and French language skills for Overseas Chinese. The dictionaries, in other words, sought to aid the overseas Chinese population in southern Vietnam by improving their proficiency in both Vietnamese and French languages. Facilitation of Western Missionary activities, in which these dictionaries served as instruments to facilitate Western missionaries that operated in the region.

A noteworthy example is Phan Duc Hoa's *Thi Luat Thong Ngon* (1886), a distinctive handwritten dictionary spanning 649 pages. It includes a French preface, a table of 214 radicals, the main text, stroke checklists, and Latin letter checklists. Intriguingly linked to Samuel Wells Williams's *A Syllabic Dictionary of the Chinese Language* (1874), the *Thi Luat Thong Ngon*, organized by 214 radicals, cross-references the page numbers of the *Poetic Rhyme Book* for each included word. The dictionary incorporates Chinese Vietnamese, Pekingese, Cantonese, and French Pekingese sounds.

The pioneering bilingual dictionary of Sino-Vietnamese characters in southern Vietnam during this period was Paulus Huinh Tinh Cua's *Dai Nam Quoc Am Tu Vi* (1895-1896). This two-volume work arranged phonetically in

Latin alphabetical order, employs a word-governing-words approach, listing relevant vocabulary under each word. The Chinese characters are annotated with 'A' for Annanmite (indicating a Nom character) and 'C' for Chinois (indicating a Chinese character). The dictionary draws from classical Vietnamese literature and contemporary spoken language, spanning ages to encompass prevalent and archaic vocabulary. Notably, *Dai Nam Quoc Am Tu Vi* includes numerous vulgar characters, reflecting common scripts in Vietnamese society and providing valuable material for modern vulgar character studies.

Huynh Thuc Tram's *New Dictionary of Chinese and Vietnamese* (1951), spanning 1446 pages, distinguishes itself from other dictionaries as a compilation of common Chinese terminology from the 1840s and 1850s. This dictionary encompasses traditional Chinese words as well as introduces numerous new terms, including some of Vietnamese origin. Covering a spectrum of fields such as science, technology, economy, finance, industry, agriculture, medicine, and the names of domestic and foreign institutions, it reflects a comprehensive linguistic landscape.

Nguyen Van Khon's *Chinese-Vietnamese Dictionary* (1960), consisting of 1,156 pages with 10,000 single words and 40,000 words under its jurisdiction, follows the layout of Dao Duy Anh's *Chinese-Vietnamese Dictionary* but incorporates additional terms. Some content is drawn from Dao Duy Anh's work, but Nguyen Van Khon introduces several new terms. Chu Van Can's *New Chinese and Vietnamese Dictionary*, spanning 2229 pages, maintains a similar style to Ly Van Hung's *New Chinese and Vietnamese Dictionary* but boasts a larger word count.

Bao Ke's *Tam Nguyen Tu Dien* (1968) draws inspiration from Ly Van Hoe's earlier work to delve into the origins of Chinese words. Alongside explaining the meanings, the dictionary incorporates relevant poems and allusions from ancient texts to support word interpretations. While the Chinese characters corresponding to the words of Chinese origin are not explicitly indicated, the dictionary provides a phonetic checklist for scholars. Moreover, it offers aliases for over 70 Sino-Vietnamese sounds, presenting a distinctive feature.

Lac Thien's *Hoa Viet Thong Dung Tu Dien* (1972) compiles 6,000 general characters. The arrangement follows the 132 radicals in Jiang Zhongqiong and Lu Yiyian's *The Little Dictionary of the Improved Radical Standard Guoyin* (1936), utilizing phonetic symbols for pronunciation and indicating Sino-Vietnamese sounds. The unique interpretation section uses Roman numerals to sort words with multiple meanings. Additionally, the author specifies the word's associated part for convenient reference and retrieval.

Nguyen Quoc Hung's *New Chinese-Vietnamese Dictionary* (1975), spanning 872 pages, employs a dual format, initially organized in Latin alphabetical phonetic order, further segmented between homophones by radical. It exhibits a close association with the *Kangxi Dictionary*, relying on it for word selection and meanings. This dictionary showcases numerous archaic and variant characters, emphasizing their original meanings, often supported by relevant examples from classical texts. With an extensive collection of Sino-Chinese words paired with Vietnamese meanings, it elucidates these terms within the context of Nom literature.

During this epoch, the Southern Vietnamese population saw a surge in Overseas Chinese residents. With the purpose for better economic and cultural communication with the local populace, numerous bilingual dictionaries were crafted to aid Vietnamese Chinese in learning and referencing the Vietnamese language. A prominent example is Pho Can Tham's *Vietnamese Chinese New Dictionary* (1955), a 428-page compendium drawing from a diverse array of Vietnamese dictionaries and literary works. Leveraging strengths from various sources, it incorporates a plethora of words prevalent in Vietnamese society, meticulously organized in Latin alphabetical order.

In this era, bilingual dictionaries of Sino-Vietnamese characters are characterized by:

1. The lexicon and its content exhibit extraordinary richness, and encompass not only reference dictionaries for the study of Chinese disciplines but also pragmatic dictionaries applicable to societal communication.

2. Diverging from the earlier phase of dictionary compilation that leaned towards capturing the original and archaic meanings of words, this stage emphasizes a nuanced equilibrium between the ancient and the contemporary. It incorporates both classical terms and words commonly employed in the societal milieu of the time.

3. The target audience has gradually broadened. While traditional Sino-Vietnamese bilingual dictionaries and those created by Western missionaries were often designed for the indigenous Vietnamese population in previous stages, dictionaries in this era also cater to Overseas Chinese. The objective is to assist them in acquiring proficiency in Vietnamese and enhancing their communication skills.

3.2.3 Contemporary Bilingual dictionaries of Sino-Vietnamese characters (approx. 1975-)

Since the complete reunification of Vietnam in 1975, there has been a remarkable surge in the compilation of dictionaries. Over the past five decades, there has been a proliferation of Sino-Vietnamese bilingual dictionaries, showcasing a diverse array of styles and content.

In this period, bilingual dictionaries of Sino-Vietnamese characters are characterized by the following.

1. The primary objective of these dictionaries has shifted from learning and promoting Sinology to acquiring proficiency in modern Chinese.

2. Unlike earlier dictionaries that drew upon ancient texts such as the *Dai Nam Quoc Am Tu Vi*, *Sino-Vietnamese Dictionaries*, and *Kangxi Dictionary*, contemporary dictionaries now reference modern Chinese dictionaries like the *Modern Chinese Dictionary* 现代汉语词典 and *Xinhua Dictionary* 新华字典.

3. The level of innovation in the dictionary compilation process has diminished, with compilers relying more on modern Chinese dictionaries, resulting in less creativity in interpreting word meanings.

4. These dictionaries exhibit clear distinctions compared to Vietnamese and Nom dictionaries.

From 1975 to the present, the proliferation of Sino-Vietnamese bilingual dictionaries has led to a diverse array of categories with detailed content. However, the originality in dictionary compilation has faced limitations, as a significant number of dictionaries are essentially direct translations of Chinese dictionaries. This can be attributed to three plausible reasons. Firstly, the passing of the older generation of Confucian scholars, coupled with the emergence of a new generation educated in a different system, less proficient in Chinese characters and Chinese studies. This has necessitated the compilation of Sino-Vietnamese bilingual dictionaries that are more suitable for Vietnamese readership through direct translations of Chinese dictionaries. Secondly, a substantial shift in the purpose of compilation. In contrast to the preceding stages, which focused on the study of Sinology and the transmission of culture, the current stage witnesses a transformation. The widespread use of the Vietnamese national script has rendered Chinese characters and the Sinology they embody subjects of study for a select group of specialist scholars. Thirdly, The diminishing demand for Chinese characters due to intermarriage with locals, a decline in the number of Vietnamese Chinese, and the emergence of a new generation unfamiliar with Chinese and Chinese characters. To cater to the specific needs of this evolving readership, direct

translation using modern Chinese dictionaries has become the most efficient approach in practice.

4   Bilingual dictionaries of Sino-Vietnamese characters: Their significance through time

It is unfortunate that Sino-Vietnamese bilingual dictionaries from the Middle Ages to the Modern Ages have not received adequate attention from the academic community and have remained in a 'frozen' state for many years. There needs to be more references to these resources in subsequent dictionaries. The exact count of Sino-Vietnamese bilingual dictionaries has yet to be officially conducted, and only sporadic studies and monographs on specific topics have emerged. As a result, the true value of these resources has yet to be fully recognized, and the present generation has not fully harnessed their potential. In this context, I aim to systematically summarize the dual value of Middle Ages Chinese character research, encompassing both documentary contributions and the dissemination of Chinese characters.

The academic community has progressively deepened its comprehension and exploration of modern Chinese characters, designating them as the Clerical stage of Chinese characters. This period, ranging from the early Han dynasty to the May Fourth Movement, emphasizes the examination of modern Chinese characters in terms of their form, structure, usage, and origin. The focus of modern Chinese character studies lies particularly on clerical variants and vulgar characters (Jing, 2013, p. 309).

Bilingual dictionaries of Sino-Vietnamese characters from the Middle Ages to the Modern Ages are situated within this historical context. Spanning from the *Chi Nam Ngoc Am Giai Nghia* (17th century) as the earliest example to the *Tam Thien Tu Giai Dich Quoc Ngu* (1915) as the latest, these dictionaries reflect the evolution from ancient Confucian scholars to Western missionaries. The transition is observable from the exclusive use of Chinese characters to the coexistence of Chinese, French, and Vietnamese characters. Even before 1945, this evolution is notable.

As authoritative, standardized, and exemplary resources, the Sino-Vietnamese bilingual dictionaries from the Middle Ages to the Modern Ages hold immense value for the exploration of Middle Ages Chinese characters. These dictionaries, serving as the most official and comprehensive materials, significantly contribute not only to the understanding of Middle Ages Chinese characters but also to the broader study of extraterritorial Chinese characters.

4.1 Glyphology

Vietnamese Chinese characters exhibit similarities with their Chinese counterparts, featuring not only the predominant use of traditional characters but also the presence of numerous national characters. These vulgar characters find application in normative and authoritative dictionaries. Although most Vietnamese Chinese dictionaries were compiled by folk Confucians, with few official compilations, it is evident that Vietnamese vulgar characters were extensively utilized by the populace and even in more formal settings. The preserved forms within these dictionaries thus offer insights into the usage of Chinese characters in Vietnam. The examination of forms in Vietnamese Chinese dictionaries also presents valuable material for the study of Vietnamese typography.

Contribution to the Compilation of Vietnamese Chinese Character Variation Patterns. In the process of transmission and printing, the constituent elements of glyphs are often transcribed into similar or identical components, contrasting with the distinctive features of Vietnamese vulgar glyphs. For instance:

食 is written as リ: 餡→ 𩜍 ;餅→𩜍 饒→ 𩞉 .

虫 is written as 黾: 蛇→ 𧓌 .

足 is written as 𤴔: 露→ 𩂠 ;路→ 𨂾 .

龍 is written as 竜: 瀧→ 滝 .

Several Vietnamese components are also present, such as 疑 is written as 亇, 凝 is written as 𣵲, 義 is written as 𦰣, 哥 is written as 𠵴 羅 is written as 𦋺 學 is written as 𡨸 .

Additional Resources on Uncollected Glyphs in the Dictionary. During the collation of ancient Vietnamese dictionaries, it has come to light that certain characters cannot be located in Chinese dictionaries. Therefore, the compilation of this resource aims to supplement materials that have yet to be included in existing dictionaries, for example 𡵸𨼌𡑝𥓀𣓀𡲵 .

4.2 Phonetics

Sino-Vietnamese bilingual dictionaries from the Middle Ages to the Modern Ages were crafted by Vietnamese scholars using various approaches. These dictionaries were composed either entirely in pure Chinese characters, interspersed with Nom characters, or as a combination of Chinese, Nom, and French characters within a Vietnamese context. They were compiled with the primary aim of facilitating the study of the ancient Chinese language and classical literature and serving as a supplementary tool for imperial examinations. Representing a carrier of traditional knowledge transmission, Sino-Vietnamese bilingual annotate individual words with phonetic readings to enhance clarity. These dictionaries predominantly rely on the phonetic systems of *Fanqie* 反切 and *Zhiyin* 直音, offering materials rich in linguistic value.

Since the Song dynasty, rhyming books have been introduced to Vietnam through various means, leading Vietnamese Confucians to inherit much of the traditional Chinese phonetic notation for their dictionaries. However, they have made slight modifications to align with local language integration. To incorporate the nuances of the local language, the phonetic transcriptions of both *Fanqie* and *Zhiyin* are typically standardized based on Sino-Vietnamese phonology. Sino-Vietnamese phonology represents a phonetic system closer to that of the Chinese language, influenced by the Chinese language during the Middle Ages. An illustrative difference between the two can be seen in the following examples:

(1) Differences on Fanqie:

Vietnamese ancient dictionaries	Chinese ancient dictionaries
嘬，楚尺切 *Nam Phuong Danh Vat Bi Khao*	嘬，楚夬切 *Guangyun* 广韵
捼，古禾切 *Nam Phuong Danh Vat Bi Khao*	捼，奴禾切 *Jiyun* 集韵
搞，近交切 *Nam Phuong Danh Vat Bi Khao*	搞，丘交切 *Jiyun* 集韵

(2) Change the script in Zhiyin:

Vietnamese ancient dictionaries	Chinese ancient dictionaries
睍，音睨 *Tu Duc Thanh Che Tu Hoc Giai Nghia Ca*	睍，音哯 *Leihai* 类海
罙，音茉 *Tu Duc Thanh Che Tu Hoc Giai Nghia Ca*	罙，浮 *Jiyun* 集韵

(3) Additions:

Vietnamese ancient dictionaries	Chinese ancient dictionaries
畧，洽又畱 *Tu Duc Thanh Che Tu Hoc Giai Nghia Ca*	畧，音畱 *Yupian* 玉篇

(4)　Change the sound:

Vietnamese ancient dictionaries	Chinese ancient dictionaries
颺，音足 *Tu Duc Thanh Che Tu Hoc Giai Nghia Ca*	颺，音蕭 *Jiyun* 集韵
胱，止了切 *Tu Duc Thanh Che Tu Hoc Giai Nghia Ca*	胱，土了切 *Jiyun* 集韵

Sino-Vietnamese bilingual dictionaries from the Middle Ages rely on homophonic substitution according to Vietnamese phonology. Prosody can be viewed as a synthesis of homophonic substitution and supplementary pronunciation.

Sino-Vietnamese pronunciation is further categorized into two readings. The textual pronunciation 文读 adheres to the evolution of medieval sounds and is characterized by regularity and standardization. Conversely, the vulgar pronunciation 白读 represents the Chinese pronunciation transmitted to Vietnam before the Middle Ages or the pronunciation that has undergone Vietnameseization. The coexistence of textual and vulgar pronunciations, persisting to the present day, results in instances where a single word may be read multiple times. Some of these sounds have fused into Vietnamese words, rendering them challenging to discern.

For instance, the word 尻 is documented in *Nam Phuong Danh Vat Bi Khao* as 丘切 but in *Guangyun* as 苦刀切. The ancient sound of 尻, belonging to the *Youbu* 幽部 category, transformed during the Middle Ages, dividing into *Haoyun* 豪韻 and *Youyun* 尤韻. Initially represented by the same ancient sound, [ɯɯ], it later evolved into [ɐu] and [ɨu]. These two distinct sounds are preserved in the Sino-Vietnamese phonetic system.

Vietnamese influence on pronunciation is evident in instances such as the word 綯, marked in *Nam Phuong Danh Vat Bi Khao* as 交切, but in *Guangyun* as 徒刀切. In Vietnamese dictionaries and rhyming books, the character 刀 is often substituted with 交 in *Fanqie*, reflecting the prevalent pronunciation [zau55] in the northern Vietnamese dialect. Due to the vulgar pronunciation of [ɹau55], where all [ɹ] is pronounced as [z], these characters are interchangeably used. Similarly, the pronunciation of the word 韜, marked as 師交切 in *Nam Phuong Danh Vat Bi Khao*, aligns with the interpretation of the word 刀 as 交 in the *Tu Duc Thanh Che Tu Hoc Giai Nghia Ca*. Consequently, Vietnamese Chinese dictionaries frequently incorporate the most commonly used sounds of the time, offering invaluable insights into the phonological study of Sino-Vietnamese characters during the interregnum.

It is noteworthy that Sino-Vietnamese bilingual dictionaries from the Middle Ages employ an incomplete *Fanqie* system, omitting one of the phonetic components. While this practice is justifiable, occasional omissions in

engraving and printing occur. For instance, the pronunciation of 嚼 and 痞 in *Dai Nam Quoc Ngu* is marked as 雀切 and 洱切, respectively. Similarly, the pronunciation of 鋪 and 橧 in *Nam Phuong Danh Vat Bi Khao* is marked as 洽切 and 登切, and the pronunciation of 頑 and 汞 in *Tu Hoc Tu Ngon Thi* is marked as 還切 and 孔切. This phonetic notation method warrants further exploration in future studies.

4.3   Glossary and exegesis

Sino-Vietnamese bilingual dictionaries from the Middle Ages exhibit distinctive features in terms of glyphs, phonetics, and extensive Chinese glossaries. These glossaries, encompassing Vietnamese words and their meanings, trace their origins to the Song, Yuan, and Ming dynasties, notably popularized through vulgar glossaries like *Sui jin xi* 碎金系 'Shattered Gold Words' and *Za zi xi* 杂字系 'Miscellaneous Words'. Dictionaries compiled during the French colonial period by Western missionaries as well as those created by Vietnamese scholars offer insights into the evolution of the usage, selection, and meaning of Vietnamese Chinese words. These resources play a pivotal role in Vietnamese Chinese lexicography and exegesis, shedding light on a field that currently lacks scholarly attention.

Contemporary scholars have conducted minimal research on the exegesis of Sino-Vietnamese bilingual dictionaries. This article aims to draw attention to this neglected area through its introduction and collation.

Dictionaries serve as the foundation of exegesis, delving into the interpretation of word meanings. Beyond this, they analyze the structure of vague words and sentences, offering material for subsequent grammatical studies. The interpretative process encompasses the analysis of sentence readings and grammatical explanations, contributing to the study of rhetoric and modes of expression. Sino-Vietnamese bilingual dictionaries primarily use their language, employing Nom characters or Chinese characters. Some dictionaries provide explanations in Chinese or French for educational purposes. The variety of interpretations includes general interpretations, Chinese-Vietnamese equivalents (including Nom characters and Chinese characters), homophonic interpretations, and comprehensive interpretations. Traditional dictionaries often integrate these interpretations with the Vietnamese folk rhyming system of alternating six and eight rhymes. Notable examples include *Tu Duc Thanh Che Tu Hoc Giai Nghia Ca*, *Chi Nam Ngoc Am Giai Nghia*, *Tam Tu Kinh Giai Am*, *Ngu Thien Tu Giai Dich Quoc Ngu*, *Thien Tu Van Giai Dich Quoc Ngu*, *Tam Thien Tu Giai Dich Quoc Ngu*, etc. Additionally, some dictionaries adopt non-poetic forms to explain word meanings, such as *Dai*

Nam Quoc Ngu, *Nam Phuong Danh Vat Bi Khao*, *Nan Tu Giai Am*, *Nhat Dung Thuong Dam*, etc. Modern dictionaries, compiled by Western missionaries and Vietnamese scholars, adhere to contemporary formats and standards.

Sino-Vietnamese bilingual dictionaries primarily draw upon extensive traditional Chinese character references and rhyme books, including significant works like *Shuowen Jiezi* 说文解字, *Erya* 尔雅, *Zhengzitong* 正字通, *Kangxi Dictionary* 康熙字典, *Tang Yun* 唐韵, *Ji Yun* 集韵, *Hongwu Zhengyun* 洪武正韵, among others. Notably, *Tu Duc Thanh Che Tu Hoc Giai Nghia Ca* illustrates various citation methods, categorized into seven types:

1. *Original text type citation:* Direct quotations from sources.

(5)	Vietnamese ancient dictionaries	Chinese ancient dictionaries
	薦，鹿藋也； 岑，山小而高 *Shuowen Jiezi* 颭，風聲；䨮，電光 *Yupian*	薦，鹿藋也； 岑，山小而高 *Shuowen Jiezi* 颭，風聲；䨮，電光 *Yupian*

2. *Abbreviated citation:* Shortened references for brevity and clarity.

(6)	Vietnamese ancient dictionaries	Chinese ancient dictionaries
	砮，堅也 晷，照也	砮,餘堅也 *Shuowen Jiezi* 晷,日照 *Yupian*

3. *Change the character citation:* Instances where characters are altered or modified from the source.

(7)	Vietnamese ancient dictionaries	Chinese ancient dictionaries
	昶，日永也 霶，辰雨	昶，日長也 *Shuowen Jiezi* 霶，時雨 *Jiyun*

4. *Mixed citation:* A combination of different citation methods for comprehensive coverage.

(8)	Vietnamese ancient dictionaries	Chinese ancient dictionaries
	鴕，鳥似雉蛇化為雉也	鴕鳥似雉 *Jiyun* ＋ 化為雉也 *Jinshu:Zhuanghuazhuan* 晋书·张华传

5. *Rewritten citation:* Reinterpretation or paraphrasing of sources.

(9)	Vietnamese ancient dictionaries	Chinese ancient dictionaries
	櫌，鋤也，布秧後以此摩之，使土合覆也	櫌，摩田器，布種後以此器摩之，使土開發處復合覆種也 *Shuowen Jiezi*

* The term 布種 in *Shuowen* was rewritten as 布秧, and the words 器, 開發復合 and were omitted.

6. *Note type citation:* Explanations provided in the form of notes, offering additional insights.

(10)	Vietnamese ancient dictionaries	Chinese ancient dictionaries
	騮，赤馬黑鬣獵馬頸毛也赤馬黑毛尾也	赤馬黑毛尾也 *Shuowen Jiezi* 騮，赤馬黑鬣尾也 *Noted by Yan Shigu* 颜师古注
	朏，月三日明生之名	朏明，月三日明生之名 *Noted Shu: Zhaogao by Kong Yingda* 书-召诰

7. *Other types of citation:* Instances where Chinese characters are explained without specifying a clear source.

(11)	Vietnamese ancient dictionaries
	橛，門兩旁挾門短限也 虹，白而赤謂之虹

These distinctions serve as a valuable resource for research and argumentation within the field of exegesis. They facilitate comparative analyses between the meanings of specific words and those found in traditional Chinese literature. Furthermore, these differences enable a comprehensive and scientific exploration of the genuine meanings of words, contributing to a deeper understanding of their evolution over time.

4.4   Documentation and communication

Sino-Vietnamese bilingual dictionaries extensively refer to ancient Chinese texts in their interpretations, providing unique insights into historical figures and narratives when viewed from the Vietnamese perspective. While many interpretations align with corresponding Chinese records, analyzing the

discrepancies allows us to comprehend the diffusion and transformation of ancient Chinese thought and culture in Vietnam.

As mentioned earlier, the connection between Vietnam and Chinese characters spans over 2,000 years, forming an inseparable bond. Chinese characters have held a significant position for more than a millennium, serving not only as a universal medium for linguistic expression but also as a carrier of profound cultural messages. Within the Middle Kingdom, it encapsulates Chinese civilization, and beyond its borders, it becomes a disseminating seed influenced by various factors. Genetic mutations necessary for adaptation to diverse geographical conditions have resulted in variances between Chinese characters within and outside the Middle Kingdom, including Vietnam. Beyond slight alterations in shape, pronunciation, and meaning, the enduring feature is their ability to convey cultural messages, showcasing the beauty of Chinese character transmission.

Sino-Vietnamese bilingual dictionaries frequently incorporate and reproduce a broad spectrum of cultural, historical, and geographical information related to the Middle Kingdom. This practice aims to enhance the clarity of Chinese word meanings and serves as a mechanism for spreading Chinese characters throughout the region. Examining the dissemination of Chinese characters in Vietnam through these dictionaries also sheds light on cultural characteristics and the evolution of cultural interactions across generations. This exploration contributes to bridging gaps in local Chinese literature, attesting to the profound influence of Chinese characters and culture on the construction of civilizations within the Chinese cultural sphere.

5   Conclusion

This article provides a comprehensive overview of the present state of research on Sino-Vietnamese bilingual dictionaries from the Middle Ages and the Modern Ages, categorizing the discussion into three main themes: an overview of current research, historical staging, and the intrinsic value of dictionary research. Additionally, it delves into the historical context of the compilation of Sino-Vietnamese bilingual dictionaries during the Middle Ages and the Modern Ages, aligning these compilations with significant events in Vietnamese history. The exploration proves instrumental for an in-depth understanding of the contemporaneity of Sino-Vietnamese bilingual dictionaries in both the Middle Ages and the Modern Ages within the realm of Chinese characters. While acknowledging certain limitations in this article, future efforts will involve an extensive review of additional literature to address any gaps and enrich the scholarly discourse.

Acknowledgments

References

Ha, D. V. (2014). *A study of the nineteenth-century Sino-Chinese dictionary* [Unpublished doctoral dissertation]. Ha Noi: Han Nom Institute.

Hoang, T.T.T (2019). Characteristics of Modern Sino-Vietnamese Bilingual Dictionaries in Vietnam. *Journal of Zhishan*, 26, 65-79.

La, M. H. (2017). Vietnamese Middle Ages Dictionary from Lexicography Theory. In H. Z. Hua & N. T. Cuong (Eds.), *A Study of East Asian Chinese and Vietnamese Chinese Ancient Dictionaries* (pp. 60–73). Beijing: China Social Science Press.

Le, A. T. (2002). Bilingual Books - Special Types of Chinese and Nom Heritage. *Han Nom Magazine, 5*, 25–30.

Liang, M. (2017). Ancient Chinese dictionaries in Vietnam. In H. Z. Hua & N. T. Cuong (Eds.), *A Study of East Asian Chinese and Vietnamese Chinese Ancient Dictionaries* (pp. 40–51). Beijing: China Social Science Press.

Phan, V. C. (1994). A Bilingual Sino-Vietnamese Dictionary Compiled by Vietnamese Scholars. *Han Nom Magazine, 1*.

Jing, S. X. (2013). A Review of Middle Ages Chinese Character Studies in the Past Two Decades. *Journal of Chinese Language History*, 1, 308–318.

Tran, T. D. (2017). Types and Characteristics of Medieval Sino-Vietnamese Dictionaries. In H. Z. Hua & N. T. Cuong (Eds.), *A Study of East Asian Chinese and Vietnamese Chinese Ancient Dictionaries* (pp. 14–39). Beijing: China Social Science Press.

Trinh, K. M. (2011). The Situation of Vietnamese Chinese Language Tools and Related Issues. *Ethnic Culture, 38*.

Tran, V. C. (2018). *A Chinese script Toolkit for China and Vietnam*. Ho Chi Minh: Ho Chi Minh City General Press.

Wen, M. (2017). The value of Sino-Vietnamese bilingual dictionaries. In H. Z. Hua & N. T. Cuong (Eds.), *A Study of East Asian Chinese and Vietnamese Chinese Ancient Dictionaries* (pp. 117–124). Beijing: China Social Science Press.

Appendix: List of dictionaries

Date	(Title in characters (Hán-Nôm)) Title in Vietnamese alphabet (Quốc Ngữ) Title in simplified Latin script	Author(s)
3rd cent. BC	尔雅 Erya	
100 AD	说文解字 Shuowen Jiezi	许慎 Xu Shen
1627	正字通 Zhengzitong	张自烈 Zhang Zilie
543	玉篇 Yupian	顾野王 Gu Yewang
618–907	唐韵 Tang Yun	
960–1279	集韵 Ji Yun	
1368–1644	洪武正韻 Hongwu Zhengyun	
16th century	安南译语 An Nam dịch ngữ An Nam dich ngu	会同四译馆 Huitong Siyiguan
17th century	指南玉音解义 Chỉ nam Ngọc Âm Giải Nghĩa Chi Nam Ngoc Am Giai Nghia	
1651	Dictionarium annamiticum lusitanum et latinum	Alexandre de Rhodes
1660	万法指南 Vạn pháp chỉ nam Van phap chi nam	释了一 Shi Liaoyi
1772	Dictionarium Anamitico Latinum	Jean-Louis Taberd
1827	日用常谈 Nhất dũng thượng đảm Nhat dung thuong dam	范廷琥 Phan Đình Hổ Phan Dinh Ho
1827	增订幼学须知杂字采珍大全 Tăng đính ấu học tu tri tạp tự thái trân đại toàn Tang dinh au hoc tu tri tap tu thai chan dai toan	范廷琥 Phan Đình Hổ Phan Dinh Ho

1831	三千字解音 Tam thiên tự giải âm Tam thien tu giai am	吴时任 Ngô Thời Nhiệm Ngo Thoi Nhiem
1838	Dictionarium Anamitico- Latinum	Jean-Louis Taberd
1839	钦定辑韵摘要 Khâm định tập vận trích yếu Kham dinh tap van trich yeu	范文谊等 Phạm Văn Nghi et al Pham Van Nghi et al
1867	Dictionnaire étymologique Chinois-Annamite Latin-Français	G. Pauthier
1868	Dictionnaire élémentaire annamite-français	Legrand de la Liraÿe
1877	Dictionarium Anamitico-Latinum	Jean-Louis Taberd
1877	Dictionnaire annamite-français	Louis Caspar
1879	字学训蒙 Tự học huấn mông Tu hoc huan mong	黎直 Lê Trực Le Truc
1880	字学求精歌 Tự học cầu tinh ca Tu hoc cau tinh ca	杜辉琬 Đỗ Huy Uyển Do Huy Uyen
1882	字学四言诗 Tự học tứ ngôn thi Tu hoc tu ngon thi	黎直 Lê Trực Le Truc
1884	Petit dictionnaire français-annamite	Trường-Vĩnh-Ký Truong Vinh Ky
1886	Index des caractères Chinois contenus dans le dictionnaire Chinois Anglais de Williams avec la pronunciation Mandarine Annamite	Phan Đức Hoá Phan Duc Hoa
1887	Vocabulaire franco-tonkinois	Gaston Kahn
1887	Vocabulaire annamite-français	Trường-Vĩnh-Ký Truong Vinh Ky
1889	Dictionnaire Annamite-Français	Charles Bailly
1890	千字文解音 Thiên tự văn giải âm Thien tu van giai am	
1893	Vocabulaire annamite-français	M.Génibrel
1895	检字 Kiểm tự Kiem tu	杨嘉训 Dương Gia Huấn Duong Gia Huan
1895	Petit vocabulaire Chinois-Annamite-Français	Edmond Nordemann

1895–1896	大南国音字汇 Đại Nam quốc âm tự vị Dai Nam Quoc am tu vi	Huình Tịnh Của Huinh Tinh Cua
1897	嗣德圣制字学解义歌 Tự Đức Thánh chế tự học giải nghĩa ca Tu Duc Thanh che tu hoc giai nghia ca	嗣德 Tự Đức Tu Duc
Unknown	古愚正误四言诗 Cổ Ngu chính ngộ tứ ngôn thi Co Ngu chinh ngo tu ngon thi	黎直 Lê Trực Le Truc
Unknown	指南备类 Chỉ Nam bị loại Chi Nam bi loai	
Unknown	村居便览 Thôn cư tiện lam Thon cu tien lam	
Unknown	难字解音 Nan tự giải âm Nan tu giai am	
Unknown	字类演义 Tự loại diễn nghĩa Tu loai dien nghia	
1898	Dictionnaire Annamite-Français	M. Génibrel
1899	大南国语 Đại Nam quốc ngữ Dai Nam quoc ngu	阮文珊 Nguyễn Văn San Nguyen Van San
1899–1900	Dictionnaire Annamite - Français	Jean Bonet
1900	Nouveau vocabulaire français-tonkinois et tonkinois-français	P.Crépin
1902	南方名物备考 Nam phương danh vật bị khảo Nam phuong danh vat bi khao	邓春榜 Đặng Xuân Bảng Dang Xuan Bang
1906	Petit lexique de poche français-annamite	M.Génibrel
1908	三千字解译国语 Tam thiên tự giải dịch quốc ngữ Tam thien tu giai dich quoc ngu	
1909	五千字解译国语 Ngũ thiên tự giải dịch quốc ngữ Ngu thien tu giai dich quoc ngu	阮秉 Nguyễn Bính Nguyen Binh

1909	千字文解译国语 Thiên tự văn giải dịch quốc ngữ Thien tu van giai dch quoc ngu	
1909	Technique du peuple Annamite	Henri Oger
1914	常谈日用汉字烈歌 Thường đàm nhựt dụng Hán tự liệt ca Thuong dam nhut dung Han tu liet ca	胡玉谨 Hồ Ngọc Cẩn Ho Ngoc Can
1931	Từ điển Việt Nam Vietnamese Dictionary	Pham Quynh, Nguyen Van Vinh, Tran Cong Kim, Duong Bac Khao (Eds.)
1932	漢越辭典 Hán - Việt Từ Điển Chinese-Vietnamese Dictionary The Chinese-Vietnamese Dictionary in Brief	Đào Duy Anh Dao Duy Anh
1936	The Little Dictionary of the Improved Radical Standard Guoyin	Jiang Zhongqiong, Lu Yiyian
1941	Tam Nguyên Tự Điển Tam Nguyen Tu Dien	Lê Văn Hòe Le Van Hoe
1942	Từ Điển Trung-Việt Chinese-Vietnamese Dictionary	Thiều Chửu Thieu Chuu
1942	Từ Điển Nam Hoa Nam Hoa Dictionary	Nguyễn Trân Mô Nguyen Tran Mo
1942	Chinese-Vietnamese Dictionary	Thieu Chuu
1951	Từ Điển Mới Trung-Việt New Dictionary of Chinese and Vietnamese	Huỳnh Thúc Trâm Huynh Thuc Tram
1955	Từ Điển Hoa Việt Mới Vietnamese Chinese New Dictionary	Phó Cận Thẩm Pho Can Tham
1957	Từ Điển Trung-Việt Chinese-Vietnamese Dictionary	Vạn Tân Van Tan
1960	Từ Điển Trung-Việt Chinese-Vietnamese Dictionary	Nguyễn Văn Khôn Nguyen Van Khon
1960	Từ Điển Mới Trung-Việt New Chinese-Vietnamese Dictionary	Lý Văn Hùng Ly Van Hung
1968	Tam Nguyên Tự Điển Tam Nguyen Tu Dien	Bảo Kế Bao Ke
1972	Từ Điển Hoa Việt Thông Dụng Hoa Viet Thong Dung Tu Dien	Lạc Thiên Lac Thien

1975	Từ Điển Mới Trung-Việt New Chinese-Vietnamese Dictionary	Nguyễn Quốc Hưng Nguyen Quoc Hung
1994	Từ Điển Trung-Việt Phan Văn Các Chinese-Vietnamese Dictionary	Phan Văn Các Phan Van Cac