LINGUISTIC INNOVATIONS IN CHINESE:

INTERNAL AND EXTERNAL FACTORS

by

XINJIA PENG

A DISSERTATION

Presented to the Department of East Asian Languages and Literatures
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2017

DISSERTATION APPROVAL PAGE

Student: Xinjia Peng

Title: Linguistic Innovations in Chinese: Internal and External Factors

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of East Asian Languages and Literatures by:

Zhuo Jing-Schmidt	Chair
Lucien Brown	Core Member
Kaori Idemaru	Core Member
Scott DeLancey	Institutional Representative

and

Scott L. Pratt	Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded June 2017.

DISSERTATION ABSTRACT

Xinjia Peng

Doctor of Philosophy

Department of East Asian Languages and Literatures

June 2017

Title: Linguistic Innovations in Chinese: Internal and External Factors

This dissertation seeks to deepen understanding towards language change by answering three questions: What is the unit of change? What is the manner of change? What are the factors of change? Three cases of linguistic innovation in the Chinese language are examined. Adopting a usage-based approach, I analyze the language data of these three linguistic innovations, and the results provide unanimous answers to the three questions. First, the basic unit of language change is a construction, and it can be of any length, such as phrasal, clausal or discourse-length. Second, these cases of linguistic innovation present a scenario of change led by high-frequency exemplars, demonstrating that language change can be abrupt rather than gradual. Third, the external factors giving rise to the exemplars prove crucial in reconstructing language change in progress. All three case studies present linguistic innovation as a response to a changing material reality. I thus advocate a usage-based constructionist approach that considers external factors in the investigation of language change, as it allows us to develop a more comprehensive understanding of the process.

CURRICULUM VITAE

NAME OF AUTHOR: Xinjia Peng

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
Beijing Foreign Studies University, Beijing

DEGREES AWARDED:

Doctor of Philosophy, Chinese Linguistics, 2017, University of Oregon
Master of Arts, Language Teacher Specialization, 2013, University of Oregon
Master of Arts, East Asian Linguistics and Language Pedagogy, 2013, University of Oregon
Bachelor of Arts, English, 2009, Beijing Foreign Studies University

AREAS OF SPECIAL INTEREST:

Chinese Linguistics
Teaching Chinese as a Foreign Language

PROFESSIONAL EXPERIENCE:

Graduate Teaching Fellow, University of Oregon, Eugene, Oregon, 2011-2014; 2015-2016

GRANTS, AWARDS, AND HONORS:

College of Arts and Sciences Dissertation Research Fellowship, Univerity of Oregon, 2016-2017

Graduate Teaching Fellow, EALL, Univerity of Oregon, 2011-2014; 2015-2016

CLTA Peking University Press Publication Award, Chinese Language Teacher Association, 2015

Best Presentation Award, Chinese Language Teacher Association-California Spring Conference, 2015

CAPS Small Professional Grant, Center for Asian and Pacific Studies, Univerity of Oregon, 2013; 2014; 2015

Siegel Graduate Travel Endowment Fund Scholarship, EALL, Univerity of Oregon, 2012; 2014

Second-class Scholarship, Beijing Foreign Studies University, 2008

First-class Scholarship, Beijing Foreign Studies University, 2006

PUBLICATIONS:

Peng, X. (2017). Book review: *Teaching Chinese as a Second Language: Curriculum Design & Instruction*, by Xiaohong Wen. *Chinese as a Second Language (漢語教學研究—美國中文教師學會學報)*. *In production.*

Jing-Schmidt Z. and Peng, X. (2017). Winds and tigers: Metaphor choice in China's anti-corruption discourse. *Lingua Sinica,* 3:2. DOI: 10.1186/s40655-016-0017-9

Jing-Schmidt, Z. and Peng, X. (2016). The emergence of disjunction: A history of constructionalization in Chinese. *Cognitive Linguistics*, 27(1), 101-136.

Peng, X. (2016). Book review: *A Handbook of Tasks and Rubrics for Teaching Mandarin Chinese*, by Miao-fen Tseng. *Chinese as a Second Language (漢語教學研究—美國中文教師學會學報)*, 51(2), 218-224.

Jing-Schmidt, Z, Peng X. and Chen, J. (2015). From Corpus Analysis to Grammar Instruction: Toward a usage-based constructionist approach to constructional stratification. *Journal of Chinese Language Teacher Association*, 50(2), 1-30.

ACKNOWLEDGMENTS

I have been fortunate enough to have many *guìrén* as companions on my odyssey towards the doctoral degree, all of whom have made the arduous journey picturesque and joyous.

I am most indebted to my academic advisor and mentor Zhuo Jing-Schmidt, who gives me continuous inspirations and encouragement during my time in the program. In the countless meetings we have in her office, online and elsewhere, new ideas are always being spontaneously generated, my frustration and procrastination are unnoticeably curbed. As many of my other previous and continuous projects, the completion of this manuscript has benefited tremendously from her guidance, suggestions, timely feedbacks, and especially her patience to work with me through multiple revisions. And such generous guidance in my academic research and career points to a future I am still about to fathom and witness.

My gratitude also goes to all my committee members. All of them gave me insightful suggestions to improve the manuscript. Specifically, I would like to acknowledge Lucien Brown for his encouragement to rethink the pragmatic function of *ye-shi-zui-le* in Chapter IV, Kaori Idemaru's suggestion to streamline the presentation of research methods, and Scott DeLancey who offers an areal and historical perspective to all three case studies examined in this dissertation.

I would like to thank audiences at the 19th WEAL conference and the EALL Colloquium at the University of Oregon where part of the manuscript was presented. Colette Grinevald has also read an early draft of Chapter II and given me my critique. Among the many fellow camaraderie graduate students who never fail me with their

uncompromising encouragement and ingenious ideas, I would like to specifically thank Matt Stave, who introduced programming to me. Our collaboration on another project has greatly improved my programing skills, without which data collection and data analysis in this dissertation would be much more strenuous.

The completion of this manuscript would not have been possible without Holly Lakey. She carefully and thoroughly edited my manuscript under time pressure, and gave me many sharp-witted critiques as a fellow linguist, which prepares me for future publications of this manuscript.

The last year of my work was supported by the College of Arts and Sciences Dissertation Research Fellowship awarded by the University of Oregon. It has allowed me to concentrate on finishing this manuscript at the final stage.

Last but not the least, I would like to thank my dear friends on the West Coast and across the Pacific Ocean that keep me accompany, especially Amy, Grace, Hsin-chin, Ningning, Teresa, Weijun and Yiqing, I would like to thank my parents for supporting me in pursuing an academic career, and my husband Zhen, who is my peace haven, my fortress, my reliable source of strength and inspirations.

TABLE OF CONTENTS

Chapter Page

Chapter Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

CHAPTER I
INTRODUCTION

1.1. Background of the Current Study

The subject of this dissertation is linguistic innovations, with special emphasis on the internal and external factors that motivate language change. By internal factors, I refer to what have traditionally been considered "internal" to the language system: the system of linguistic signs and the cognitive capacity of its users. The term has also presupposed the opposite of "internal," which are factors that are "external" to the linguistic system—namely, the social factors of language use. It is the latter consideration that this dissertation will strongly advocate, and this is in line with the philosophy of at least certain research areas in linguistics, especially sociolinguistic research and studies on language contact.

The consideration of external factors was exemplified most remarkably in the work of variationist sociolinguist Labov.[1] In his influential study of Martha's Vineyard (1963), he made two important arguments: First, there are social meanings assigned to linguistic signs; and second, these linguistic signs imbued with social meanings can be utilized under the pressure of changing social circumstances. The abundance of sociolinguistic research in the 60s and 70s, especially what later came to be termed as the *variationist* approach, greatly advanced the integration of social considerations in linguistic research and dispelled the "sociological agnosticism in structural linguistics," as Weinreich, Labov and Herzog (1968, p. 177) put it. These authors, departing from the increasingly influential social approach to linguistic research, established seven principles as the "empirical foundations" of language change, with the last one stating:

Linguistic and social factors are closely interrelated in the development of language change. Explanations which are confined to one or the other aspect, no matter

[1] In Labov's (1963) study of Martha's Vineyard, he observed an increase of centralized diphthong among younger speakers compared to elder ones, by which he predicted that there is change in progress in the language (the apparent-time construct). He proposed that there was a social meaning associated with the variant, and by using the centralized diphthong the younger generation were proclaiming an islander identity which was under threat at the time.

how well constructed, will fail to account for the rich body of regularities that can be observed in empirical studies of language behavior. (p. 188)

The field has seen subsequent works that took up the suggestion, though not in prolific terms. Among them were the brief discussion in Aitchinson (3rd edition 2001, pp. 133–200; first edition in 1981), the work of Thomason and Kaufman (1988) on language contact and creoles, as well as the essay collection of Gerritsen and Stein (1992) which is dedicated entirely to exploring the internal and external factors of syntactic change. Emerging from these works is the realization that external factors can offer powerful explanations to language change. For instance, Mithun gives an example of how varying degrees of cultural conservatism as exhibited by various Iroquoian language communities (in North America), as well as the differential power dynamics between these communities and their English-speaking neighbors, have resulted in an uneven innovation of coordinating conjunctions within the language family (1992, pp. 89–130). Fujii presents the case of the standardization of modern Japanese, particularly the invention of compulsory marking on the grammatical subject, following the country's foreign language contact when Japan opened for trade (1992, pp. 257–294). Gerritsen observes the stabilization of SOV order for Dutch in conjunction with the technological advancements in printing (1992, pp. 355–394). Jing-Schmidt has also hypothesized that the emergence of the *ba* construction (which primarily has the function of discourse dramatization) in colloquial literature in Chinese since the Song Dynasty is driven by the literary development of story telling—when it became a popular form of entertainment, it called for linguistic forms that "create narrative tension and accuracy" (2005, p. 286). All these more recent studies show that as researchers, when considering social (or external) factors of language change, we can and should go beyond identity as the only type of social meaning associated with linguistic forms, and exhibition of an alternative social identity as the motivation behind linguistic variants and language change. The social contexts surrounding language use, the intention of language use, the need for expressing certain ideas, even the medium language use is carried on, can all relate to language change. I particularly agree with Thomason and Kaufman (1988), who state that external factors should not be sought "**only** [emphasis added] when all efforts to find an internal

motivation for some change have failed […] the possibility of multiple causation should be kept in mind [and] an explanation should be as complete as possible" (pp. 57-8). My investigation of the internal and external factors of language change in this dissertation does not spring from a shortage of internal explanations, but merely from a desire to be thorough. The discussion of these factors does not seek to be exhaustive, but I aspire to tell a story as multi-dimentional and multi-faceted as I can to recreate the process of language change.

But as Weinreich, Labov and Herzog have correctly pointed out, "[l]inguistic and social factors are closely interrelated in the development of language change" (1968, p. 177). As I intend to show throughout this dissertation, the so-called "external" factors are not in opposition to the internal factors and they cannot be treated as independent causes of language change. On the contrary, this study will show that change in linguistic signs can be associated with a different way to cognize the world, which can in turn be triggered by changing social and cultural realities. The specific linguistic expressions (which are internal to language) that demonstrate the changing abstract linguistic knowledge have their roots in society, in their use in social settings. The internal and external factors of language change can hardly be teased apart.

1.2. Construction Grammar as the Framework

For this investigation, three case studies of observed linguistic innovations will be examined, including a grammatical category, a formulaic phrase and a syntactic pattern. However, instead of proposing that these innovations develop in and by themselves, I consider them all to have respectively emerged in larger linguistic constructions which contain them. For instance, I argue the formulaic phrase *ye-shi-zui-le* which became viral on a Chinese social media platform did not emerge by itself but through its usage in a discourse construction, the YSZL construction. The larger construction that contains the phrase (or other types of linguistic innovations) not only provides a semantic anchor but also structural features that allow the linguistic innovation to acquire a stable form-meaning pairing. In other words, the current study thus considers the construction as the basic unit of language change (innovation). The entire dissertation will be situated in the

theoretical framework of Construction Grammar—a linguistic theory which recognizes the construction to be the basic unit of grammar. In this section, I will give a brief summary of key proposals of the theory, especially those of particular relevance to the current discussion on language change.

Construction Grammar has developed within it approaches that differ from each other, but there are several basic tenets that underlie most approaches, including (Goldberg, 2013):

> (a) a construction is a conventional pairing of form and meaning;
> (b) a construction's semantics is associated directly with its surface form;
> (c) knowledge of language is a system formed by a network of constructions, in which nodes are related by inheritance links.

An additional tenet not found in all approaches but exemplified in and central to Cognitive Construction Grammar (Goldberg, 1995, 2006) and Radical Construction Grammar (Croft & Cruse, 2004) is that language structure emerges from language use. This is recognized as the usage-based approach to language. These four tenets represent the primary assumptions within Construction Grammar that this study adopts. These assumptions form the basis to understanding how grammar emerges, how grammar is stored, and how grammar is structured.

Regarding the emergence of grammar, a usage-based approach to language dictates that the linguistic patterns of language (traditionally understood as "grammar") emerge as language users make generalizations over language input. Goldberg calls this the "surface generalization hypothesis," which posits that "surface forms have more powerful generalizations than 'deep structure'" (2006, p. 23). It implicates that the kind of language input users are exposed to can shape their generalizations towards a linguistic pattern. For instance, the lexical meaning of verbs can affect the constructional meaning of the constructions they occur in. Take the classic ditransitive construction [SUBJ V O_i O_d] as an example: Meaning subtypes such as causing the O_i to receive O_d (e.g. 'SUBJ promise O_i O_d'), enabling the O_i to receive O_d (e.g. 'SUBJ allow O_i O_d'), or even cause O_i not to receive O_d (e.g. 'SUBJ refuse O_i O_d') all surround the central sense of "transfer of an object," and this sense relates to the meaning of the prototypical verb *give* that

occurs with high frequency in this construction (Goldberg, 1995). Collocational analysis has further enabled researchers to support this point with quantitative evidence. This approach, exemplified in Stefanowitsch and Gries (2003), shows the lexical input of the verbs that tend to appear in the *into*-causative construction (e.g. 'to make one's way into') is why the construction conveys the semantic notion of "force" or "coercion." Words like *trick*, *coerce*, *force* are among those that show the strongest collocational strength with the construction. These results suggest that constructions are not neutral and unbiased; they possess idiosyncrasies, and there is an inclination for certain words to occur or not occur in certain constructional slots. On the one hand, the usage instances (or instantiations) of the construction establish and shape the generalizations of a linguistic construction; on the other hand, the emergence of a linguistic generalization allows a user to further instantiate the pattern, sometimes with a creative usage, which can either reinforce or reshape the construction. In this sense, neither grammar nor vocabulary (as they were defined traditionally) is static because they are constantly changing and evolving in an interactive manner.

In this process of pattern generalization from surface form, frequency plays an important role, and it is now agreed that token frequency and type frequency are two parameters that can influence the process. Token frequency represents how often a single usage instance is instantiated in a given construction. How often the lexeme *give* occurs in the verb slot in the ditransitive construction [SUBJ V O_i O_d] is an example of token frequency. High token frequency of an item strengthens its cognitive representation through repetition, and when it has cognitive salience, it provides stable features that have an anchoring effect, strengthening the memory of the linguistic construction (Bybee, 2006, 2013; Goldberg, 2006). Research on child language acquisition has shown that high token frequency of a single item can enhance generalization (Goldberg, 2006, p. 83), and that children can "leap to make generalizations on the basis of a single exemplar" (Tomasello, 2003, p. 173).

Type frequency counts how many different usage instances are instantiated with a given construction. In the context of the ditransitive construction, it would mean how many different verbs can be found in the verb slot in a given corpus. For instance, *give*,

pass, *throw* would count as a type frequency of three. It was proposed that the higher the type frequency, the more productive a construction can be (Goldberg, 1995, p. 134) as it establishes "more bases […] for the item-based analogy [to] create novel instances of the construction" (Bybee, 2013, p. 62).

The emphasis on token frequency and type frequency suggests that usage instances have a place in the mental representation of linguistic knowledge. Acquisition research by Goldberg (2006, p. 55) reveals that, contrary to the belief that our linguistic knowledge consists only of abstract generalizations of language use (and the lexicon), instances of usage are stored alongside abstractions over specific usage instances. Such findings thus suggest an exemplar model to be a more likely language learning model than the prototype model, because the former permits the retention of details in exemplar representation which shape the idiosyncrasy of the construction.

Several methodological implications stem from the usage-based approach to the emergence of grammar. First, because grammar comes from usage, an examination of usage instances is essential to understanding the kind of generalizations that emerge. In all my case studies, I use authentic language data from a corpus for the analysis. Second, usage instances with high token frequency can, in particular, drive towards generalizations prominently associated with these instances; token frequency will thus be an important parameter I consider in the analysis of high-frequency usage instances of a construction. Third, high type frequency of a construction is an indication of the productivity of that construction, but additionally, as previous studies have shown, there are usually semantic clusters among these types (see, for instance, Israel, 1996 for the semantic clusters of verbs in the *way* construction). In my analysis, I will look into the semantic clusters of the various subtypes of the construction to gain insights into the constructional meaning of the construction. Nosofsky (1988) found that categorization is greatly facilitated by both high-frequency exemplars and members that are similar to these exemplars. In other words, the combined effect of frequency and similarity between high-frequency instances will inform my analysis of the new linguistic construction.

Now, regarding the structure and storage of grammar, a usage-based approach to language holds the construction as the most basic unit of language, and they can be found

in different levels of analysis, from morphemes to clauses (Croft, 2001; Fillmore, 1989; Goldberg, 2003; Langacker, 1987). Constructions can also vary in their degree of concreteness from more substantive (having concrete forms) to more schematic (having abstract generalizations) (Traugott & Trousdale, 2013, p. 12). This study consists of an investigation of constructions at different levels: a phrasal construction (Chapter II), a clausal construction (which has both subject and predicate, see Chapter III) and a discourse construction which can span across several clauses as a single or several sentences (Chapter IV). The constructions examined here invariably have an abstract component—a schema or linguistic pattern which emerges as users make generalizations over many instances of usage. The linguistic constructions that come in different sizes and levels of schematicity together form a network of constructions, and this is what a constructionist approach perceives as the language system. In this network of constructions, constructions at different levels can be connected to each other through inheritance links. The following only gives a few examples of the most common types of links. For instance, the abstract pattern of a construction can be connected to specific usage instances (an inheritance link), but it can also be connected to another abstract pattern which is part of the bigger construction (a subpart link). Even though the different parts of a construction can be independent constructions themselves, the bigger construction is not compositional: the constructional meaning is larger than the sum of its parts. The non-compositional nature of a construction is one feature identified in Goldberg's conception of 'construction' (2006, p. 5), and is a position the current study adopts. Additionally, an abstract construction can be connected to different constructions with formal similarities but meaning differences (polesemy links); constructions that are related by metaphorical mapping are also connected in this network (metaphorical extension links). In the current investigation, the constructions in the three case studies, although characterized as "new" constructions, emerge in a network of constructions and have inheritance links with existing constructions that are familiar to language users. Their form-meaning mappings create new nodes in the network, and they do not emerge in isolation, but in connection with existing nodes in the network. The concept of a network of constructions in the current study requires attention be paid to related

constructions linked to the new construction through various inheritance links. As the case studies will show, analysis of these related constructions will prove crucial to the proper understanding of the new construction.

Overall, Construction Grammar with a usage-based approach is distinct from the traditional understanding of grammar as *a priori* knowledge. In this renewed view, linguistic knowledge is a system that constantly undergoes evolution and development. In this sense, it is an Emergent Grammar (Hopper, 1998). Because grammar emerges in usage, the source of the generalizations over linguistic patterns effectively shapes the mental representation and structure of linguistic knowledge, and the source of language use is human experience. One of the most fundamental claims made by Cognitive Construction Grammar is that the structure of language is organized around the categorization of basic embodied human experience (Langacker, 1987, 2008, 2012). Language users understand the meaning and function of linguistic forms through their experience interacting with others using language. Tomasello (2003) proposes that there is a psycholinguistic mechanism, an intention-reading skill, that helps children to understand the communicative intention of language users, and henceforth to understand language. Utterances are thus comprehended, analyzed and structured around their communicative intentions (pp. 296–7). Therefore, when talking about form-meaning mappings, linguistic forms can be mapped to not only semantic generalizations, but also to information structure (Goldberg, 2006, p. 9) and other concerns of broader usage contexts. The mental representation of knowledge is thus concerned not only with form, but also with *when* and *how* the form should be used.

At this point, I would like to briefly introduce Hyme's (1974) comprehensive model that investigates the "ethnographies of communication" as a conceptual orientation of what linguistic forms can be mapped to. According to Hyme, the contexts that are relevant to communication include all the elements captured by the acronym SPEAKING—S standing for setting and scene, P for participants, E for ends, A for act sequence, K for key, I for instrumentalities, N for norms, and G for genre. In presenting the three case studies in the following chapters, I strive to give an account of the various components of these constructions in the context of communication, including the

historical backdrop in which their emergence is situated, the users and audience of their instantiations, the intention and the result of the speech act, and the medium in which the communication is conducted, to paint a more comprehensive picture of the scope of emergence. The linguistic constructions will be presented not only as form-meaning pairs, but also a form-function pairs that fulfill communicative functions in a given social-historical setting.

1.3. Language Change from a Constructionist Perspective

Language change in a constructionist view refers to the shifting of form-meaning pairs. The concern of this study is linguistic innovations, or the emergence of new linguistic constructions. In the constructionist framework, it is conceived as the creation of a new node in a network of related constructions, or defined as "constructionalization" in the work of Traugott and Trousdale (2013). In their work *Constructionalization and Constructional Change*, which aims at re-conceptualizing form-meaning pair change from the perspective of Construction Grammar, they see constructionalization as "typically the outcome of language users making not one, but a succession of multiple, new links between constructs and constructions or schemas" (p. 92). For instance, usage instances in Middle English through present day English suggests that the constructionalization of the *way* construction is paved by many gradual changes on multiple paths. According to Traugott and Trousdale, the two subschemas of contemporary *way* construction (one with verbs used in manner of motion constructions, and other with verbs used in sound or activity accompanying motion constructions) develop from two sets of precursors (pp. 83-91). One of them concerns instances in Middle English using deictic motion (instransitive) verbs like *go*, in [$V_{\text{deictic motion}}$ POSS way] (DIR), which gives *way* a reanalysis as a "fake object" because of the following directive complement; the other is a transitive set that is found later in Early Modern English, with causative (transitive) verbs like *take* and *make* as the prototype. And following this path, by the end of the seventeenth century, a *way* construction with a transitive verb has emerged, in the form of [$SUBJ_i$ [V $POSS_i$ way] (DIR)]. Here, *way* is no longer an objective, but a fixed part of a causative construction; in this construction,

there are also verbs of "path-creation" such as *battle* and *drag* which form the basis for later constructional developments because of their implied manner of action. The *way* construction example shows that when constructionalization occurs, it is preceded and followed by gradual micro-changes of the construction, or "constructional changes." By their definition, constructionalization creates an entirely new form-meaning pair, $form_{new}$-$meaning_{new}$, which results from many constructional changes of the form-meaning pair in only one dimension ($form_{old}$-$meaning_{new}$ or $form_{new}$-$meaning_{old}$). The gradualness of the process and the gradience exhibited between stages of development had also been espoused in an earlier collection of essays edited by the same authors (Traugott & Trousdale, 2010). However, they have also remarked that the gradualness of change is more widely observed in the emergence of more procedural constructions, and for more contentful construction, the process can sometimes be instantaneous, though they are still largely gradual (2013, p.22).

In the case of English, the difference between constructionalization and constructional change is not always apparent, but it is still discernible. To these authors, constructionalization occurrs when "morphosyntactic changes appear in the textual record as well as semantic ones" (p.27). In the present study, the judging criteria based on morphological and syntactic marking does not apply because written Chinese does not have morphological markings. And when morphosyntactic changes are not visible in changes composed by a succession of micro-constructions, the only evidence is semantic change. In a diachronic study of the Chinese disjunction, we find that the differences in semantics can sometimes be difficult to discern if only examined locally, but when the larger pragmatics of a construction is considered, the ambiguity between stages of development disappears. It is the contextual inference at the pragmatic level that gives rise to reanalysis of the structure and change (Zhuo Jing-Schmidt & Peng, 2016). In the following case studies, the analysis of the usage contexts of the construction will also be emphasized.

Language change from a constructionist perspective is thus not in conflict with grammaticalization and lexicalization—if anything, language changes in constructionist terms under the definition of constructionalization and constructional change can capture

more types of linguistic changes that do not fit into the canonical definition of grammaticalization (a lexical item comes to serve grammatical functions) and lexicalization (creation of a contentful lexical item). The type of linguistic change as proposed in the formulation above (including form change and semantic change) can all be observed in processes of grammaticalization and lexicalization. Theories emerging from grammaticalization and lexicalization about language change will also inform research under the constructionist framework. The directions and features of change (Hopper, 1991; Lehmann & others, 1995), the mechanisms of change (such as reanalysis, analogy in (Hopper & Traugott, 2003)) and the role of contexts and pragmatic inferencing (Bybee, Perkins, & Pagliuca, 1994; Heine, Claudi, & Hünnemeyer, 1991; Himmelmann, 2004; Hopper & Traugott, 2003) are all theoretical proposals the current study will engage and adopt in different capacities.

1.4. Overview of the Dissertation

Departing from the theoretical assumption of Construction Grammar with a usage-based approach, I plan on engaging with existing observations about language change within this framework, and specifically, linguistic innovation in Chinese. I do so (1) by focusing on constructions as the units of change, (2) by examining the extent to which changes are gradual or abrupt, and (3) by investigating the factors that contribute to the change. In the case of this study, I propose that because linguistic change is factually a change in the mental representation of a linguistic pattern, it can only be identified by a changed lexical input and collocational pattern in the context of Chinese constructions. The degree of innovation in the input item thus influences how gradual or abrupt the process is. The linguistic innovation in the case studies exhibited in the current study tend to fall on the abrupt end.

As an overview, the following is a summary of the content in each chapter.

Chapter II works with a phrasal construction in Chinese. More accurately speaking, I propose that the new grammatical category “numeral classifier” emerges within the QUANTIFICATION construction [Noun Numeral Unit]. Numeral classifier is the quantification unit in a subconstruction of the QUANTIFICATION construction that

counts count noun by individual unit. I use text data between 16th and 8th century BCE because this is the time frame generally considered to predate the establishment of numeral classifiers as a grammatical category. By examining the instantiations of the construction in the language data and their frequency of occurrence, I argue that the emergence of the grammatical category starts from one specific instantiation—the counting of individual horses, or *'ma* (numeral) *pi*'. This is an innovative instantiation of the QUANTIFICAITON construction, and it is motivated by the need for accurate and efficient counting of horses. Additionally, the cultural significance of horses (reflected by its frequent occurrence in the corpus) propelled the generalization towards this instantiation to impact the linguistic system.

Chapter III presents the development of a clausal construction in the Internet language of modern Chinese. I propose that two subconstructions of the Chinese *bei* passive emerge from the Internet—particularly, the news genre online. The usage of the new *bei* construction [N_{person} "*bei* X"] to indicate that a condition has been imposed against/without the will of the experiencer extends from a series of usage events. Observers (possibly the experiencer) of the event denounces the misrepresentation of X imposed upon N_{person} by an institutionalized discourse that speaks through the news media. With quantitative evidence, I will show that high frequency instances constitute the exemplars that drive the generalization of such a new construction. The new construction acquires pragmatic functions from the social contexts of the usage instances and can only be explained in concrete social contexts.

Chapter IV also deals with a new construction emerging in Internet language, but specifically on a social media platform. The construction in question is a discourse construction that became widely known through the viral spread of memes first on this platform. I call this discourse construction the YSZL construction, and it is composed of both a schematic component that can be filled with user-generated content and a substantive component—the formulaic phrase *ye-shi-zui-le* which literally means 'I'm also drunk' but used metaphorically. Different from some assumption that *ye-shi-zui-le* is a formulaic chunk that functions by itself, I suppose that it emerges within the entire construction as users dynamically interpret the inter-relation of these two components,

and their meanings feed each other to facilitate comprehension of the entire construction. Overall, the constructional meaning conveys a sense of disorientation and a state of impaired judgment, and it has the pragmatic function of being an evasive device that exempts the user from opining on this public space. I believe the social context where the construction emerges offers an explanation for its emergence. Briefly put, the Internet users' experience in this new space where public and private life has been blended has induced an embodied experience analogical to what is induced by alcohol intoxication, and in such a blended space, they also feel the need to suppress their opinions.

Finally, in Chapter V, I will discuss the two questions I pose at the beginning of this section, namely how new constructions emerge and how this emergence is motivated, drawing on implications from these three case studies.

CHAPTER II
EMERGENCE OF A PHRASAL CONSTRUCTION: THE CASE OF THE EMERGENCE OF NUMERAL CLASSIFIERS

2.1. Introduction

This chapter investigates the emergence of numeral classifiers as a grammatical category in Chinese. 'Numeral classifier' is a full-fledged grammatical category in modern Chinese, the systematic development of which can be traced back as far as the second century BCE (Huang, 1961). In this chapter, I will go into text data predating the second century BCE, which are the earliest available written records of Chinese, to examine what are possibly the earliest instances of numeral classifiers. I propose that their emergence should be considered within the QUANTIFICATION construction, and that the emergence of numeral classifiers at the same time creates a subconstruction of the QUANTIFICATION construction, which expresses quantification through the numeral classifiers.

A special grammatical category that is prominent in languages of Asia and Oceania is the numeral classifier. According to a typological study by Aikhenvald (2000, p. 121), numeral classifiers are present in Tibeto-Burman languages, in Chinese languages, and in most Austroasiatic languages as well as Japanese, Korean and Ainu. They are used obligatorily in counting count nouns. The standard view on the function of numeral classifiers is that they "make count nouns enumerable by individualizing and classifying them" (Bisang, 1999, p. 113). The idea of "individualization" as a function of numeral classifiers can be traced back to Greenberg (1972) in his typological study of the grammatical category found in more than 100 languages. While agreeing that the so-called 'classifier languages' in general lack plural markings (Sanches, 1971), Greenberg proposes that the functional motivation for the rise of classifiers is in response to the need to distinguish collective nouns and singular nouns which are usually marked morphologically or lexically in many other languages, such as Russian and Classical Arabic. As isolating languages lack the means to mark singularity and plurality in morphological terms, and a stand-alone noun represents a collective concept in context,

these languages need a way to "individualize" noun referents to make them "enumerable." The close link between numeral classifier and enumeration can also be testified to by the syntactic integration of the numeral classifier and the numeral. In Greenberg's investigation, there is no case of a numeral classifier construction where the numeral classifier and the numeral are separated by the head noun. "Individualization" is thus to Greenberg the main function of numeral classifiers, and semantic classification is more of a by-product in this process.[2] However, as languages evolve, new functions have stemmed from this prototypical usage of classifiers. Some languages, such as Cantonese, Hakka, and some Min dialects of Chinese languages, have also developed the function of referencing and expressing possession in classifiers (Bisang, 1999).

Greenberg pointed out early on the connection between the numeral classifier and measure words. He hypothesized that "unit counters [a.k.a. numeral classifiers] are modeled after the construction of mass nouns which cannot stand directly with numerals but require a measure or quasi-unit-counter as an intermediary" (1972, p. 16). Here, he is referring to the use of measure words in quantification, such as *cup* in *a cup of water*. Not coincidentally, early descriptions of Chinese grammar also see the Chinese numeral classifier as being related to measure words. When describing classifiers from Archaic to Contemporary Chinese, Chinese grammarians consider numeral classifiers as falling under the category of measure word along with measure unit (度量衡單位), container unit (容量單位), and counter unit (集體單位) (Chou, 1959; Ding et al., 1953; Guan, 1953; Wang, 1958). One reason to put these under the same category is the fact that they exhibit similar syntactic behavior. As the examples below show, numeral classifiers and other measure words appear in a construction that has the form [NUMERAL UNIT NOUN]. For example:

[2] The usage of numeral classifiers exhibits some kind of semantic clustering, or conceptual categorization. In Chinese, for instance, the numeral classifier for cows is *tou* 'head'; it is also used with other large livestock. The numeral classifier for trees is *ke* (written with a tree radical); it can also be used for grass and sprouting. The numeral classifier for rivers is *tiao*, and it can be used with many other long and bending objects, including concrete objects such as scarfs, and also more abstract objects such as a piece of news. The sharing of numeral classifiers based on appearance feature, functional feature, and affective distance is common across languages that have numeral classifiers.

(1) *yi_sheng_shui*
one _liter_water
NUMERAL_MEASURE UNIT_NOUN
'a liter of water'

(2) *yi_bei_shui*
one_cup_water
NUMERAL_CONTAINER UNIT_NOUN
'a cup of water'

(3) *yi_ge_ren*
one_NC_person
NUMERAL_NUMERAL CLASSIFIER_NOUN
'a person'

(4) *yi_qun_ren*
one_group_person
NUMERAL_COUNTER UNIT_NOUN
'a group of people'

What these expressions share, besides their surface structure, is their function of counting nouns. Such form-function mapping thus nicely constitutes what Construction Grammar defines as "a construction." We may tentatively name this construction the "QUANTIFICATION construction" based on its primary function. It has the following form:

(5) QUANTIFICATION construction: [NUMERAL UNIT NOUN]

Depending on the semantics of the modified noun and the need for different levels of specification, different types of quantification units occur in the UNIT slot: Measure units are standardized unit for measurement; container units are non-standardized measure unit, used mostly with noncount nouns; numeral classifiers are used to count the individual item of count nouns; counter units count count nouns in a unit of multiples.

Exemplified previously is the current state of the Chinese QUANTIFICATION construction. Some scholars have suggested that numeral classifiers as a category might not have emerged when the other three types of quantification units were in use in a certain period of Archaic Chinese (Wang, 1958). The absence of numeral classifiers amidst the presence of other quantification units not only represents a diachronic slice of

the Chinese QUANTIFICATION construction, but also corresponds to the status quo of most languages outside of Asia and Oceania. This chapter investigates the motivations behind the peculiar emergence of numeral classifiers in Chinese.

2.2. The Establishment of Numeral Classifiers as a Grammatical Category

Previous scholarship has identified the significant growth of numeral classifiers as a grammatical category in Chinese during the second century BCE and the sixth century CE. In the first stage of development (during the second century BCE and the second century CE), the use of numeral classifier become more obligatory as more nouns began to use the numeral classifier when counting (Huang, 1961). Two forces are at work here. On the one hand, the emergence of a general numeral classifiers *mei* (later *ge*) creates a channel for more nouns to take on classifiers in the QUANTIFICATION construction; on the other hand, a number of more specialized numeral classifiers developed to accommodate these nouns (Chen, 2002; Xing, 2013). The increasing frequency of use, induced by both the increasing number of collocating nouns and an increasing number of numeral classifiers, reinforces the semantic features of numeral classifiers, which may include the fact that (i) classifiers do not denote any concrete object in the world, but represent an individualization of the counted object, and (ii) sometimes they exhibit classifying functions based on the physical attributes of the modified nouns. At this stage, the numeral classifiers are instantiated in an early form of the QUANTIFICATION construction in the order of [NOUN NUMERAL UNIT].

The later phase of development features the flourishing of classifiers between the second and the sixth century CE (Liu, 1965). The structural shift of the QUANTIFICATION construction from [NOUN [NUMERAL UNIT]] to [[NUMERAL UNIT] NOUN] changes the position of the expression of quantification [NUMERAL UNIT] relative to the NOUN. For instance, to say "three pieces of arrow" would change from (6) to (7), which gives the QUANTIFICATION construction its current form.

(6) *shi_san_mei*
arrow_three_NC:mei
NOUN_NUMERAL_NUMERAL CLASSIFIER
'three pieces of arrow'

(7) *san_ge_jiantou*
three_NC:ge_arrow
NUMERAL_NUMERAL CLASSIFIER_NOUN
'three pieces of arrow'

This structural shift is largely credited as the main trigger for the reanalysis of the quantification unit, especially numeral classifiers, as a grammatical category that performs a specific modifying function. Peyraube (1998), for instance, argues that when the structural shift took place, and the quantification phrase [NUMERAL UNIT] became prenominal, it tended to be interpreted as a modifier for the following noun (versus when [NUMERAL UNIT] is postnominal, it tends to be interpreted as the predicate of the preceding noun) (p. 103). In this process, the structural feature of the numeral classifier, along with its high type frequency, came to be analyzed as a word category that modifies the noun in counting.

The reanalysis induced by the structural shift has to be understood in the way grammar and the lexicon interact in the context of Archaic Chinese. Bisang notes that particularly in Archaic Chinese, the language exhibits "precategoriality," which means that a word in written Chinese has relative freedom to enter different syntactic/constructional slots without being constrained by its typical usage pattern. For instance, a prototypical noun *you* 'friend,' can be used as a verb to mean 'to be friends with,' as in example (8):

(8) *you_qi_shi_zhi_ren_zhe* (《論語・衛靈公第十五》)
friend_DEM_gentleman_ASSOC_kind_NOM
'Befriend people who are humane.'

Here, *you* is coerced (in the sense of constructional coercion, see Michaelis, 2004) into being a verb by both the local clause and the larger discourse context, and exemplifies what Croft proposes about the fluidity of word categories: "There are no grammatical categories independent of constructions"(2001, p. 170). In other words, the "precategoriality" of Chinese at this stage does not imply the complete lack of word categories in the language system. Rather, it elucidates the disassociation between a specific lexical item and a designated word category. Word categories do exist, and they

exist within constructions. Therefore, it is essential that the examination of the emergence of numeral classifiers be situated in the unit of a construction, especially when examining data in Archaic Chinese.

In the previous discussion, the two stages of development of the numeral classifier thus point to its establishment as a grammatical category prior to the second century BCE.. A theory about the emergence of the numeral classifier can be found in Huang (1964) (also An, 2009; Dai & Jiang, 2005; Li, 2000 inter alias). He suggests that the emergence and development of the category numeral classifier follows a path as (9) shows. The different stages of development indicate the development is diachronic:

(9)
a. Stage I [NOUN NUMERAL]
b. Stage II [NUMERAL NOUN]
c. Stage III [NOUN NUMERAL NOUN] (or the "repeater construction")
d. Stage IV [NOUN NUMERAL NUMERAL_CLASSIFIER]
e. Stage V [NUMERAL NUMERAL_CLASSIFIER NOUN]

In this proposal, numeral classifiers first emerged in a numeral classifier construction (Stage IV) before it became a full-fledged category (Stage V). Leading to Stage IV is a repeater construction (Stage III) evolving out of quantification expressions in Stage I and II, which count nouns by combining the noun and the numeral.

I found this proposal problematic in several respects. First, as the data presented here show, the forms in 9a through 9c (even possibly 9d) were already present in the earliest written records of Chinese in oracle-bone inscriptions; without access to earlier data, it is impossible to argue whether, prior to oracle-bone inscriptions, their earliest occurrences follow the suggested order chronologically. Second, the developments through these stages is reminiscent of a transformational approach. But even using a transformational approach, it does not specify which one is the "deep structure" and lacks a motivation of change. Third, as the following data will reveal, both the type frequency and the token frequency of the repeater construction is low across texts predating the second century CE, and I doubt they could have had such an effect that led to the emergence of Stage IV. Finally, the proposition of numeral classifier construction in the

form of [NOUN NUMERAL NUMERAL_CLASSIFIER] suffers from the lack of a semantic definition of NUMERAL_CLASSIFIER, so to say that the numeral classifier emerged from this construction is circular.

In the following sections, I seek to revise the proposal about the emergence of the numeral classifier in Archaic Chinese by systematically examining actual usage data. Specifically, its emergence will be situated within the QUANTIFICATION construction in the form of [NOUN NUMERAL UNIT], following Greenburg and others' proposal (section 2.1) about numeral classifiers' affinity with other types of quantification units.

2.3. Data and Methods

Drawing on insights from usage-based Construction Grammar, in particular how surface forms have the most powerful generalizations and the role of frequency in forming generalizations (see section 1.2), I will examine the usage instances of numeral classifiers and their frequency of occurence. But, as I have pointed out earlier, the precategoriality of Archaic Chinese dictates that the numeral classifier is a semantic and grammatical category in relation to the noun, and thus it should be examined within a construction where it modifies the noun. To avoid the circularity observed in Huang's approach, I decided that the search of the numeral classifiers should start from examining the usage instances of the QUANTIFICATION construction [NOUN NUMERAL UNIT] because the construction has clear structural features, and it prevents inclusion and elimination based on subjective semantic judgment. The goal is to locate high-frequency and exemplary usage instances of the QUANTIFICATION construction in the data, as measured by occurrence frequency.

2.3.1. Data

The data used for the current analysis are from the oracle-bone inscriptions and bronze inscriptions, which are the texts used from around the sixteenth to the second century BCE. Oracle-bone inscriptions, which compose the earliest written records of Chinese, correspond to the Shang dynasty (16^{th}-11^{th} century BCE) historically. They are considered to predate bronze inscriptions.

Oracle-bone inscriptions are inscribed texts on oracle bones. They are used in rituals of fortune-telling, where numbers are inscribed on bones which priests would then burn to see the cracks on the surface as a way to tell good from evil. The numbers found on the bones are the indicators of the sequences of the rituals. Narratives are inscribed after the rituals to include the details of the rituals and the results of the fortune-telling. Topics of fortune-telling include weather, harvest, hunting, illness, childbearing, war, and other rituals. Bronze inscriptions are found on bronze vessels and date back as far as the Shang dynasty, but the majority have been linked to Western Zhou (11th century - 771 BCE) and later periods. In the Bronze Age of ancient China, ritual bronzes were produced for an individual to use in ritual offerings to his ancestors. The inscriptions on these vessels were usually to document the purpose of rituals around important social events such as edicts, war, hunting, treaties, and other events happening in the court that were worth commemoration. In this study, I will use oracle-bone inscriptions from the Shang Dynasty and bronze inscriptions from Western Zhou as the material to investigate the emergence of numeral classifiers from the sixteenth to the eighth century BCE.[3]

My primary sources of data are two collections of work. For oracle-bone inscriptions, I used *Jiaguwen Heji*, or the *Oracle-bone Inscriptions Collection*, which was published between 1978 to 1982 by the Institute of Archeology of the Chinese Academy of Social Sciences. It is a thirteen-volume compilation of 41,956 pieces of oracle bone excavated by the published date. It contains both rubbings of the bones and transliterations of the oracle-bone inscriptions. The archaisms of the text pose great difficulty for researchers. The grammar and lexicon in oracle-bone inscriptions are vastly different from Archaic Chinese, and to make it worse, characters are sometimes lost in these artifacts, which hinders the comprehension of the text as a coherent whole. The fragmented texts have made it quite a challenging task for scholars to understand each and every one of the characters of the oracle-bone inscriptions. Although they had made good progress with deciphering the texts and were able to give transliterations of the

[3] Since the contents and orthography of the inscription can determine the periodization of the texts relatively accurately, I was able to tease apart data from different periods and use only those from Western Zhou based on periodization identified by previous scholars.

characters to the best of their knowledge, the data that I worked with are texts sometimes filled by lacunae to indicate missing or unknown characters. The *Oracle-bone Inscriptions Collection* has been made available online, searchable by text and includes scanned images of the rubbings alongside the transliteration of the texts.[4] The data I present in the following examples use numbering from this collection.

For bronze inscriptions, I used *Yinzhou Jinwen Jicheng Shiwen*, or the *Annotated Collection of Bronze Inscriptions from Yin(Shang) and Zhou*, which is an annotated version of the 18-volume *Collection of Bronze Inscriptions from Yin(Shang) and Zhou* published during the '80s and '90s by the same institute. This collection also includes 12,113 pieces of rubbings of the scripts and transliteration of the inscriptions. Compared to oracle-bone inscriptions, bronze inscriptions in Western Zhou have more similar orthography and grammar to Archaic Chinese, which makes the transliterations more comprehensible. However, they also have the same problem of missing characters. An online version has also been made available with searchable text and scanned images of the rubbings.[5] The data I present in the following examples use numbering from this collection. Additionally, I also referenced an *Index* of the collection (Zhang, 2001) to quickly locate and identify scripts containing certain words.

2.3.2. Methods

The first step of data collection is an initial exploratory analysis to identify usage instances of the QUANTIFICATION construction in these two bodies of texts, guided by the example instances identified in previous studies, including Chen (1956), Da (2004),[6] Guan (1953), Huang (1964), Peyraube (1998), Wang (1958) and Yang-Drocourt (2004). My search was narrowed down to a handful that are relevant to the current discussion of the numeral classifier, especially those that occur in high frequency. Their frequency of

[4] *Oracle-bone Inscriptions Collection* online: http://www.guoxuedashi.com/jgwhj/

[5] *Collection of Bronze Inscriptions from Yin(Shang) and Zhou online:* http://www.guoxuedashi.com/yzjwjc/

[6] Da (2004) is a thesis that collects and presents the quantification expressions in Archaic Chinese of different periods. Regardless of theoretical contribution, the thesis is exhaustive in presenting the different words and structures used to quantify in Archaic Chinese

occurrence (in the form of [NOUN NUMERAL UNIT]) is counted by performing a keyword search on the online corpus. The occurrence frequency is a piece of important information (based on the assumption that productivity is an indication of the level of grammaticalization). that was largely missing in this line of work, as most studies focused only on presenting the variety of quantification units I believe my supplement of this information has allowed me to make observations that was not available before.

From the same prior studies (as listed in the last paragraph), I knew that there were different ways to express quantification (see (9)). I thus conducted another round of keyword searches based on the NOUN in these instances and counted the total number of occurrences where these nouns occur in any kind of quantification expressions to find out how often they are counted using the QUANTIFICATION construction.

The non-obligatory use of the QUANTIFICATION construction in counting, as observed in oracle-bone inscriptions and bronze inscriptions, also leads to another question concerning the pattern of distribution. Drocourt (1993) noted that in Archaic Chinese, the difference of the other two forms of counting ([NUMERAL NOUN] and [NOUN NUMERAL]) is mainly in their exposition of information structure, with the new information in [NOUN NUMERAL] being the numeral, and the new information in [NUMERAL NOUN] being the noun (she did not comment on the information structure of the QUANTIFICAITON construction though). Droucourt also found different distributional patterns in different genres (record-keeping as in the bronze inscriptions, narratives as in the *Book of Songs*). Inspired by such findings, I thus posit that the context of the text is also a variable that could affect the choice of quantification expressions. In the exploratory analysis, I noticed that various kinds of quantification expressions can be found most prominently in texts about a) hunting trips and sacrificial ceremonies in oracle-bone inscriptions, and b) bestowal and war booties (in bronze inscriptions). Keywords indicating these themes are thus identified and used to annotate the context of occurrence. The texts annotated for their themes, and the findings based on these annotated texts, will be discussed in greater details separately in section 2.4 and section 2.5.

A note on the presentation of the inscriptions: Because the phonological information of these inscriptions is not available, and their lineage with modern Chinese is through their orthography, I will use the modern Chinese characters they have been translated into to represent the data. For important instances, pinyin will be given instead for the convenience of reference. Also, because there was no punctuation in these inscriptions, the punctuation presented was added by compilers of the data.

2.4. Findings in Oracle-bone Inscriptions

2.4.1. Contexts of Cccurrence

All three types of quantification expressions, [NUMERAL NOUN], [NOUN NUMERAL], and [NOUN NUMERAL UNIT], can be found in the oracle-bone inscriptions, and they occur predominantly in documenting two types of content, namely hunting trips and sacrificial ceremonies. This observation is based on random searching of nouns in quantification phrases as documented in previous scholarship, as well as a search of numbers (one through ten and a hundred).

In the dataset, relevance to the theme “hunting trips” is indicated by the character 隻, which is an archaic form of 獲, meaning ‘to get, to obtain, to acquire objects’ as a result of human labor. In this content category, quantification is used to indicate the animals acquired during a hunting trip. For instance:

(10) 乙未_卜[：]今日_王_□_光[，]□[。]
(date)_to divine[:]today_king_?_(location?)[,]?[.]

允_隻_馬_二[、]兕_一[、]鹿_二十一[、]豕_二[、]
indeed_get_horse_two[,] female rhino_one[,] deer_twenty one[,] pig_two[,]

麋_百_二十七[、]虎_二、兔_二十三、
river deer_hundred_twenty seven[,] tiger_two[,] rabbit_twenty three[,]

雉_二十七[。]十一_月[。] [10197]
pheasant_twenty seven[.] eleven(th)_month[.]

In this example, a bountiful hunting trip was documented, evidenced by the inventory list of a variety of prey and their number. While the length of the list is not typical in this content category, this is a representative example, as the long list of hunting prey shows that the quantification of the animal prey is consistently in the order of [NOUN NUMERAL].

"Sacrificial ceremony" is the other content category that contains huge numbers of quantification instances. This category is usually indicated by means of live sacrifices, such as 卯 'to kill (animals) with a knife,' 伐 'to behead,' 宜 'to display meat on the table,' and 血 'to drain blood out of an animal and use the blood.' For instance:

(11) …大丁[、]大甲[、]且乙_百_鬯[、]百_羌[，]
(name)[,] (name)[,] (name)_hundred_wine(,) hundred_*qiang* people

卯_三百_□ [301]
kill_three hundred_(?oxen)

This example documents a ceremony in remembrance of several ancestors. Wine, live persons, and three hundred animals (possibly oxen; the character was not identifiable in the inscription) are used for the ceremony. Here is another example of quantification in the context of sacrificial ceremonies:

(12) 癸卯_宜_于_□_羌_三_人[，]
(date)_to display meat_LOC_(location)_*qiang* people_three_person[,]

卯_十_牛[。] [6068]
kill_ten_oxen[.]

This example documents a different sacrificial ceremony, less extravagant in scale, that used ten oxen and three live persons.

The two examples generally reflect the types of quantification expressions used in this context. That is, unlike the hunting trip context shown above, this context does not use a single construction consistently when quantification is called for. However, it seems that which construction to use is somewhat dictated by the nouns (much more so than by syntax). From what is observed, the counting of live persons is in the

QUANTIFICATION construction, and the counting of the oxen is in [NUMERAL NOUN].

For sacrificial objects and animals, a majority use the order [NUMERAL NOUN], although a handful of [NOUN NUMERAL] instantiations can be found; for instance, a rare case of the latter order is shown in (13).

(13) …血_三_羊[，]□_伐_二十[，]
…to drain blood_three_goat[,] ?_behead_twenty[,]

鬯_三十[，]牢_三十[,] … [22231]
wine_thirty[,] cow_thirty[,]…

I conducted a corpus search querying common objects used in sacrificial ceremonies, including animals (e.g., 牢 which refers to cows and goats raised specifically for sacrificial purpose, as well as regular cows 牛 and goats 羊 not raised for the purpose), wine 鬯 (wine that is specifically used in ceremonies). This tendency to use [NUMERAL NOUN] rather than [NOUN NUMERAL] in this usage context generally holds true.

Combining the context of occurrence and the information structure of the pattern, the distributional pattern of the different means of quantification is thus this: In the context of documenting hunting acquisition, where the form [NOUN NUMERAL] is favored, the emphasis is on the number of each type of prey—just like an inventory list. In the context of documenting sacrificial ceremonies, which more often uses the form [NUMERAL NOUN], the emphasis is on the type of objects used, which signifies the grandeur of the ceremony—although quantity also matters, it is more likely to be prescribed by existing rules.

However, the fact that these two forms do not strongly correlate with usage contexts and that they are used somewhat interchangeably suggests that the different expressions of quantification are not fully grammaticalized. This is further evidenced by the occurrence of a third form [NOUN NUMERAL UNIT], which, in the context of sacrificial ceremony, is predominantly observed in the quantification of *qiang* people, as

example (12) above shows.[7] I have counted the occurrence of different ways to count *qiang* people in this context, and as the results in Table 1 below indicate, its occurrence by context fails to exhibit any predictable patterns.

Table 1. Distribution of the quantification of *qiang* people across texts in different constructions

	qiang + (numeral)	(numeral) + *qiang*	*qiang* + (numeral) + *ren*
Sacrifice	33	75	27
Hunting	2	0	0
Others (unknown)	22	53	10

Note: This table shows the distribution of observable quantification expressions of *qiang* people—observable in the sense that the transcriptions are legible enough to render texts both before and following the quantification expressions.

From the data provided by the inscriptions, I believe the use of quantification expressions is not sanctioned by grammar but is driven by the pragmatic need to present different aspects of the information, depending on the usage context. For this reason, there exists an explanation for the different distributional pattern between [NOUN NUMERAL] and [NUMERAL NOUN]. But with this assumption, what might be the pragmatic reasons driving the usage of the form [NOUN NUMERAL UNIT], for instance, "*qiang* + (numeral) + *ren*," in the context of sacrificial ceremonies?

2.4.2. The QUANTIFICATION Construction

As defined previously in this chapter, the form [NOUN NUMERAL UNIT] instantiates the QUANTIFICATION construction. Its status as a construction is clearly informed by its instantiations in Medieval Chinese and Modern Chinese. However, in oracle-bone inscriptions, its instantiations are quite limited, found only in a handful of instances. Table 2 shows the instantiations of the QUANTIFICATION constructions in

[7] *Qiang* was a major group of non-*han* people that is mentioned frequently in the oracle-bone inscriptions. According to texts in the oracle-bone inscriptions, it is thought that they resided in an area neighboring the territory of the *han* people, and they were constantly at war with each other (Ma, 1984; Wang, 2008). In times of confrontation, the *qiang* people could fall captive to the *han* people and they were kept alive as slaves or used in human sacrifices; this is what they were for in this example and the next example. Corresponding to such a status of the *qiang* people, two orthographies of *qiang* can be found in oracle-bone inscriptions, one similar to the orthography in modern Chinese characters, and the other with an additional radical representing chains and shackles, indicating that they are slaves.

the oracle-bone inscriptions, including the token frequency of the instance as listed in the first column, and the total frequency of occurrence of the same noun that occurs in counting in the entire corpus of oracle-bone inscriptions.[8]

Table 2. Instantiation frequency of the QUANTIFICATION construction in oracle-bone inscriptions

Instance	Meaning	Token Frequency[c]	Frequency of quantification of the noun
qiang (numeral) *ren* 羌(numeral)人 "*qiang* (numeral) person"	Counts the number of *qiang* people by *ren* "people"	27	223
chang (numeral) *you* 鬯 (numeral)卣 "ceremonial wine (numeral) container"	Counts the number of ceremonial wine by container/vessel	21	41
ma (numeral) *bing* 马(numeral)丙 "horse (numeral) one horse-drawn chariot unit"	Counts the number of horses by the number of chariot units they draw	4*	9
bei (numeral) *peng* 贝(numeral)朋 "cowrie shells (numeral) ten cowrie shells"[a]	Counts the number of cowrie shells by a unit of *peng*, a collective of ten cowrie shells	2**	2
ju (numeral) *bing* 車(numeral)丙 "chariot (numeral) one horse-drawn chariot unit"	Counts the number of chariots by the number of chariot units	1*	2
chang (numeral) *sheng* 鬯 (numeral)升 "wine (numeral) liter(?)"[b]	Counts the number of wine by volume (liter)	1	41

[a]Cowrie shells were the primitive money in use starting from the Shang dynasty (Wang 1980, 64). It is reasonable that they were also the main form of endowment from the emperors.
[b]Whether the character *sheng* was in fact a measure unit equivalent to 'liter' is uncertain.
[c]Asterisks are used to indicate levels of obligatory usage.

First, the use of quantification units in quantification is not obligatory. This is contrary to the impression created by studies listing only one example for each instance. The column "token frequency" shows the total number of occurrences of this instance in

[8] Disputable instances were left out, such as 丰 and 珏.

the entire *Collection of Oracle-bone Inscriptions*. When compared against the number in the next column, which indicates how many times the same noun is used in any quantification expressions, this shows how often the QUANTIFICATION construction is used in quantification for the given noun. Among these five quantification instances, only the quantification for cowrie shells (used as currency) obligatorily uses the QUANTIFICATION construction (indicated by two asterisks). The quantification unit in this construction is a counter unit, which counts cowrie shells in a unit of ten. Next in degree of obligatoriness to use quantification units seems to be the counting of horses and horse-drawn chariots in a chariot unit (indicated by one asterisk), both of which occur witih very low frequency. With regard to ceremonial wine measured by container, I can only say the use of the construction is only somewhat obligatory, given the overall high frequency of occurrence. The least obligatory is the counting of *qiang* people by individual. The counting of wine by volume *sheng* can possibly be ignored because it is a *hapax legomenon* which does not constitute sufficient evidence of it as a measure unit.

The different degrees of obligatoriness of this construction, I believe, can help construct the function of this construction. The focus should be placed on the obligatoriness of the quantification units. A basic assumption of Construction Grammar, inherited from the functional approach, is that different surface forms are connected to different meanings and functions. In this view, the instances instantiating the QUANTIFICATION construction thus have different meanings than their counterparts that appear in the form [NOUN NUMERAL] and [NUMERAL NOUN]. Apparently, the difference is in the presence and absence of the quantification units. The obligatoriness of the QUANTIFICATION construction is eventually the obligatoriness of the quantification units.

Now let us look further at cases where the quantification units are obligatory, versus those that are less so. It seems that the focus is not on whether something can be counted off one by one, but rather on marking its functional features. The cowrie shells were compulsorily counted in a unit of ten, even though the counting in ten naturally means that they can be counted individually. This was not done, possibly because, in an imperial bestowal, the value of cowrie shells is only in units of ten, and the individual

counting of cowrie shells would cause them to lose their ritualistic value. The use of a collective quantification unit with cowrie shells is thus to realize their function as a bestowal. Similarly, for horses and chariots, the functional unit is in the assembly of both horses and chariots; not counting them in a functional unit of *bing* might indicate that the horses or the chariots are not discussed in the context of a functional horse-drawn chariot. That the function of the quantification unit is in marking the object's social function, and not in making objects countable, is further supported by the counting of ceremonial wine. It is probably difficult for us to imagine how liquid can be counted without a container or volume (except, for example, "a beer," where the noncount noun comes in a predictable container or size), but half of the instances of counting ceremonial wine were done without a quantification unit to make them countable. The use of a quantification unit marking social function is optional in this case because the social function has been embedded in the noun itself. 鬯 is a word that refers specifically to wine for ceremonial purposes, and wine not for ceremonial but personal consumption might have a different word. From this, I would like to propose that the function of the quantification unit was not quantification; instead, the quantification unit serves to add metadata onto the counted items.

My theory about the nonobligatory use of the QUANTIFICATION construction in the counting of *qiang* people is thus as follows: Because *qiang* people were used in live sacrifices like animals and are listed in the inscriptions alongside other sacrificial objects and animals, their status as a superior category needs to be identified to reveal the devotion of the initiators in this ceremony. Although in everyday life, their existence as slaves and captives means that they were considered inferior and thus "unhuman" compared to civilians, it is in this special context as human sacrifice that their humanness needs to be highlighted again. The occasional use of the quantification unit *ren* "person/people" is thus a marker of its functional category.

My second observation based on the occurrence frequency in Table 2 is that the frequencies of occurrence of these instances are highly uneven. I would like to direct the reader's attention to the occurrence frequency of "*qiang* (numeral) *ren*." Notice that, although its occurrence is rare in ratio to the overall counting of *qiang* people, its

frequency of occurrence is rather high compared to other instances (types), the same as the counting of ceremonial wine in containers. These two items have high token frequency and therefore could be powerful exemplars that drive future developments of the QUANTIFICATION construction. However, the role of these two high-frequency instances might not be clear at this point. I will return to the discussion of these elements as exemplars after I examine the developments of the QUANTIFICATION construction in bronze inscriptions. The other instances (the counting of cowrie shells, chariots, and horses), although they seem to have rather low token frequency, developed a high frequency of usage in bronze inscriptions, as the following section will show. I suspect this has to do with the cultural significance of these items.

2.5. Findings in Bronze Inscriptions

2.5.1. Contexts of Occurrence

In the bronze inscriptions, especially from the sixteenth to the eighth century BCE (corresponding to the period from the Shang Dynasty through the Western Zhou Dynasty), expressions of quantification are mostly found in the vessels that document imperial bestowals and war booties. In the first category, common things being bestowed to a person could include ceremonial wine, currency of different kinds (cowrie shells, gold), armor and weapons, chariots and horses, jewelry and silk, etc. Sometimes, land and people would be given in a bestowal as well. In the second category, common things that are reported as war booty include captives, chariots, and horses. To create a corpus of these two content categories, I used keywords that occur in these categories to search the online corpus. Based on keyword searching with 易, 賞, and 舍, which all mean "to bestow," I collected 545 entries. Based on keyword searching with 孚, which means "to capture," I gathered 27 entries. Combining these results, I built a small corpus consisting of 566 entries (with reduplicates removed). Each of the 566 entries represents an entire piece of text from a single bronze vessel, the length of each piece of text ranging from one to 490 characters.

Among these 566 entries, 136 use some kind of numeral quantification expression, and some entries contain multiple types of quantification expression. In an

inventory list style, the quantifications for different items can take a variety of forms, including (a) [NOUN NUMERAL] and (b) [NOUN NUMERAL UNIT]. In addition to these forms with numerals, there are two other quantification expressions that are without numerals but I did not take them into consideration. These are (c) a single [Noun] to indicate it is one count, and (d) [UNIT NOUN] or [NOUN UNIT], which also indicates one count. Of these four ways to express quantification, forms (b) and (d), similar to the situation of quantification with quantification units in oracle-bone inscriptions, only occur with specific items (nouns). Form (d) occurs quite restrictively, almost exclusively with the quantification of silk and horses, and also is extremely rare.[9] For now, the focus is given back to form (b), the QUANTIFICATION construction.

2.5.2. The QUANTIFICATION Construction

In the corpus containing 136 entries with some forms of numeral quantification, I identified 111 entries that have at least one instance of the QUANTIFICATION construction. What I observed from these 111 entries includes the following findings.

First, the inventory of quantification units has greatly expanded as my 111 entries basically cover all the 32 quantification units identified in Da (2004).

Second, despite the expansion, the QUANTIFICATION construction is still the marked (and not the default) way to quantify. As I remarked earlier, it is used only on specific items. The pairings of quantification units and nouns are more predictable than productive, meaning that in most cases, a quantification unit is used only with a specific noun, and a noun tends to only use one or two quantification units. In this period, the nonproductive aspect of the QUANTIFICATION construction remains similar to its usage condition in oracle-bone inscriptions.

Third, the obligatory use of the QUANTIFICATION construction in counting some of these items has been reinforced from the usage condition in the previous era. In the oracle-bone inscriptions, among the five instantiation types of the

[9] This form has been discussed elsewhere as contributing to the emergence of numeral classifiers (Gong 2007; Zhang 2002), but since the current study is opposed to a transformational approach, I will save discussion of this form for another paper in the future.

QUANTIFICATION construction, only one type has obligatory usage (cowrie shells), two show some obligatoriness (chariots and horses), and the other two are all quite flexible (*qiang* people and ceremonial wine; refer back to Table 2 for details). In this period, more items enforce the obligatory use of the QUANTIFICATION construction. For instance:

(14) 王_易_呂夫_鬯_三_卣[。]
king_bestow_NAME_ceremonial wine_three_wine container[.]

貝_卅_朋[。] [2754]
cowrie shells_thirty_*peng* (ten cowrie shells)

Table 3 lists the instances that uses the construction obligatorily when a quantification larger than one needs to be expressed in the entire *Collection of Bronze Inscriptions*. Comparing this table to Table 2 (instantiation frequency of the construction in oracle-bone inscriptions), we observe that:

(i) cowrie shells have remained obligatory in the use of QUANTIFICATION construction in counting, and it has gained token frequency in this period.

(ii) Ceremonial wine, which was previously flexible in the use of the construction has also become obligatory. It had a high-frequency of occurrence in oracle-bone inscriptions and remains among the high-frequency instances in bronze inscriptions.

(iii) horses, and chariots have also developed obligatory usage of the QUANTIFICATION construction. These are the items that were only somewhat obligatory and occur with low frequency in oracle-bone inscriptions. It seems that their cultural significance and relevance has, at this stage, manifested in a higher frequency of occurrence in the text.

According to the calculation of the *Index* by Zhang (2001), nouns that occur with the highest frequency in the entire body of bronze inscriptions are (from high to low)人/*ren* ‘person’ (299 tokens), 車/*ju* ‘chariot’ (216 tokens),馬/*ma* ‘horse’ (207 tokens), and 貝/*bei*

Table 3. Items that use the QUANTIFICATION construction obligatorily in bronze inscriptions

Instance	Meaning	Token frequency	Frequency of quantification of the noun
bei (numeral) *peng* 贝(numeral)朋 "cowrie shells (numeral) ten cowrie shells"	Counts the number of cowrie shells by a unit of *peng*, a collective of ten cowrie shells	56	56
ma (numeral) *pi* 马(numeral)匹 "horse (numeral)one unit"	Counts the number of horses by an individual unit	18	18
chang (numeral) *you* 鬯 (numeral)卣 "ceremonial wine (numeral)container"	Counts the number of ceremonial wine by container/vessel	15	15
tian (numeral) *tian* 田 (numeral)田 "farm land (numeral) farm land"	Counts the number of farm land by farm land	6	6
ju (numeral) *sheng* 車(numeral)乘 "chariot (numeral) one horse-drawn chariot unit"	Counts the number of chariot by the number of chariot units	3	4
ju (numeral) *liang* 車(numeral)兩 "chariot (numeral)one horse-drawn chariot unit"	Counts the number of chariot by the number of chariot units	1	

'cowrie shells' (187 tokens)—chariots and horses are the highest frequency items ony following people. The use of QUANTIFICATION UNIT in the construction with chariots and horses is espeically worth noting. In oracle-bone inscriptions, chariots and horses share a quantification unit *bing*. But in the bronze inscriptions of Western Zhou, they share two quantification unit *liang* and *sheng*, and the horse additionally has its own quantification unit *pi*. As Table 3 shows, in my corpus, the counting of horses in *pi* and the counting of chariots in *liang* or *sheng*, are among the high frequency quantification items.

I went beyond the data of the small corpus I built and examined all the instances of horses and chariots in bronze inscriptions. Table 4 shows that all instances of horse quantification use one of the three quantification units, and more than half of the instances of chariot quantification use one of the two quantification units (when the quantification of chariots co-occurs with horses, they share the obligatory use of a single quantification unit). The counting of horses across contexts in the entire corpus of bronze inscriptions is thus obligatory.

Table 4. Obligatory use of quantification units with horses and chariots in bronze inscriptions in counting

	Horses	Chariots
pi: unit for horses	32	N/A
sheng: unit for horse-drawn chariots	11	11
liang: unit for horse-drawn chariots	8	2
No quantification unit	0	7
Total number of occurrences	51	20

Remember that in oracle-bone inscriptions, the two items share one quantification unit 丙(*bing*), which I attribute to the functional integration of horses and chariots as a vehicle—a horse-drawn chariot. Not using this quantification unit might indicate that their function in the context is irrelevant to the horse-drawn chariots. I assume this is similar to the condition found in the bronze inscriptions. When the two items share a quantification unit, they are used to indicate their function as part of a horse-drawn chariot. However, in the bronze inscriptions, there are two shared quantification units: *sheng* and *liang*. Additionally, horses have their own quantification unit, which counts an individual horse and is irrelevant to the counting of chariots. I believe the coexistence or possibly the competition of these quantification units to quantify horses and chariots reveals a problem the linguistic system tried to solve. Eventually, the solution of this problem affects how quantification is expressed in the language. Specifically, the creation of *pi* as an individual quantification unit for horses attempts to create a method to count horses in an accurate and economical way in a changing socio-historical context that requires an update of language specification in this aspect.

2.5.3. The Case of "*Ma* (numeral) *pi*," Chariots and Horses

Archaeological findings point to the existence of only one form of chariot, drawn by two horses, in the Shang dynasty. As data in Wu (2009) show, excavations of chariots and horses at thirty-five Shang burial sites are remains of either one chariot with two horses or two horses without a chariot.[10] Following Shang, Western Zhou sees the emergence of chariots drawn by four horses amidst the presence of the previous two-horse chariot form. Seven out of thirty-five burial sites from Western Zhou feature remains of four-horse chariots (Wu, 2009, pp. 212–17).

Such findings are coherent with the linguistic representation regarding chariots and horses. With such a solid piece of evidence, I can confidently say that as two-horse chariots are the only form that existed in the Shang dynasty, the only quantification unit for counting chariots and horses, *bing*, in oracle-bone inscriptions counts two horses/one chariot. In Western Zhou, as there were two forms of chariots, *liang* and *sheng* respectively each represent one form. And because all later literature has accepted the interpretation of *sheng* as counting four-horse chariots, it is reasonable to deduce, given the historical facts and the linguistic contexts above, that in bronze inscriptions, *sheng* counts a chariot (drawn by four horses) or four chariot-drawing horses, and *liang* counts a chariot (drawn by two horses) or two chariot-drawing horses. There has also been etymological evidence supporting this point of view.[11]

---

[10] Wu (2009) presumably is not an exhaustive study of all the excavated sites to date, and the selection criteria of her data are unclear. However, the quantity and range of the sites presented in the data are substantial and sufficiently representative.

[11] *Liang* has an early etymology of 兩. Its variations as shown in *Shuowen Jiezi Gulin* (Ding 1927:(32)3371) invariably suggest that it is an iconic representation of a pair. *Shuowenjiezi Gulin* also records an account which attributes its later development in the form of 两 due to the rise of meaning of "a pair forms a whole" ("兩者，積雙而為一，故從一也"), hence the horizontal stroke that covers it. Some early records interpret its usage with chariot from its "double" denotation, "because a chariot has two wheels, it is called *liang*; just as shoes have two to make a pair, they are also called *liang*" ("車有兩輪，故称為兩，犹履有两隻，亦称为兩"). The use of *liang*, and how it is interpreted as a whole, is thus coherent with its etymological evolvement. If this is the case, then *liang* in bronze inscriptions would have the same meaning as 丙. In addition to their orthographical connection, it is likely that *liang* evolves from *bing*.

The fact that horses and chariots share the same set of quantification units (*bing*, *liang*, and *sheng*), as exhibited in the two types of inscriptions, commonly illustrates the conceptual and functional tie between chariots and horses. From the records of fortune telling in the oracle-bone inscriptions, we can see that in the Shang dynasty, the function of horses is for drawing chariots. Questions like "which side should a new horse be used for" and "which horse should be used to replace an injured horse on the left-hand side" were extensively recorded—all hinting at the function of horses as drawing chariots (Liu and Cao 2004). The conceptual binding of chariots and horses to some extent hinders the free expression of the concept "horse." The realization of horse counting with a quantification unit referencing and used for the chariots is conceptually and linguistically cumbersome and confusing, and this could constitute a motivation for creating an expression to count horses on their own.

Subsequent to Shang dynasty in Western Zhou, there is evidence showing that the perceptual separation between horses and chariots emerged due to a change in the function of horses. Primarily used for drawing chariots in the Shang dynasty, horses started to be used independently for riding in Western Zhou (Yang, 2007). This constitutes yet another real-world necessity for the invention of a quantification unit solely for counting horses. The emergence of *pi* in bronze inscriptions represents a well-motivated functional specialization. The counting of horses became independent of the counting of chariots and achieved one-to-one correspondence between a sign and what it signifies, thereby increasing the economy of communication.

Data in Table 4 also shows that *pi*, once introduced, is well adopted into the system. In bronze inscriptions of Western Zhou, *pi* is already a preferred quantification unit over *sheng* and *liang* during Western Zhou. Orthographical evidence from bronze inscriptions between the sixteenth and the second century BCE also suggests that fourteen of the twenty-two occurrences of "*ma* (numeral) *pi*" count a set of four horses, and nine occurrences of "four *pi*" are written as "combined characters" (*hewen*). The frequency effect of having four horses together in reality might contribute to both phenomena. However, the fact that *pi* is chosen over *sheng* in the QUANTIFICATION

construction of horses suggests that *pi* has gained conventionality for the counting of horses.

2.6. Discussion

Before discussion specifically aiming at numeral classifiers, I hope that through the data presented above, it is now clear that the larger construction—the QUANTIFICATION construction from which numeral classifiers emerged—was itself undergoing drastic developments. In the oracle-bone inscriptions, its usage is not obligatory with the few instantiations it has, but in the bronze inscriptions, it has become obligatory in the expression of certain items. This is before it became the default way of quantification in Chinese, when the inventory of quantification units has grown to an impressive lot, and the usage of quantification units has become productive rather than exclusive with only a noun.

Therefore, the emergence of numeral classifiers is in conjunction with the emergence of other quantification units and the development of the QUANTIFICATION construction overall. In these two periods (of oracle-bone inscriptions and bronze inscriptions), the instantiations of the construction provide concrete usage instances that allow users to make generalizations about the form and meaning/function of this construction that affect the further instantiations of this construction. Here, I would like to discuss the effects produced by these instantiations. Examples (15) and (16) reprint the high-frequency instantiations of the constructions in the order of high frequency to low.

(15) Oracle-bone inscriptions:
a. *qiang* (numeral) *ren* (counting *qiang* by person)
b. *chang* (numeral) *you* (counting wine by container)
c. *ma* (numeral) *bing* (counting horses by a chariot drawn by two horses)
d. *bei* (numeral) *peng* (counting cowrie shells by unit of ten)
e. *ju* (numeral) *bing* (counting chariots by a chariot drawn by two horses)

(16) Bronze inscriptions:
a. *bei* (numeral) *peng* (counting cowrie shells by unit of ten)
b. *ma* (numeral) *pi* (counting horses by horse)
c. *chang* (numeral) *you* (counting wine by container)

d. *ju* (numeral) *sheng* (counting chariots by a chariot drawn by four horses)

In these instantiations, we can actually find prototypes (in the sense of early forms) for the different categories of quantification units identified by Chinese grammarians, as those identified in section 2.1 of this chapter. For instance, *ren* 'person' is a prototype of a numeral classifier; *you* 'wine container' is a prototype of a container unit; *bing* 'a unit of a chariot drawn by two horses' and *peng* 'a unit of ten cowrie shells' are both prototypes of counter units (*peng* is probably a more suitable prototype because its members are homogenous, while the components of *bing* are heterogenous, containing both horses and a chariot station). All of these quantification units have contributed to a common generalization of the QUANTIFICATION construction, as well as showing the units' idiosyncrasies. The generalizations enabled by these quantification units include various aspects, such as the kind of items to be counted with the constructions, the relationship between the item and the quantification unit, the lexical source of the quantification unit, etc. The variety of these usage instances, or the type frequency of the construction in oracle-bone inscriptions, has consolidated its status as a linguistic construction, and the strengthening of the construction has in turn warranted its further instantiations.

In the bronze inscriptions, we witness the entrenchment of "*bei* (numeral) *peng*" from low frequency to high, and the entrenchment of "*chang* (numeral) *you*" from optional to obligatory; they occur naturally and predictably from their counterparts in the oracle-bone inscriptions because of their prototypicality in their category (cowrie shells as the typical object that needs to be counted in multiples, and wine as a valuable liquid that needs to be counted with containers that are countable) and also their cultural significance that continues to exist in the society. As I pointed out in Chapter I, language acquisition research has shown that a single exemplar can induce powerful generalizations that give rise to an entire category, and I believe this was the case with these two instances. They have respectively created subconstructions of the QUANTIFICATION construction: one that quantifies count nouns by a unit of multiple items (counter unit) and the other that quantifies noncount nouns by its countable

container (container unit). The conceptual validity of these two categories can be seen in other instantiations with similar functions in bronze inscriptions.

In a similar fashion, I believe it is from the instantiations and entrenchment of "*ma* (numeral) *pi*" that the conceptual category of numeral classifier emerges. Possible generalizations that form around this specific instance could include that the quantification unit representing an individual unit of horse, or that the lexical source of the unit does not bear apparent semantic connection to the horse. In other words, the quantification unit is an abstract representation of an individual representative of the noun. This is the characteristic that numeral classifiers exhibit today and throughout history since their emergence, as mentioned in section 2.2. This is also a characteristic that distinguishes the category from other quantification units. Such a characteristic could not have emerged in association with the counter units and the container units—if anything, it could only have emerged in opposition to the counter units and the container units by defying the collectivity of counter units and the concrete representation of the container units. But this does not mean that the emergence of "*ma* (numeral) *pi*" is completely baseless. As I have hypothesized in section 2.4.3, it emerges against the collective counting of horses by *sheng* and *bing* or *liang*, but it also draws analogy from a precedent instantiation, the exemplar "*qiang* (numeral) *ren*" in the inscriptions. I believe this is the legacy of "*qiang* (numeral) *ren*" as the highest-frequency instance of QUANTIFICATION construction in the oracle-bone inscriptions. *Ren* as a prototype (in the sense of basic and common) of the numeral classifier is a strong exemplar that provides analogy for the counting of horses individually.

Whether *ren* as it occurs in the instance of "*qiang* (numeral) *ren*" can technically be considered a numeral classifier is widely debated. It was supposed that because *ren* 'person' is not an abstract representation of *qiang* people, it does not technically count as one. However, in my opinion, *ren* 'person' already exhibits some level of abstraction, especially when we compare it to *qiang*, which occurs in the noun slot. *Qiang* makes specific reference to a type of human, while *ren is* a generic term; by definition, a generic term is closer to being a concept, and, understandably, more abstract. The fact that there are only instances of "*qiang* (numeral) *ren*" rather than "*ren* (numeral) *qiang*" allows the

users to make the generalization that in this construction, the quantification unit has less concrete denotation than, if not the same level of concrete denotation as, the noun. The individual counting of *ren*, as well as its abstract representation of the counted, sanctions the innovation of an individual-counting and abstract-representing quantification unit for horses when counting in a counter unit does not work anymore. Although "*qiang* (numeral) *ren*" no longer exists in the language because the concept of *qiang* people has lost relevance in the social context, the generalizations it initiated and the prototypical characteristics of a quantification unit it established had been embodied in "*ma* (numeral) *pi*," which established more concrete generalizations for numeral classifiers.[12]

2.7. Summary

Overall, the analysis above shows that, informed by written records of Archaic Chinese, the emergence of the numeral classifier did not occur in isolation. It is first situated in the development of the QUANTIFICATION construction. It develops as the QUANTIFICATION construction creates a grammatical slot for all types of quantification units, and it develops simultaneously with other quantification units, as different types of instantiations in the quantification unit slot.

When examining the type frequency of the quantification unit slot in oracle-bone inscriptions and bronze inscriptions, I did not find a rich variety of types in this slot, but only a handful of usage instances at this stage. And when further considering the token frequency of these types, my attention is drawn to the counting of *qiang* people and chariot and horses. The quantification units of both "*qiang* (numeral) *ren*" (in oracle-bone inscriptions) and "*ma* (numeral) *pi*" (in bronze inscriptions) can be considered early forms of numeral classifiers, and they are both high frequency tokens in the corpus. I believe they are the items that drive the development of numeral classifiers as a type of quantification unit. In other words, my analysis based on available data suggests that the historical development of numeral classifiers that has been observed in the written

[12] *Ren* as a quantification unit continues to be used in bronze inscriptions, its usage distributed in occurrence with different nouns with very low frequency, and I doubt it could create any substantial effects that impact users' generalizations towards the QUANTIFICATION construction or the mental representation of numeral classifiers.

records of Chinese seems to locate the beginning of this process at these two exemplars. The generalizations formed around the numeral classifier as a specialized type of quantification unit which remains current in modern Chinese do not emerge out of transformations of different ways to express quantification, and specifically not from different ways to count individual horses. Rather, its true beginning is found in a single usage instance, especially the instance "*ma* (numeral) *pi*" in bronze inscriptions, which takes analogy from "*qiang* (numeral) *ren*" in oracle-bone inscriptions. Contrary to the proposal ofTraugott and Trousdale (2010, 2013) that constructionalization occurs in a gradual fashion, the current study shows a case of linguistic innovation that takes place in a less gradual, more abrupt manner.

I should point out that the discussion of this chapter is based on Chinese written records, and I did not rule out the possibility that a) the emergence of numeral classifiers is a result of language contact, and as a result, b) the earliest usage is within the colloquial language before spreading to the written. However, judging by the usage contexts and the early instances of numeral classifiers (the Qiang people and horses), which are cultural artifacts that are reserved for the aristocracy, (b) is unlikely. But the robust numeral classifier systems present in Austroasiatic languages strongly point to the possibility of a).

What has been examined in this chapter provides a case of linguistic innovations whose high-frequency exemplars reveal the internal and external factors of change. The development of the QUANTIFICATION construction overall has provided a constructional slot for the emergence of the numeral classifier, and this could be considered as one internal factor, but there are other internal and external factors at work that lead to the sudden or abrupt emergence of "*qiang* (numeral) *ren*" and "*ma* (numeral) *pi*." With the case of "*qiang* (numeral) *ren*," an external factor would certainly be the presence of *qiang* people in the society. Conceptually, they constitute a special category that is half object and half human in the sense that when reduced to slaves, they are labor powers just like other livestock and can be sacrificed in ceremonies. Their existence, or the need to quantify them, thus eventually led to inclusion of humans in this kind of quantification which had been reserved for the counting of objects for various purposes.

Their presence in the world external to the linguistic system triggers an internal reconceptualization of object/human category. Regarding the case of "*ma* (numeral) *pi*," as I have proposed, the trigger was initially the use of a quantification unit denoting a chariot structure to count horses. This is disrupted by the conceptual and functional separation of horses from chariots, and also the presence of multiple quantification units for the counting of horses due to the changing structure of chariots (from one drawn by two horses to one drawn by four) which causes confusion and redundancy in counting. In this scenario, it is difficult to tease the internal and external apart, and this is an excellent example that shows how a linguistic system actively interacts and responds to reality in the society, resulting in linguistic innovation.

CHAPTER III
EMERGENCE OF A CLAUSAL CONSTRUCTION: THE CASE OF THE INNOVATIONS OF THE *BEI* CONSTRUCTION

3.1. Introduction

This chapter will examine a familiar Chinese clausal construction, the *bei* (被) passive, which has developed new subconstructions on the Internet. The changes in this construction have been quickly identified and adopted by its users, and it is recorded and studied as a social/linguistic phenomenon of some scale by scholars (See Xiao, 2016 for instance, for an English publication). However, a common linguistic premise that the majority of the scholarship is based upon is problematic, and this chapter will offer a revision to this linguistic premise.

3.1.1. Canonical *Bei*

Before delving into the discussion of *bei*'s new uses, some relevant introduction to the familiar usage of *bei* is needed. *Bei* is canonically used as a passive marker that appears in the passive construction:

(17) [$N_{patient}$ PSV (e.g. *bei,zao*) (NP) VP]

In this construction, the grammatical subject is the semantic patient, and an optional semantic agent comes after the passive marker, with the predicate–typically a transitive verb that can take two arguments–following. For instance:

(18) *ta_bei_ren_da_le*
he_PSV_person_hit_PFV
'he has been beaten by someone'

Here, a simple event was described. The grammatical subject of the sentence, *ta* (3SG), has been beaten by an unnamed person. Several semantic characteristics strongly associated with the *bei* passive can be found in this example. First, there are two participants in this event, and an action is being transferred from one to another. In this case, the action of beating is initiated by the oblique *ren* 'someone' and is received by the grammatical subject *ta* 'he.' Second, from the construal of the event, it is implied that the

grammatical subject does not have control over the action acted upon him. Third, the event described has been completed.

Da 'to hit/to beat' is inherently an activity verb, but with the perfective marker *le*, the activity becomes bounded. Although it is possible that the same action can take place again, the individual instance has been completed. The implication of transfer, volition, and telicity all suggest transitivity (Hopper & Thompson, 1980) as a notion highly relevant to the *bei* passive (a similar observation can be found in Li & Thompson, 1981, p. 501).

For pragmatic reasons, the semantic agent can be omitted in the *bei* passive, and the only difference is that information on the semantic agent now does not have a formal representation. For instance:

(19) *ta_bei_da_le*
he_PSV_hit_PFV
'he was beaten'

I shall call this the suppression of semantic agent, because sometimes the information on the semantic agent can still be inferred from context. The *bei* passive thus factually has two formal representations—one with the formal representation of the semantic agent, and the other without it:

(20) a. [$N_{patient}$ *bei* N_{agent} VP]
b. [$N_{patient}$ *bei* VP]

Prior to modern era, *the Grammar of Modern Chinese* (Wang, 1943) stated that the Chinese *bei* passive primarily had an adversative meaning—that the occurrence of VP to the grammatical subject (semantic patient) is deemed infelicitous and undesirable to the subject (see reprints in 1985, p. 132). For instance, the example of "being beaten" is a prototypical usage of the adversative *bei* passive. The adversative passive is in fact a typologically-common construction that is found not only in Chinese, but also in several East Asian and Southeast Asian languages and beyond (Abraham & Leisiö, 2006). In Contemporary Chinese, however, it has been observed that there is a neutralization process due to the influence of European languages (Wang, 1958; Chao, 1968; Li & Thompson, 1981), which is confirmed by corpus data (Xiao, McEnery, & Qian, 2006).

While the overall development of *bei* passive is towards de-adversatization, the adversative meaning is present in certain usages. For instance, Chappell (2009) remarks that the adversative meaning of *bei* is largely present in colloquial speech, and Jing-Schmidt and Jing (2009) observed the adversative usage of *bei* prominently in Internet news headlines compared to news texts and novellas. This suggests that the so-called neutralization of *bei* is context-dependent and not systematic. The adversative meaning of *bei*, rather than the neutralized meaning of *bei*, is what is inherited in the new *bei* construction.

3.1.2. Problems with the Proposal of New *Bei*

The new developments of *bei* swept across the Chinese-speaking community, and not just as a linguistic phenomenon that interests Chinese linguists as it has caught the attention of the general public, especially in cyberspace. In 2009, *bei* was voted by 200,000 Chinese netizens as the 'Word of the Year.' The new usage that led to the public acknowledgement was quickly updated in the *Modern Chinese Dictionary:*[13] "used before verbs or nouns to indicate the (described) circumstance is **not true**, or is being **forced upon** (with a sarcastic undertone)" (first updated 6th edition, 2012; same in 7th edition, 2016, p. 59). Generally speaking, the grammatical and pragmatic characterization of the "new usage of *bei*" as defined in the MC dictionary was accepted in Chinese academia. It was agreed that the new usage has a formal representation of "*bei* X" (Shi, 2013), which indicates that :

(i) The grammatical category of the VP slot following *bei* has expanded. Typically taking only transitive verbs, the VP slot can now take intransitive verbs (including adjectival verbs) and nouns (see also Wang, 2009).

(ii) The suppression of the semantic agent is now compulsory.

In terms of its pragmatics, it is suggested that the new usage conveys a strong attitudinal stance towards the event, rather than being a mere objective description. In particular, it

[13] The most widely circulated dictionary of Contemporary Chinese. Its entries include both characters and vocabularies.

is associated with a sense of infliction, coercion and falsity—all conveying a negative stance (Shi, 2013; Wang, 2011).

The conceptualization of the meaning of the new usage is less unanimous than the conceptualization of its grammatical and pragmatic features. Although the two meanings identified by the MC dictionary ("untrue" and "forced upon") are relevant in most discussions, the interplay of the two meanings are constructed differently.

For instance, Shi (2013) proposes that "voluntarity" and "factuality" are two "pragmatic contents/features" of the new usage of *bei*. These two "pragmatic features" have binary values (each has a positive and negative value), that produce four combinations, with three of them observable in actual usage data. For instance, the combination of [- voluntary] and [+ factual] produces the meaning of "X being forced upon" the grammatical subject, such as:

(21) *zhongduo_haizi_* **"*bei_ticao*"** [21427][14]
many_child_**PSV_gymnastics**
lit: a lot of children are gymnastics-ed
'a lot of children are forced (by their parents) to learn gymnastics'

In this example, the word *ticao* 'gymnastics,' which is typically a noun, appears in the VP slot of the *bei* passive. The sentence is making the observation that a lot of school-age children are being sent to a gymnastics school at Xiantao, China, mostly because their parents are hoping that their children will be future Olympic champions like the four trained by this school during the last three decades. The emphasis of this sentence is thus on the grammatical subject, *zhongduo haizi* 'many children,' who were sent to the school to start training as gymnasts without having the opportunity to make their own decision, and were possibly even sent against their will. It is therefore considered [- voluntary]; meanwhile, because the practice of gymnastics has taken place, it is [+ factual]. Another combination, [- voluntary] and [- factual] produce the meaning of "not true". For instance, in the following example, one of the interpretations of "*bei zisha*" would mean 'it is not true that the whistle blower has committed suicide':

[14] When a number follows the examples, it indicates that it is from the sample corpus of *bei.* See details in section 3.2.3.

(22) *jubaoren*_**"*bei_zisha*"** [66988]
whistle-blower_**"PSV_commit_suicide"**
lit: the whistle-blower was "committed" suicide
'the whistle-blower allegedly died of suicide' (versus the fact that suicide is not the real cause of death, or the fact that the person is still alive)

A third combination, [+ voluntary] and [-true] produce a meaning seen less often (but his two examples fail to show the voluntarity of the participants involved). According to Shi (2013), "*bei* X" is polysemous: All three meanings proposed can be a possible interpretation of a certain usage instance. He argues, for instance, that the phrase "*bei zisha*" in example (22) can also have the meaning of "forced upon," and the sentence would mean 'the whistle-blower was forced to commit suicide' (p. 14).

A feature-based approach to the meaning of "*bei* X" is also seen in Yuan and Liang (2016). Based on a semantic analysis of 158 usage instances of "*bei* X" they collected from newspapers and a neologism phrasebook, they conclude that there are three basic semantic features of "*bei* X": falsity, imposition and adversity. These semantic features can be found in the usage instances singularly or in combinations, producing seven semantic categories.

In Xiao (2016), "*bei* X" is thought of as a product of language change within the preposition *bei*. She proposes that as *bei* becomes a pseudo-prefix that is attached to X, it is accompanied by semantic bleaching, losing its old meaning of the "passive experience caused by a concrete and recoverable source" and gaining the new meaning of "a feeling of being helpless, or grievously wronged by an unknown/unfathomable source" (p. 92).

Regardless of the approach, the most problematic aspect common to all three analyses is the idea that the meaning of "forced upon" or "imposition" is optional or lost in new *bei*. The examples these analyses are based on, however, show exactly the opposite of such a proposition, all revealing a sense of something imposed onto the grammatical subject without its consent or even against its will. The adversative meaning, contrary to what certain scholars have proposed, is also an inherent implication in the act of imposition. It is thus wrong to presume that "forced upon" and "adversity" are features of the new *bei*. Additionally, Xiao's construal of the new *bei* based on its formal change (the suppression of agent) is also problematic. Instead of supporting her claim, each of

her examples point clearly to one (or multiple) imposing agent(s), which is obvious when situated in the social context of language use.

These obvious flaws of the current analyses of the new *bei* (especially exemplified by the studies mentioned above) thus necessitates a re-examination of this linguistic innovation. In the following, guided by research methods in a usage-based Construction Grammar framework, I will examine naturally-occurring language data of the new *bei* and present an analysis based on its pattern and context of occurrence. The data and methods will be detailed below.

3.2. Data, Methods and Findings from Corpus

3.2.1. Source of Data

For this investigation, I chose the Simplified Chinese TenTen corpus hosted by Sketch Engine. It is 1.7-billion-word corpus, created by web-crawling performed in 2011 and it contains contents in and prior to 2011. As several previous studies have pointed out, the new *bei* construction finds diffusion mostly through online news media. The source and compiling date of the corpus is most suitable for this study, primarily because it is a corpus built entirely of content published online, including 2.6 million online documents from news websites, official websites (of institutions), informational websites and online forums. Second, the corpus was built after the emergence of the new *bei*. *Bei* was elected "word of the year" in 2009; a corpus built in 2011 can thus capture language data prior to and after this public acknowledgement of the new usages, and it provides a vantage point for me to analyze the linguistic changes of *bei*. The corpus has also been word-segmented and POS-tagged to allow searching based on CQL grammar in the online interface.

3.2.2. Query Terms

From language data presented in earlier studies, I notice that the new *bei* construction frequently appears in double quotation marks in the text ("被 X")—a feature generally left undiscussed in previous studies. In Chinese, double quotation marks are the default form when quotation marks are needed. In this context, the use of a pair of wrapping double quotation marks indicates the expression in it is a coined (novel)

expression.[15] Using the pattern ("被 X") as a search template thus helped quickly identify the new *bei* construction in the corpus.

3.2.3. Procedures

I first conducted an exploratory and qualitative analysis of the data based on a BEI Sample Corpus I built from the Sketch Engine Simplified Chinese TenTen Corpus. The TenTen corpus contains 2,134,389 instances of *bei*, and I generated a random sample of 100,000 instances of the construction from the two million instances (about 5% of the total occurrences, with the sample size being 100,000 because this is the maximal amount of instances allowed for download). They are saved locally as a text file with each instance taking up a single line (but in each line there might contain more than one instance of *bei*). This data, consisting of 100,000 lines of *bei* occurrences, is the BEI Sample Corpus. Following previous proposals on the syllabic structure of X in "*bei* X" being mostly disyllabic and trisyllabic (Xiao, 2016; Yuan & Liang, 2016), I identified 472 occurrences of *bei* followed by an X wrapped in double quotation marks (the length of X is one to three characters) in the BEI Sample Corpus. Browsing through the 472 occurrences, I noticed that the functions of the double quotation marks include marking the phrase as a quote, a reference of a technical term, or a coined expression that is not conventional in the lexicon. I then proceeded to remove the noise in the data, including quotes of sentences (e.g. "*bei ma-dao le*" 'being numbed'), instances of "*bei* X" that could be understood in the traditional passive usage (e.g. "*bei kuishi*" 'being spied on') and *bei* as a prefix (e.g. "*bei xuke-ren*" 'the person who has been approved'), leaving 198 occurences of *bei* in the form of "*bei* X" as the actual data I examined closely.

I manually went through these instances one by one and examined their context of occurrence to understand their meaning and usage—as language users would in encountering a new usage. The usage of the 92 distinct types (instances) in 198 occurrences fall into two subtypes, which can roughly be mapped to the two meanings identified by the MC dictionary. I thus coded each of these 198 occurrences into two

[15] The identification of non-standard usage is important in the formal register. See the use of quotation-marked expressions in a national report given by the premier of China at http://news.ifeng.com/a/20150305/43278612_0.shtml

categories. The details of token frequency and type frequency of the two subtypes can be found in Table 6 and Table 7 in the next section.

Parallel to the qualitative analysis conducted in the BEI Sample Corpus, I also conducted a quantitative analysis in the entire TenTen corpus. I made the following two queries using CQL grammar, and asked the platform to produce a frequency list based on the queries:

(a) Simple query: " 被 ?? "
(b) Simple query: " 被 ??? "

In the first query, I asked Sketch Engine to find any phrase that starts with a beginning double quotation mark followed by the character *bei*, which is then followed by a disyllabic word and then an ending double quotation mark. The second query is similar, except the word that follows is trisyllabic. The findings are compared to the findings in the BEI Sample Corpus, and the details can be found in Table 8 in the next section.

3.2.4. Preliminary Findings from the BEI Sample Corpus and TenTen Corpus

My coding of the data in the BEI Sample Corpus reveals that the new *bei* construction has at least two idiosyncrasies. First, the element following *bei*, or X, is a lexical word. From what I observed, the semantic category of the lexical word is the opposite of verbs that shows transitivity and telicity, meaning they are verbs that take no arguments, such as intransitive verbs (e.g. *zisha* 'to commit suicide'), adjectival verbs (e.g. *xingfu* 'be happy'), nouns (e.g. *jingshenbing* 'mental illness') and even adverbs (e.g. *ziyuan* '(to act) voluntarily').[16] Second, the grammatical subject of the phrase "*bei* X" has a person bias—of the 92 types of the new *bei* construction, only 17 types occur with a non-person grammatical subject, but 6 of them are some kind of subjective entity, such as a school, corporation and the court. Putting these idiosyncrasies together, they indicate that the new *bei* construction is more than a phrasal construction, but is a clausal construction that exhibits constraints on the grammatical subject and the predicate. The use of of double quotations marks for the *bei* phrase is also a marked structural feature. The new *bei* construction has the constructional frame:

[16] In Chinese, verbs include both action/activity verbs and stative verbs (what would be considered adjectives in other languages); see the discussion in Chao (1968).

(23) [N_{person} "*bei* X"]

Sharing the same constructional frame, there are two subtypes of the new *bei* construction.

The first subtype is a more distinct category, which I call the MISREPRESENTATION construction. I will further discuss this subtype in section 3.3. Briefly, expressions in this category are used in circumstances when people are implicated in a claim about X. Table 5 lists some of the usage instances that I identified in the BEI Sample Corpus, along with their frequency of occurrence in this corpus. The type frequency of this subtype is 38 types, and they occur in 135 tokens (see the full list in Appendix A)

Table 5. X in "*bei* X" in the MISREPRESENTATION construction

token	definition	token frq	token	definition
re-occurring items			items that occur twice	
jiuye	'to be employed/employment'	25	*manyi*	'to be satisfied'
jingshenbing	'mental problem"	13	*daiyan*	'to endorse (a brand)'
zengzhang	'to increase'	12	*piaochang*	'to pay for prostitution'
zisha	'(to commit) suicide'	12	*qushi*	'to pass away'
daibiao	'to represent'	11	*kaixin*	'to be delighted'
xingfu	'(to be) happy / happiness'	7	*shishi*	'to pass away'
			lihun	'to divorce'
juankuan	'to donate money'	7	*jianqiang*	'to be strong'
xiaokang	'to be moderately prosperous'	3	*fumian*	'negative (review)'
			shouru	'income'
shizong	'(for a person) to go missing'	3	*huaijiu*	'to be nostalgic'
jiujia	'DUI'	3		

The second subtype is what I call the IMPOSITION construction, which I will further investigate in section 3.4. The function of this category is to present how people are implicated in a situation about X. The type frequency of this category is larger than the first subtype, with a total of 54 types, but they generally occur less frequently as they are only found in a total of 63 occurrences. Table 6 lists some of the instances I found in the BEI Sample Corpus (see a full list in Appendix B)

Table 6. X in "*bei* X", in the IMPOSITION construction

token	definition	token frq	token	definition	token frq
gaotie	'high-speed train'	4	*shangke*	'take courses'	1
shanglou	'to move into high-rises'	2	*maifang*	'to buy a house'	1
xiaosan	'mistress; lover'	2	*xiujia*	'to take some time off'	1
buke	'to make up lessons'	2			
zhangjia	'to raise the price'	2	*xinyongka*	'credit card'	1
xiguan	'to get used to; habit'	2	*guanxin*	'to care about'	1
baohu	'to protect'	2	*chuming*	'to be famous'	1
xiaofei	'to spend/consume'	2	*gongli*	'to be pragmatic'	1
fuwu	'to serve'	2			

Comparing the statistics of these items to that of the entire TenTen Corpus (in Table 8), I find the ranking based on the frequency of occurrence in the sample corpus to be reliable. The high-frequency items I noted in the sample corpus have also occurred with high frequency in the TenTen Corpus. To have a visual understanding of the distribution of the two subtypes of the new *bei* construction I highlighted instances in the MISREPRESENTATION construction in darker shade marks, and those of the IMPOSITION construction in lighter shade marks in Table 7. It is apparent that the darker shade dominates the top section of the table, indicating that high-frequency tokens of the new construction are primarily instances of the MISREPRESENTATION construction.

Of the two subtypes, the MISREPRESENTATION construction shows lower type frequency but higher token frequency for each type compared to the IMPOSITION construction, which has higher type frequency but very low token frequency for each type. The occurrence pattern of these two subtypes of the new *bei* construction prompted a hypothesis regarding the developments taking place in the *bei* construction. The cluster of high-frequency tokens in the MISREPRESENTATION construction suggests that the new developments of *bei* might have been initially driven by exemplars in this usage context. *Bei jiuye* 'being employment-ed,' *bei zengzhang* 'being increase-ed,' *bei jingshenbing* 'being mental-illness-ed' are all instances of a usage that emerged in response to specific events that are widely known. The knowledge of these events,

Table 7. High-frequency "*bei* (disyllabic) X" and "*bei* (trisyllabic) X" instances in the Simplified Chinese TenTen Corpus

"*bei* + (disyllabic) X"			"*bei* + (trisyllabic) X"		
token	definition	token frq	token	definition	token frq
jiuye	'to be employed'	226	*jingshenbing*	'mental problem'	89
zengzhang	'to increase'	121	*xinyongka*	'credit card'	28
daibiao	'to represent'	88	*jitihua*	'to collectivize'	25
ziyuan	'to act voluntarily'	72	*chengshihua*	'to urbanize'	21
zisha	'to commit suicide'	51	*bianyuanhua*	'to marginalize'	13
shanglou	'to move into high-rises'	46	*qianguize*	'hidden rules'	12
xingfu	'to be happy'	42	~~*guanlizhe*~~		
juankuan	'to donate money'	40	~~*boduogan*~~		
gaotie	'high-speed rail'	38	~~*shenqingren*~~		
siwang	'to die'	35	~~*tongzhizhe*~~		
xuyao	'to need'	31	~~*dailiren*~~		
xiaofei	'to spend/purchase'	29	~~*yanjiuzhe*~~		
~~*shidai*~~			~~*yapozhe*~~		
xuanze	'to choose'	25	~~*xukeren*~~		
liuxue	'to study overseas'	23	~~*beishuren*~~		
fuwu	'to serve'	23	*shiminhua*	'to civilize'	6
qushi	'to pass away'	22	*guojihua*	'to globalize'	6
liyong	'to utilize'	22	*gaopingjun*	'high average'	5
zhongchan	'middle class'	19	*jinronghua*	'to financialize'	5
manyi	'to be satisfied'	18	~~*xuanzezhe*~~		

Note: Items with strikethroughs are words of the form *bei-XX(X)* (not phrases) that were parsed erroneously by Sketch Engine. Words that are not highlighted are what I identified as the new usages from the sample corpus.

experienced personally and in one's immediate social surroundings, provide the dimension of embodied experience to the users, which allows them to construe the meaning/function mapped on the linguistic form. The other subconstruction, the IMPOSITION construction, might have only emerged in relation to the the MISREPRESENTATION construction, for instance, *bei daibiao* 'being represent-ed.' Because the latter subconstruction has a wider scope of concern (extending beyond the concern of "a claim about X" to "an event of X"), it became more productive, hence yielding more type frequency as observed in the corpus.

In the following sections I attempt to analyze the usage context of some of the high-frequency instances to present the kind of generalizations that emerge through these instances. Section 3.3 examines the MISREPRESENTATION construction, and section 3.4 provides a brief review of the IMPOSITION construction.

3.3. The MISREPRESENTATION Construction

By the term MISREPRESENTATION, I would like to argue that as one of subconstructions instantiating the form [N_{person} "*bei* X"], the MISREPRESENTATION construction describes the condition in which the grammatical subject person(s) N_{person} is misrepresented by an institutionalized discourse about X. This use of the construction has the function of satirizing such a misrepresentation. The subconstruction can be found in four usage contexts, but the most conspicuous one, and possibly the one that brings the new usages of *bei* to the attention of the wider public, is in encountering the fabrication of national statistics.

3.3.1. Misrepresentation by National Statistics

Governments regularly rely on national statistics to make claims. For instance, "our college graduates have jobs," or "our citizens are quite well-off." But when people notice that their personal circumstances are not consistent with such claims, as in the case of Chinese netizens, they begin using the construction [N_{person} "*bei* X"] to expose the falsity of statistics about X. *Bei jiuye* 'being employment-ed,' *bei zengzhang* 'being increase-ed,' *bei xingfu* 'being happy-ed,' *bei zhongchan* 'being middle-class-ed,' *bei xiaokang* 'being moderately-propsperous-ed,' *bei manyi* 'being satisfied-ed' all emerged in this context.

Bei jiuye 'being employment-ed' is the "*bei* X" instance with the highest frequency of occurrence in the corpus. It is tied to a news event, which constitutes the key to understanding the function of the construction. In my corpus, an article remarks that:

(24) *qunian_biye_ji,* ***"bei jiuye"_****yi_ci_*
last year_graduation_season, **"PSV employment"**_a_ word_

kai_"bei shidai"_xianhe [11762]
start_"bei era"_beginning

'Last year, the word "being employment-ed" marks the beginning of the era of *bei*.'

Note: the word *bei shidai* 'the era of *bei*' was coined to denote the rampant usage of "*bei* X" expressions and correspondingly the social phenomena that had stirred up such usages.

The news story that captured media attention was prompted by a 2009 college graduate, who, upon graduation, discovered in his academic records an employment agreement between himself and a company. He was not aware of its existence, he had never signed the agreement himself, and the company as a functioning economic entity never existed. Similar documents were also found by fellow students in their files. The student posted an article about his situation, coining the expression *bei jiuye* 'being employment-ed.' His article provoked public investigation into the validity of the college's reported employment rate because he believed the document was forged to boost the employment rate on paper.[17]

The exposure revealed through the expression of *bei jiuye* seems to have successfully achieved its purpose. It had taken its toll on the credibility of statistics released by the Ministry of Education less than a month later when it published a reported 68% national employment rate among college graduates that year.[18] The statistics, delivered immediately following college graduation, has been called into question by the public because of this particular exposure.

(25) *yexu, gongzhong_zhiyi_shang_bu_neng_zhengming_gaoxiao_gao_jiuyelv*
maybe, public_doubt_yet_not_can_prove_collge_high_employment rate

*de_xujia, danshi_ci_fan_biyesheng_zi-bao_**"bei-jiuye",***
NOM_false, but_this_time_graduate_self-expose_**"PSV-employment",**
no-doubt

[17] As it turns out, this is also a tactics for a non-local graduates to obtain residence status in the area. Even though the student might not be aware of such a method, he has paid and given permission to the department to make the arrangement. The full story is here: http://edu.sina.com.cn/j/2009-07-27/1050175130.shtml

[18] This figure comes from a news report, written with statistics provided by the Bureau of Statistics: http://edu.sina.com.cn/gaokao/2009-07-09/0742209198.shtml

wu-yi_gei_huore_jiuye-lv_jiao_le_yi_peng_lengshui. [22109]
no-doubt_give_flourishing_employment-rate_pour_PFV_one_bowl_cold water.

'if public doubts cannot falsify the high employment rate of college graduates, this time, when college graduates came forward about the false (document on) employment, it definitely invalidates the flourishing employment rate'

What language users were able to generalize from the entire usage context surrounding *bei jiuye* could include the following:

(i) The meaning of "*bei* X," as used by the graduate who exposed the event, is that X (*jiuye* "employment") represents a condition of misrepresentation of his personal circumstances, specifically his employment situation. A claim about him being employment was made by the school even though he was not in fact "in employment." The claim has possibly contributed to the national statistic of 68% employment as reported by the National Bureau of Statistics. In other words, the misrepresentation of his employment was established and sustained by the school and the bureau. Contrary to what Xiao (2016) argues (see section 3.1.2), the usage here points clearly to an agent who initiates such a claim.
(ii) The function of "*bei* X," as the follow-up report in (8) shows, is that it has the power not only to expose the occurrence of such misrepresentations, but also questions the credibility of institutions that are responsible for creating the discourse. Specifically, statistics as evidence that supports such claims are under attack.
(iii) An implication about the event related to "*bei* X," is that, because X is something that can be (mis)represented by national statistics, it is a circumstance that is relevant to most people in the society. The occurrence of "*bei* X" as an event–the misrepresentation of X on certain people through a discourse–can thus occur as a large-scale event that implicates not just a handful of individuals.

(iv) An another implication about the occurrence of the event is that it took place without the knowledge of the N_{person}, a.k.a. the college graduate. And even when he became aware of the existence of such a claim, he was not able to reverse the situation, nor could he get himself uncounted from the reported 68% employment rate.

The form-meaning mappings as observed in the high-frequency instance of *bei jiuye*, can also be observed in other usage instances. I will present more usage instances within the same context of the fabrication of national statistics in the following examples. But the generalizations just given also hold true for other usage contexts in the following sections.

(26) *zai_guoqu_duo_nian_shiqi_nei*,
LOC_past_many_year_period_in

*guojia_tongji_jiguan_de_pingjun_gongzi_shuzi_de_*
country_statistics_department_NOM_average_salary_figure_NOM

zhenshixing_yizhi_shoudao_shehui_gongzhong_de_zhiyi,
credibility_constantly_receive_society_public_NOM_doubt

*bing_bei_shehui_chi_zhi_wei_shi_**"bei_zengzhang"***. [10097]
also_PSV_society_scorn_it_as_COP_ **"PSV_increase"**

'in the last few years, the credibility of the figure on national average salaries as released by the national statistics departments has been questioned by the public. The society at large condemns such (distortions of the truth) as "being increase-ed".'

The figure released in July 2009 reported that the national average monthly salary was RMB 14638 (about USD 2100). But according to a majority of Chinese netizens who commented on the issue, that figure was outrageously overstated. In other words, the reported salary was not an accurate representation on a nationwide level. A claim of an increased average national salary has been imposed on all workers in China.

This example supports my previous argument that the use of the construction "*bei* X" is seen to carry the power as a discourse–discourse in the Foucauldian sense. The use of word *chi* 'scorn' in example (26) shows that by captioning the possible fabrication of

national statistics as *bei zengzhang* 'being increase-ed,' it shows the denunciation of the fabrication of national statistics, it questions the evidentiality of national statistics at large, and it challenges the credibility of statistics departments.

There are other instances of this construction being used in the context of misrepresentation through statistics, such as *bei xingfu* 'being happy-ed,' *bei manyi* 'being satisfied-ed,' *bei xiaokang* 'being moderately-prosperous-ed' and *bei zhongchan* 'being middle-class-ed.' *Bei xingfu* was coined when a national report on "China's Happiest City" shocked residents in the top cities on the list. By the phrase, they implied that their happiness was only on paper and not experienced on an individual level. *Bei manyi* emerged in a story of how residents were instructed to give standardized answers to a national phone poll. They could only answer that they were "satisfied" so that the results would look good on paper, and the residents' real opinions did not matter. A report on middle-class households in Beijing also infuriated the public, because some who had been categorized as "middle class" based on their income vowed that their disposable income was less than sufficient to enjoy a typical middle-class lifestyle. They were thus categorized and defined by numbers, while their actual living conditions were not evaluated.

All these instances are examples of members of a society being subjected to the misrepresentation of an official discourse–a report about their livelihood based on national statistics. The instances in the X slot in this usage context, *jiuye* 'employment,' *zengzhang* 'to increase,' *xingfu* 'happiness (to be happy),' *manyi* 'to be satisfied,' *xiaokang* 'to be moderately prosperous,' *zhongchan* 'middle class,' constitute the key indexes that measure the success and prosperity of a society, including the number of college graduates having found jobs, the amount of pay increase achieved in the last year, the number of residents feeling happy with their lives and satisfied with the work government has done in their communities, and the country's progress towards a "moderately prosperous society" measured by the financial well-being of families. The reports generated on the basis of statistics create a picture of prosperity, and they want to constrain members of the society not only to a discourse of prosperity, but also to a discourse of happiness measured by only a fraction of the overall living conditions: economic achievements, employment rate, income level, etc. At the same time, the

construction of these reports and statistics attempt to subsume unique individuals into a discourse of statistical representation. Even if the statistics and the reports based on them are not exaggeratedly fabricated, the actual living conditions of individuals have been evened out by a number. The gap between their conditions and the reported overall condition is exactly what triggers the creation of the expression, and what enables it to have a ripple effect to reach the mass.

3.3.2. Misrepresentation by News Media

The usage instances above are mostly found in online news reports. The instances, as used in this context have the function of exposing fabricated statistics is achieved through the news media as a channel of information, which cites the populous as its source of information and disseminates the information to the wider public. News media has a strong influence in the modern era because we rely more and more on information, and the media controls the flow of that information. By taking control of a discourse, the media gains the power to stand up against political entities to expose the fabricated statistics they announce; in the meantime, such power is also susceptible to abuse. In this section, I will showcase instances of the *bei* construction that are related to the manipulation of discourse by news media. In this context, news media is both the creator and imposer of misrepresentation, and possibly the channel where the resistance of such misrepresentation is publicized and circulated.

Among the higher-frequency tokens identified in Tables 6 and 8, *bei qushi* 'being pass-away-ed' and *bei lihun* 'being divorce-ed' are examples of such usage contexts. The two expressions refer to the spreading of sensational news about the condition of celebrities by the news media. They can be, for instance, about a renowned author secretly passing away or a celebrity couple's sudden divorce. The writing of rumored sensational news about celebrities is nothing new in the media industry. However, more and more recently, these kinds of rumors have been taken more seriously by the public. After 'being pass-away-ed' twenty times, the *wu-xia* author Jin Yong still must personally debunk fake news to stop people from writing eulogies to remember him.

The responsibility of the news media was initially the truthful reporting of current events. Such an assumption has given news media the power to construct a discourse that is taken as reality. In a way, whether anything the media reported actually happened is

not important. What is important is that the topic has come into being through their discourse; it is shaped and constructed however the media desires it. Even when the topic is one with volition and a voice, for instance, a person, its will is totally submerged in the institutionalized news media, and its voice consumed in the institutionalized discourse. When confronted with the news media, a single individual has no power to construct a discourse against that of the media. On the contrary, news workers, even as individuals, can acquire the power of discourse through affiliation with the institution. A notable event about one such news worker for the official news media of China led to the creation of the term *bei daibiao* 'being represent-ed.' In November 2010, on the 2010 G20 Summit news conference, a famous Chinese media person, Rui, stole the mic from Korean reporters and insisted on asking President Obama of the U.S. a question. He made multiple attempts to "represent" China, Korea, and Asia, saying "I think I get to represent the entirety of Asia," and "will my Korean friends allow me to ask a question on your behalf?"[19] The Chinese netizens joked that by this act, not only were the Chinese *bei daibiao* 'being represent-ed' by the reporter, but the Koreans and the entire Asian population were also *bei daibiao* 'being represent-ed.'

Jokes aside, for the audience, the relationship of representing does not stand because they were well aware that Rui was not given the right to represent Korea and Asia by these parties. The expression *bei daibiao* 'been represent-ed' in this context is more of a ridicule of his attempt to falsely represent the region by creating such a discourse. However, was the audience equally certain from what he spoke and how he acted that he was not representing China and the Chinese people at large? Perhaps they were unsure. In this context, by being at the news conference and attending as a representative of the Chinese media, the news reporter assumed the role of representative of China and the Chinese. Whether the Chinese people actually elected him as a representative does not matter. Whether the public was against such representing does not matter. The relationship of representing was established the moment the reporter announced that he would like to "represent" China to ask a few questions–an illocutionary act in the Austinian sense. What is imposed is not only a claim of

[19] To watch the video online: https://youtu.be/i_nEZzdyvec?t=42s

representation, but also a relationship of representation. In this way, *bei daibiao* is possibly also an instance of the IMPOSITION construction.

Following the occurrence of the event, the expression of *bei daibiao* 'being represent-ed' came to represent a condition that is generalizable. In one incident, a policy maker cited a positive poll result of 67% and suggested a reform of certain capacities towards Chinese characters. In response to this, Chinese netizens declared "our opinions have 'been represent-ed.'"[20] The use of the term not only documents another occurrence of misrepresentation, but also adds to the general distaste of the public towards such misrepresentation.

In this usage context, the general public are largely vulnerable to the hegemonic discourse that is created by and circulated within the news media. Similar to the misrepresentation by a discourse of national statistics, the general public are also likely to be misrepresented by a "reality" created by the discourse of the news media. For example, when the media declares that someone is deceased, or divorced, or expresses certain opinions, it is likely these reports are taken as facts rather than being examined critically by the public. It is not that the general public does not want to combat those received distortions, but the institutionalization of discourse makes any extensive critical engagement utterly futile. "*Bei* X," with its resistant impulse, is at best a protest against the tyrannous act that seeks to subsume the general public into an institutionalized discourse rather than an endeavor to subvert the *status quo*.

3.3.3. Misrepresentation of Official Statement

In this section I will present another usage context of "*bei* X" that points to the hegemonic discourse of official statements. One of the most notable instance of the subconstruction is *bei zisha* 'being commit-suicide-ed,' which can be found in example (22) in section 3.1.2. In the event this particular expression is associated with, the phrase *zisha* 'to commit suicide' appeared in a forensic report that identified the death of an inmate in a prison hospital as "suicide." The person had previously been involved in exposing corruption that would jeopardize the political career of a government official. His sudden and bewildering "suicide" thus raised suspicion, and the public began to

[20] To read the news article: http://news.xinhuanet.com/politics/2009-08/23/content_11929031.htm

question the credibility of such a claim, urging the government to look into the real cause of death. Evidence eventually pointed to a corrupted government official who had arranged the homicide to remove the witness. The seemingly justified claim of *zisha* 'to commit suicide' is a rhetorical device to exonerate oneself from the responsibility of being held accountable for violence. By using the word *zi-sha,* literally meaning 'self-killing' or 'to kill oneself,' the responsible agents found a rhetorical excuse to exonerate themselves of the victim's death in their jurisdiction. *Bei zisha* 'being commit-suicide-ed,' stemming from this instance, captures the entire mentality behind the meticulously staged homicide and construes the innocent person as a victim of political violence.

The morpheme *zi* 'self' in *zisha* 'to commit suicide' is not an idiosyncratic one in the vocabulary of Modern Chinese. The morpheme is typically associated with free will, volition, independence as found in autonomous acts, or having the freedom to do something according to one's desire. By labeling an event as self-initiated and self-inflicted, the responsible agents blame the victim for the calamities and exempt themselves from the responsibility of the consequences. In general, this kind of wording is frequently found in the discourse of official statements that intend to abdicate responsibilities.

Ziyuan '(to act) voluntarily' is another high-frequency word in the phrase "*bei* X." It also contains the morpheme *zi* and expresses a sense of volition and autonomy that is similar to *zisha*. The word *ziyuan* 'voluntarily' is found in the official statement of a public school when it was accused of arbitrarily charging fees to newly admitted students in a report called "Teachers' Day Bonus." In response to the accusation, the school issued a statement and claimed that the parents had "voluntarily" paid money in support of the faculty. However, "voluntarily" is the exact opposite to the actual condition, which is "compulsorily;" *ziyuan* (as appeared in the statement) is only a rhetorical device. Thus *bei ziyuan* is a direct response to the fictional and self-contradictory *ziyuan*, meaning 'being forced.' Following this event, *bei ziyuan* is extended from this specific case and finds its way into much larger contexts. Yet instead of enjoying a prevalence across contexts, its occurrence is predominantly found in discourse about unregulated fees in the educational sector. For instance:

(27) *xuesheng_bei_qiangpo_ziyuan_goumai_jiaofushu*
student_PSV_force_voluntarily_purchase_supplementary materials

*hai_you_duoshao_**bei_ziyuan*** [60372]
still_have_how many_**PSV_voluntarily**
"Students are forced to 'voluntarily' purchase supplementary materials. How many more (cases of) 'being voluntarily-ed' will happen?"

The compulsory nature of the miscellaneous educational expenses is verbally revealed through the word *qiangpo* 'force' in this example, which also points out that the case of *bei ziyuan* is not an isolated occurrence. The development of the *bei ziyuan* construction attests to the persistence of the discourse of *ziyuan* as well as the protests against it.

Juankuan 'to donate money,' as it has appeared in the *bei* construction, does not have the morpheme *zi*, but it retains the same sense of forced consent. *Juankuan* can be translated as 'donation,' but in this context it means 'forced donation,' an oxymoron. The expression comes from a case about forced donation that happened after the 2005 Wenchuan earthquake. In some state departments, civil servants got a paycheck that was less than the usual amount, and the decrease of payment was explained as humanitarian aid to the earthquake survivors. This is not to say that the civil servants did not want to donate voluntarily to the cause, but the institutionalized and manufactured consent led to frustration.

In all these words, there is a sense of acting voluntarily–either it is represented by *zi* or by a word that has inherent volition like "donation." Constrastingly, as a result of institutional violence, the volition of these words goes to the opposite of itself, and becomes forced, involuntary and provoking frustration. The phrase itself is oxymoronic. In this context, "*bei* X" is thus a misrepresentation of the grammatical subject's volition and will.

3.3.4. Misrepresentation by Law Enforcement Accusation

In this section, I will identify the last usage context of "*bei* X" within the ongoing discussion of misrepresentation by institutionalized discourse. In this case, the claim of X has the power to bring negative impact on a person's life.

Bei jingshenbing where *jingshenbing* means 'mental illness' is a high-frequency token of the "*bei* X" construction. *Bei jingshenbing* refers to the practice of alleging mental illness in a healthy person.

(28) *zhengchang_xin_bei_shuo_cheng_jingshen-bing* [38092]
normal_mind_PSV_say_as_mental-illness
'a healthy-minded person is identified as insane'

It is associated with multiple incidents in which a mentally-healthy person was forcibly detained in a mental health facility due to a witness's false accusation. The loosely defined category of *jingshenbing* can potentially render anybody a suspect. According to existing laws in China, a mental health facility takes the description of a single family member/friend as sufficient evidence to compulsorily admit a patient. Both economic interest and political violence can be the culprit behind the charges. Here, the person's mental health is not the real concern. Instead the spurious claim is used to detain political enemies and economic competitors. In one instance, a netizen was taken to the police station and later sent to a mental hospital by the police because s/he took a few pictures of people petitioning the government.[21] In another instance, a mother was put away in a mental facility by her daughter. While she was in the facility, the mother's assets were secretly transferred by her son. In some extreme cases, the involuntary commitment to an institution can turn into a prolonged torture of six years:

(29) *yi_wei_zhengchang_ren_* ***"bei_jingshen-bing"*** *chang_da_liu_nian_ban*
one_CL_normal_man_ "PSV_mental-illness" long_reach_six_year_half
[48023]
'a normal person was mentalized for as long as six years'

In terms of misrepresentation by accusation, *bei jiujia* 'being DUI-ed,' is used in the same vein. It means that traffic police wrongfully allege a charge of Driving Under the Influence (DUI). Such instances are mainly due to hypersensitive and inaccurate

[21] The central government of China installed an administrative system that allows civilians to legally petition the government if their problems cannot be resolved by local governments. This creates a pressure on the local government, resulting in their interceptions of such petitioning acts through various means, including illegal imprisonment.

breathalyzers (devices for breath alcohol tests), which could yield positive results even when there is no alcohol consumption.[22] Based on these tests, drivers are identified as DUI and given penalties even though they are not driving under the influence of alcohol.

In both cases, individuals who are *bei jingshenbing* 'being mental-illness-ed' and *bei jiujia* 'being DUI-ed' are given a label that is inconsistent with the actual circumstances and that personally affects their lives. Persons identified as mentally ill were deprived of physical freedom and put through unjust and undeserved treatments. The unfortunate drivers who were wrongly charged with DUIs were given tickets, penalties, and possible short-term prison sentences.

Bei jingshenbing 'being mental-illness-ed' and *bei jiujia* 'being DUI-ed' are both examples of the usage of [N_{person} "*bei* X"] that not only identifies the wrongful attribution of X as associated with the subject N_{person}, but also implies the severe consequences of such misrepresentation in the form of wrongful charges. The attribution of X has a visible effect on the person, because the claims were made by law enforcement who have executive power. The pragmatic function of these two phrases, besides revealing the falsity of the claims, acts as a denunciation of the authorities who have power in the discourse, but who abuse such power.

3.3.5. The MISREPRESENTATION Construction in a Network of Constructions

In the previous sections, I have examined four different usage contexts with high-frequency instances of what I term the MISREPRESENTATION construction. It is an important subconstruction of the construction [N_{person} "*bei* X"] because its instantiations constitute exemplars that give rise to the generalizations about the form-meaning mappings of [N_{person} "*bei* X"]. The MISREPRESENTATION construction has a meaning of identification, where the grammatical subject has been identified to be an X / to have X / to have the quality of X / to have done X – or broadly speaking, to be associated with X. Across these contexts, the identification is made within an institutionalized discourse—for instance, reports of national statistics, news articles, official statements and law enforcement accusations–that misrepresents the grammatical subject N_{person}. A

[22] It was known that certain foods and oral products contain ingredients that produce a false positive. http://auto.sohu.com/20160405/n443392217.shtml

claim that misrepresents the N_{person} is imposed without consent, and in certain cases, the misrepresentation can also bring about negative impact.

As a construction, [N_{person} "*bei* X"] is situated in a network of constructions with formal similarities but different functions. The MISREPRESENTATION construction is thus connected to the *bei* passive in various ways. First, as indicated above, it inherits the adversative meaning from the canonical usage of the *bei* passive, in terms of the meaning of imposition and negative impact. Second, I suspect that the meaning of identification might have come from the more neutralized *bei* passive found in the formal register of Modern Chinese. In (20), I have shown the constructional form of the *bei* passive is [N_{patien} *bei* (N_{agent}) VP] where *bei* can be directly followed by an N_{agent} or an VP (predicate). A quick search of the TenTen corpus shows that high-frequency verbs immediately following *bei* largely belong to the lexical category of "recognition of status as warranted or considered by the authorities and people" (see Table 8).[23] This is consistent with findings in Jing-Schmidt and Jing (2011) in which "acts of designating by authoritative or popular appraisal and selection" is determined to be one of the three lexical categories for the high-frequency verbal collocates of *bei*. For instance, *pingwei* 'select as,' *chengwei* 'address/term as,' *renwei* 'regard as,' *yuwei* 'praise as,' *shiwei* 'consider as,' *shouyu* 'endow with,' *liewei* 'rank as,' *lieru* 'rank among.' The morpheme *wei* (roughly meaning 'as') which re-occurs among the high-frequency verbs in "*bei* + VERB" also helps construe "identification" and "recognition" as important constructional meanings of the *bei* passive. At the same time, a corpus search of high-frequency nouns immediately following *bei* finds that these nouns correspond to the lexical category of designation and selection.[24] The high-frequency semantic agent of the *bei* passive include the populous, such as *renmen* 'the mass/people,' *ren* 'people,' *bieren* 'other people,' *guoji* 'the international community,' *shehui* 'the society,' *shichang* 'the market,' *shijie* 'the world,' but also authoritative entities such as *guojia* 'the country,' *zhongguo* 'China,' *meiguo* 'the USA,' SCI 'the Science Citation Index' (an authority in the field of academic

[23] CQL: [word="被"] [tag="V.*"]

[24] CQL: [word="被"] [tag="N.*"]

Table 8. High-frequency immediate collocates of *bei* in TenTen Corpus

bei + VERB			*bei* + NOUN		
VERB	definition	token frq	NOUN	definition	token frq
pingwei	'to select as'	82230	*ren*	'people'	30571
chengwei	'to call as'	39377	*guojia*	'country'	13050
renwei	'to consider as'	26645	*renmen*	'the mass'	10732
yuwei	'to praise as'	22683	*zhongguo*	'China'	9181
shiwei	'to regard as'	15820	~~*shangsuren*~~		
shouyu	'to award'	15512	~~*shenqingren*~~		
liewei	'to rank as'	14328	*meiguo*	'USA'	5083
~~*gao*~~			*bieren*	'other people'	4729
lieru	'to include into'	11669	*guoji*	'the world'	4370
faxian	'to discover'	8528	*SCI*	'SCI (index)'	4321
yong	'to use'	8475	~~*diaochazhe*~~		
queding	'to confirm'	8146	*meiti*	'the media'	3950
~~*po*~~			~~*Shenji*~~		
luqu	'to recruit'	8001	*Shehui*	'society'	3784
zhengdi	'to expropriate land'	7219	*Sheng*	'province'	3425
taotai	'to eliminate'	7172	*xuexiao*	'school'	3194
rending	'to give credential to'	7069	*jiaoyubu*	'dpt. of Education'	2949
pizhun	'to approve'	6603	*guowuyuan*	'the state council'	2924
~~*dao*~~			*shichang*	'the market'	2873
xuanwei	'to elect as'	5728	*shijie*	'the world'	2716

Note: Items with strikethroughs are words (not phrases) that were parsed mistakenly by Sketch Engine.

publications on science), *meiti* 'the media' (who controls the "truth"), *sheng* 'the provincial (government),' *xuexiao* 'the school,' *jiaoyubu* 'the department of education,' *guowuyuan* 'the state council.' All of these entities have the capability of constructing a discourse.

Among these entities, there are apparent overlaps with the semantic agents implied in the MISREPRESENTATION construction, including government departments as the source of manufactured statistics, the media as the source and filter of news stories, educational agencies that release official statements aiming at fostering consent, and government departments who have executive power to criminalize potential subjects.

Institutions, with their power to create and interpret discourse, can become the source of institutional violence.

The meaning association between the MISREPRESENATION construction and the neutralized *bei* is possible because the the collocates of both constructions are rooted in the formal register. The neutralized usage of the *bei* passive is more prominent in the formal register (Chappell, 2009), and in this usage, *bei* also tends to collocate with newly-coined words associated with modern concepts, which are usually disyllabic (Li & Thompson, 1981, p. 497). Similarly, the X words of the MISREPRESENATION construction have all come from discourses using a formal register: "the employment rate," the level of "satisfaction," the number of "middle class" families, "divorce," "voluntarily," "suicide." Both "mental illness" and "DUI" are charges that can decide the fate of a person. They are also predominantly disyllabic (none of them is monosyllabic, and thus they are less likely to be associated with the colloquial register). The generic resemblance of collocates in the two *bei* constructions constitutes a formal resemblance, allowing the generalization of meaning to be extended into the new construction of *bei.* Using words from an institutionalized discourse has given rise to the meaning association of "identification" that dominates the neutralized *bei* passive construction, and with it the implication of semantic agent—the "authoritative" and the "populous" (Jing-Schmidt & Jing, 2011), but most importantly the former.

However, while the root of X in "*bei* X" is on the more formal end of discourse types, the usage of the construction is largely informal. The MISREPRESENATION construction is used by the public, found on media and is particularly popular on the Internet. In this sense, it also connects to the canonical usage of the adversative *bei* passive found in colloquial language. Rather than being neutral, the MISREPRESENATION construction has an evident adversative meaning. The meaning of the construction is "identification," but the pragmatic inference that emerges from context is the negation, denunciation, and refusal of such identification. Frequently, when the *bei* passive has an adversative reading, it is because the action is deemed undesirable. For the MISREPRESENATION construction, the claim of X is also deemed undesirable, not only because it is imposed against the will of the person, but also because such identification can lead to dire consequences.

When viewed in this light, having connections to both the neutralized *bei* passive and the canonically adversative *bei* passive, the MISREPRESENATION construction has the following characteristics.

First and foremost, it has an adversative meaning relating to the canonical usage of the *bei* passive that is reflected in many aspects. The construction has a great contrast of agentivity and volition between the semantic patient and the semantic agent. In such a contrast between the low agentivity grammatical subject/semantic patient and high agentivity semantic agent that is suppressed syntactically, the former is completely subject to the control of the latter, and the latter can rather freely exert a force onto the former without much resistance. In this particularly construction, it means a claim of X can be imposed on the grammatical subject without its consent, sometimes even against its will; it means that even when the grammatical subject becomes aware of the claim as a misrepresentation, there is no revision to it. When putting together the fact that the grammatical subjects in my sample are all individuals, and the misrepresentation is established by some kind of institutionalized discourse, it makes the observation above even more noteworthy. The contrast of agentivity in the *bei* passive has specifically transposed onto two types of entities in this subconstruction. This is a generalization that is peculiar to this construction, and I presume it will be resistant to the neutralization process that the canonical *bei* has been going through.

Another peculiarity about this construction, when compared to the *bei* passive, is that instead of describing an action being imposed on the grammatical subject, this construction specifically deals with a claim of identification being imposed, or a discourse that intends to misrepresent the grammatical subject. The word X is a quote from the claim, and it enters into the construction as a metonym, representing the entire discourse event of misrepresentation imposed by an institutionalized discourse. The wrapping double quotation marks as a marked structural feature of the construction indicates such a source, but it also functions to mark the phrase as a neologism, a coined expression. In the corpus, I have found that some of the high frequency instances (such as *bei jiuye*, *bei xingfu*) have lost the quotation marks in their usage because their meaning and function has been so entrenched, but for other less entrenched items, especially those

created on the spot, the use of quotation marks is still essential in indicating its source and its status as a neologism.

These are the two most important idiosyncrasies of the MISREPRESENTATION construction compared to the canonical *bei* passive. The MISREPRESENTATION construction, as a subconstruction of the new *bei* construction [N_{person} "*bei* X"] is specifically concerned with the dynamics between participants in an institutionalized discourse. On one end, there is a hegemonic power that constructs and controls the discourse, and on the other end, there are the addressees who do not participate in the construction of discourse but who are shaped by the discourse.

What further distinguishes this construction from the canonical *bei* passive which describes a one-time and finished event, is that the MISREPRESENTATION construction portrays the occurrence of imposed identification as multipliable, which has the potential to permeate the given society. The affected entity is thus not singular but collective, construing the undesirably-imposed identification as a social phenomenon.

On the pragmatic level, this construction indicates the language user's resistance to such a non-negotiated misrepresentation. It is out of the scope of this chapter to discuss whether such a resistance yields any positive results from the language user's side, but the appearance of such a usage already expresses a deeply-seated desire to not be coerced, to not be forced, and to not be interpolated with violence.

The MISREPRESENATION construction constitutes the foundation of the new *bei* construction. The IMPOSITION construction, in my opinion, is an extension based on the former. It shares a lot of fundamental generalizations with the MISREPRESENTATION construction, such as the adversity, the multipliability of the event, and the pragmatic function of resistance. The only difference, I presume, is that the former construction concerns the discourse as the arena of power dynamics, while the later has a wider scope of concern.

3.4. The IMPOSITION Construction

The most apparent distinction of the IMPOSITION construction compared to the MISREPRESENTATION construction is the imposition of an event onto a person instead of an identification. In fact, an identification of X is also a kind of event—a

speech event. In this sense, I consider the IMPOSITION construction to have a wider scope of concern because it deals with general types of event. In terms of the operational mechanism of X, in the former construction X is a keyword in the identification and a quote from the claim. In the IMPOSITION construction, X is a phrase that can represent the whole event. In other words, it can be considered a metonymy of the whole event. By metonymy here, I am referring to a metonymic structure which uses a high-frequency phrase to generalize the whole event. In such a working subtype, the phrase is metonymized to the generalizable event. I will now give a few examples to further explain.

The phrase *bei shanglou* 'be move-into-a-high-rise-ed' refers to an influential and widespread phenomenon in contemporary China. Due to the rapid urbanization happening all over China, land becomes a space of power assertions. The single-story building is giving way to high-rises, and residents who used to live in these buildings are being forced to move into high-rises with compensation/subsidies from the government. The phrase *bei shanglou* is a direct response to the phenomenon. *Shanglou* literally means 'to go upstairs.' In this usage of the construction, the residents were unwilling to move into high-rises but were forced to against their will. The villagers were forced to give up not only the ownership of their land for relatively small compensation but also their traditional lifestyle.

(30) *sichuan_de_nongmin_**"bei_shanglou"**_si_nian*,
Sichuan_NOM_peasant_"**PSV_up-building"**_four_year

*ning_shui_yangjuan_bu_zhu*_loufang [59289]
rather_sheep-pen_not_live_building

'the peasant land owners in Sichuan have been 'move-into-high-rise-ed' for four years, but they would rather live in their sheep pen than in the buildings'

In this rather emotional commentary, the villagers would rather live in a much dis-privileged and uncomfortable location, *yangjuan* 'the sheep pen,' than live in high-rises that are alien to them. Here, *shanglou* is being used metonymically to indicate the whole imposed event. It corresponds to a larger social context where urbanization,

development and privatization of land has been imposed upon the common people without much room for debate.

Bei gaotie 'being high-speed-train-ed' is another result of the economic development happening in contemporary China. For the past few decades, China has updated and expanded its trains and railway lines tremendously. *Gaotie*, meaning 'high-speed train,' is the fastest and most comfortable train between any two locations in China. However, that comfort comes with a higher cost for passengers. In many areas, the "slower" trains are no longer in use as a result of such updates. Many passengers who are not willing to pay a higher price to take a high-speed train have no option to take a slower and less expensive train. That is, the passengers are deprived of other options—they are forced to take the high-speed train, and thus "*bei gaotie*."

Another extension deals with an emerging concept, the credit card. The credit system is a relatively new phenomenon, developed only after the Cultural Revolution and marking the transition from a planned economy to a market-oriented economy. In recent years, commercial banks have enjoyed an exponential expansion and development. Rapid economic development also generates fast-changing fluctuations of wealth. Credit cards, not being well regulated, seem to enter every household, sometimes without the owner's knowledge. The over-issuing of credit cards and the lack of effective supervising can lead to credit card fraud. *Bei xinyongka* 'being credit-card-ed' here means one is issued a credit card without being informed and sometimes even without an application, and one only finds out many months after the issue date. The relevance of imposition is quite apparent here. In *bei shanglou*, a policy to move peasants away from their land and into buildings is imposed; in *bei gaotie*, the monopoly of high-speed trains in the railway transportation is imposed; in *bei xinyongka*, a credit card without rightful ownership is imposed.

Conceptually speaking, these instances of the IMPOSITION construction are very similar to the canonical adversative *bei* passive. They both describe how an action or event is initiated by another, but affects the grammatical subject regardless of the patient's desire (if it has one), producing a negative effect. However, the two constructions are also distinct. While the *bei* passive requires a VP and refers to a one-time imposition of an action, the IMPOSITION *bei* uses a metonymic X to represent the

imposition of something bigger. X can initially be a metonym of an event. For instance, *Shanglou, gaotie* and *xinyongka*, are all bracketed from the event, and at the same time represent three different and prevalent social events. But the metonymy scheme does not stop at this level. The phrase also refers to a social phenomenon that is larger than the singular case where the phrase comes from. For example, *bei shanglou* not only addresses the specific event of peasant land owners moving into high-rises, it also hints at their lifestyle changes and the bigger context of rural reform. *Bei gaotie* concerns not only the monopoly of high-speed trains, but also brings out the inaccessibility of lower-speed trains and an overall increase in the cost of transportation. *Bei xinyongka* presents more than credit card fraud, it also underlines the lack of supervision in the banking industry and the peril of identity theft. The chain of metonymization can go on and on and on. In this sense, these phrases are metonyms of these three contextualized social phenomena. The IMPOSITION construction is concerned with various types of imposition that originate from the social structure in general.

This subconstruction's pragmatic function is quite similar to that of the MISREPRESENTATION *bei,* which is to allow user to comment on social conflicts, express discontent, and satirize absurdity in a contemporary China that is undergoing urbanization, transportation updates and economic reforms of all sorts. In fact, I suspect that the IMPOSITION construction might have emerged from an implication in certain usage instances of the MISREPRESENTATION construction. For instance, the word *xiaofei* 'consume' in the expression *bei xiaofei* 'being consume-ed' can be looked at in both ways. In the event associated with the term, a water company of a city in China set the monthly water usage of a household at a minimal of five tons. This means that when actual usage is below the minimum level, consumers will still be charged at the volume of five tons. Here, *xiaofei* is an imposed claim because the act of consumption is established through the water bill issued by the City Water Bureau–an institution that is not to be negotiated with. Also it is a metonymic representation of the event: forced consumption because of forced bill payment. Thus the new *bei* construction extends from the specific emphasis on imposition of a discourse to the more comprehensive coverage of imposition of social phenomena at large.

3.5. The New *Bei* Construction

Contrary to the theory that the new *bei* construction is a revival of the adversative usage of *bei* passive as suggested by previous scholarship (Wang, 2011), I propose that the new *bei* construction has emerged in connection with the *bei* passive. As a subconstruction of the *bei* passive, it inherits the meaning of adversity from the *bei* passive, in the sense that something has been imposed onto the person against his/her will, something which has a negative impact on the person. The construction has a constructional frame [N_{person} "*bei* X"], and it has developed two sub-schemas, each formed by a cluster of exemplars. By analyzing high-frequency tokens of the construction in the corpus, I showed that the two sub-schemas have distinguishable constructional meanings and operate with different mechanisms. The first subconstruction is the MISREPRESENTATION construction, where X is a misrepresentation imposed on the grammatical subject. The second subconstruction is the IMPOSITION construction, where X is a condition imposed permanently on the grammatical subject. The former is a specific type of imposition event (imposition of a discourse), while the later is a more generic type of imposition event. The constructional meaning of the new *bei* construction, as emerged in usage instances, can be summarized as follows:

(31) A condition has been imposed upon the semantic patient (and grammatical subject) N_{person}. The imposition of such a condition as represented by X is not about a particular incident that occurs to one individual, but it implies that it is a pattern of behavior and is a widespread social phenomenon imposed on collectives. Its occurrence has a negative effect on the patient, and its usage contains resistance.

The new *bei* construction, in terms of its usage context, is thus a specialization of the adversative *bei* passive, as it is used when expressing the adversative circumstances faced by humans in the wider society. In terms of formal representation, it is an expansion of the VP predicate of the *bei* passive in that it allows more grammatical categories to occur in the VP's position. Essentially, it is an extended specialization of the *bei* passive, which means that this new *bei* construction is still very much connected to the canonical *bei* passive.

3.6. Final Remarks

3.6.1. The Role of Grammatical Categories

A linguistic feature that has been stressed in previous studies, but not attended to with much elaboration in this study, is the grammatical category of X. As I briefly introduced in section 3.1.2., earlier scholarship observes that the grammatical category of X encompasses transitive verbs, intransitive verbs, adjectival verbs and nouns, which is a notable expansion from the typical transitive verb in the VP slot in the *bei* passive. Some of these studies have remarked that it is the "oddity" of such collocations that generates the new interpretation of the *bei* passive. This shift relates to the concept of constructional coercion in Construction Grammar. In constructional coercion, type shifting occurs when a lexical item that belongs to an unexpected category enters a constructional slot and is coerced into the canonical category (Michaelis, 2004). In other words, the meaning of the intransitive verb, adjectival verb or noun that enters into the slot X is converted into a transitive verb with the meaning of 'forced to (do) X.'

I disagree with this proposal. The lexical input from X interacts with the constructional meaning of "*bei* X" and its meaning is not simply "forced to do something," but also "forced to accept a permanent condition." For instance, compare example (21) and a related passive sentence (32):

(32) *haizi-men_bei_qiangpo_lian_ticao*
children-PL_PSV_force_practice_gymnastics
'the children were forced to practice gymnastics'

Example (21)'s reference of *ticao* "gymnastics" represents an entire way of existence dominated by gymnastics, including the strong stance toward not just a single event but an entire ideology of parenting, the patterned imposition of parental will upon children, the totality and tyranny of its ruthlessness, the helplessness of the affected, and their total loss of voice and self-determination. What is imposed is not a single training session of gymnastics but a lifestyle that is permeated by gymnastics.

On the one hand, the canonical grammatical category of X contributes to the generalization of the construction. For instance, it is in a highly transitive construction that the use of a canonically intransitive/adjectival verb foregrounds the presence of the other participant in an intransitive act and suggests that it is not self-initiated. The use of

adjectival verbs and nouns, lexical items without telicity and the exact opposite of the transitive causative verbs typically found in the canonical *bei* passive, extracts temporality out of the act and presents the occurrence of the event as something that is more stable, permanent and absolute. In fact, a few of the X words can be interpreted both as verbs and nouns, like *jiuye* 'employment; to be employed' or *xingfu* 'happiness; to feel happy.' While satisfying the requirement of the old constructional form of [*bei* VP], it is also possible to have the surface generalization of [*bei* NP]. X, rather than being a verb that takes an argument, is a concept, a discourse, an entire way of living that interacts with the grammatical subject by imposing itself on it. The lexical input of X establishes it as something similar to an NP, a regional concept, to use Langacker's term (1987, p. 198).

On the other hand, as generalizations of the construction emerge, the construction can coerce X with its constructional meaning. At this time, the canonical grammatical category of X does not matter because X has been constructionalized and turned into a concept, something permanent, ubiquitous. The innovation of the new *bei* construction changes the idiosyncrasies of X, from a grammatical slot that is constrained by grammatical categories, to something that is not a grammatical slot, but part of a construction. The constructionalization of the new *bei* construction, including the decategorization of the slot X, is an innovation that achieves pragmatic inferences in usage contexts.

3.6.2. Contributions of the Current Study

Because of the quantity of scholarship already produced on the new *bei* construction, I feel obliged to quickly summarize the contributions of the current study. The biggest contribution of the current study is the use of the new *bei* construction as a case study to answer the questions I posed in this dissertation concerning the role of the construction in linguistic innovation, the manner of change and the internal/external factors leading to change. In this chapter, I established that the new *bei* construction is not a random innovation from the *bei* passive, but a subconstruction of the *bei* passive, which comes into being through usage instances of two related subconstructions that have the same form: the MISREPRESENTATION construction and the IMPOSITION

construction. Specifically, the findings of the current study benefit from an exemplar-based approach in processing language data. By examining the instantiations of this construction in a corpus, especially their occurrence frequency, a few high-frequency instances emerge as exemplars of the construction, providing a conceptual orientation about the usage of the new construction for language users. The exemplars of each subconstruction also exhibit a high level of similarity which form clusters in each subconstruction. Emerging in such a fashion, the new *bei* construction develops in a catastrophic, rather than gradual, manner. The innovation of the new *bei* construction, in instantly creating a new form-meaning pair, is made possible by pragmatic inferences that emerge in specific usage contexts. This is consistent with the observation of Bybee who posits that, when a new construction develops out of old constructions, "[e]xemplar representation is necessary [. . .] it is the specific contexts in which the construction is used that add the pragmatic implications" (2013, p. 63). And because these exemplars are the driving force of linguistic innovation, I would propose that an external factor leading to the emergence of the new *bei* construction is the presence of imposition in the social structure at large and the need to expose such a condition, which is made possible by Internet as an anonymous communication platform. The language users didn't invent an entirely new construction to express their meaning, but expanded an existing construction within the system through analogy: the *bei* passive. These are the internal and external factors of linguistic innovations that I have identified in this case.

Another contribution of the current study is a rebuttal of the perception of the new *bei* construction as polysemous and illusive. As I point out, it has two subconstructions, and the meaning of a usage instance can be reliably decided based on the context and background information of the usage. While the same construction can have different meanings, an actual usage instance of these subconstructions is not polysemous because it is inferred from the context of usage.

Through analyzing the instances of actual usage of the new *bei*, I was able to form an understanding of its constructional meaning through the generalizations that emerge from usage and avoided modeling parameters to fit into the descriptions. In specific, I hope it is clear now why binary descriptors cannot fully capture the meaning of the two

subconstructions—in this case, “untrue” is not a helpful descriptor, because truth is constructed differently by different participants and in different usage contexts.

CHAPTER IV
EMERGENCE OF A DISCOURSE CONSTRUCTION: THE CASE OF *YE-SHI-ZUI-LE*

4.1. Introduction

This chapter investigates a construction that operates at the sentence/discourse level. The construction under study emerges in the genre of Internet language and was first accepted into the Chinese speech community as an "Internet neologism" through its popular use on the Chinese social media platform Weibo. Section 4.2 will discuss how the platform provides crucial mechanisms that ensure the acceleration of that emergence. For now, I will begin with an example of a Weibo blog post to form a first impression of the construction:

(33) *mingtian_you_yi_lun*_zhibo[,]
tomorrow_another_one_round_live show[,]

*caipai_dao_xianzai_hai_mei_jieshu_**ye-shi-zui-le***[.]
rehearsal_to_now_still_not_end_**also-be-drunk-PFV**[.]

🌙_🌙_🌙 _*wo_yao_shuijiao*[.] 😂_😂_😂
(emoticon)_(emoticon)_(emoticon)
I_need_sleep[.](emoticon)_(emoticon)_(emoticon)

"Another round of live shows starts tomorrow. We are still in rehearsal and it's not ending yet, [emoticon: moon] x 3 (?) ***I'm also drunk***. I need some sleep[.] [emoticon: tears of joy] x 3"

In this example, the user (author of the blog post) shares an ongoing life event, time-stamped at 23:58 (two minutes before midnight). The post is both a description of the event (with text and pictures) and a personal comment towards the event. The substance of this event is quite straightforward from information revealed through the blog post: The user will be on an important live show tomorrow, and he is in rehearsal to prepare for it; the rehearsal has lasted for a long time and now it's midnight. Of particular interest in the example is the statement *ye-shi-zui-le*, which literally means 'I'm also drunk.' However, being drunk is not a factual state of the user. What the blog post describes is a context where drinking is irrelevant—if not inappropriate. And, it turns out,

the user is a finalist in a Chinese reality television singing competition and a possible rising celebrity. It is against common sense that a person in such a position would be drinking prior to such an occasion, even more so to announce that drunkenness on a public platform. Besides, his selfie (Figure 1 below) does not show any signs of alcohol intoxication. Not just in this particular example, but in most usages of *ye-shi-zui-le* (as the readers will see in the following analysis), contextual information that justifies a drunkenness interpretation is not provided. The point is, the phrase *ye-shi-zui-le* is not used for its literal meaning here. So what is its meaning here?

Figure 1. An example of the YSZL construction from a Weibo blog post
Note: Image blurred to protect identify of the user

As the analysis unfolds, the examples will demonstrate that this expression is similar to saying "I don't know what to say," with an attempt to conceal certain attitudes. For instance, in (33), the user, by using *ye-shi-zui-le,* makes an effort to conceal his impatience towards the dragging rehearsal. Contrastingly, on a deeper emotional level, there is also the intense gratification, pride, and excitement about having the opportunity to rise to fame that he is eager to hide. This last assumption about the existence of such emotions is apparent when one considers the context of the posting. The user is under pressure to build an online celebrity persona—this is evident from pictures of the performing stage and of him dressing up like an actor. I presume that creating an

impression of fame is a more primal motivation behind the speech act of posting the blog post), and *ye-shi-zui-le* makes his true communicative intent less pronounced.

In the following section, I will argue that *ye-shi-zui-le* is a metaphorical usage of the pyscho-somatic state of alcohol intoxication that serves social and pragmatic purposes. This chapter sets out to explore how this meaning emerges, how it is learned by language users, and how such a functional phrase imbued with metaphorical meaning can quickly gain popularity and become part of the public discourse without formal instruction on its use. I hypothesize that *ye-shi-zui-le* is not learned as a stand-alone expression. Its function and meaning can be acquired because it is part of a bigger construction, a discourse construction. The discourse construction allows the meaning and function of the phrase to emerge in association with other components in the entire construction. Adopting a usage-based approach, I will examine synchronic data of its usage in a corpus and formulate generalizations that people might make when exposed to its usage. The following section will discuss how discourse grammar can intersect with Construction Grammar, and validates the concept of a discourse construction. In addition, proper understandings of the conceptual basis of *zui* 'alcohol intoxication' and the mediality of Internet social media as a connective platform are also essential to understanding the emergence of this construction, as social media possesses mechanisms that enable and accelerate the process. These topics will also be addressed in section 4.2.

4.2. The Conceptual and Technological Basis of the Emergence of *Ye-shi-zui-le*

4.2.1. The Embodied Experience and Cultural Memory of *Zui*

Metaphor is an important mechanism in language change. In research of grammaticalization, metaphorical thinking helps explain the direction the meaning change takes place (Bybee et al. 1994; Heine et al., 1991; Hopper & Traugott, 2003[1993]; Jing-Schmidt & Kapatsinski, 2012; Lakoff & Johnson, 1980; Panther, Thornburg, Barcelona, 2009; Sweetser, 1990). In this case, the phrase *ye-shi-zui-le* is a metaphorical extension of the experience of *zui* (醉) 'alcohol intoxication.' Although consuming alcohol is a common practice across cultures, alcohol intoxication is not a basic embodied experience—the wide range of psychophysiological effects induced by alcohol intoxication varies with the context of consumption, the type of alcohol, and the

mental and physical state of the users. In dealing with a metaphor that emerges from such a complex embodied experience, there are different types of stimulation and inhibition effects to consider, and I will need to pin down to one, if not several, entailments associated with a state of alcohol intoxication to clearly enunciate the meaning and function of *ye-shi-zui-le.*

In Contemporary Chinese, *zui* describes a state of alcohol intoxication. It is also a morpheme associated with the mental state of 'pleasant indulgence,' which occurs in words like *taozui* 'to be carried away by' and *chenzui* 'to indulge in.' However, neither of these meanings represent the meaning of *zui* in *ye-shi-zui-le*. *Ye-shi-zui-le* as a construction has developed a constructional meaning that is associated with another aspect of the psycho-somatic experience of alcohol intoxication. Specifically, how alcohol intoxication impairs the cognitive functions of humans is relevant in this phrase.

A large number of studies conducted in the '60s through the '80s showed that "alcohol intoxication impairs nearly every aspect of information processing," including memory and the abilities to abstract, to register cues, to encode information, to make associations, and so forth. (Peterson, Rothfleisch, Zelazo, & Pihl, 1990; see reviews of more studies in Ryan & Butters, 1983; also Steele & Josephs, 1990). These are all crucial cognitive functions in handling verbal tasks, and a few studies have found empirical evidence that supports the claim that alcohol intoxication interferes with verbal processing. In one study in which intoxicated subjects were asked to process sentences with opaque meanings, researchers determined that in a state of induced intoxication, the lack of contextual cues would greatly affect a subject's ability to interpret the sentences accurately and efficiently (Birnbaum, Hartley, Johnson, & Taylor, 1980). This finding is consistent with a model of memory storage and retrieval that emphasizes the role of contextual cues (Birnbaum & Parker, 1977; Tulving & Thomson, 1973; see a similar proposal in Pernanen, 1976). In other words, this study implies that, in a state of alcohol intoxication, one's ability to interpret ambiguous meaning could be greatly impaired due to the difficulty of constructing an appropriate schema to retrieve and interpret sentences. In a similar study, it was shown that verbal tasks were most affected by the deficit of recognition memory induced by alcohol intoxication; its effect on perceptual priming and

picture recognition tasks was almost minimal (Söderlund, Parker, Schwartz, & Tulving, 2005).

As these studies showed, *zui*, or alcohol intoxication, has a conspicuous effect on the cognitive functions of humans in ways that particularly impair language processing. *Ye-shi-zui-le*, a proclamation of "alcohol intoxication" as a response to a verbal description, is thus used metaphorically to account for one's inability to process the information. The actual experience that one undergoes when enunciating *ye-shi-zui-le* is comparable to the cognitive impairment one experiences in a state of alcohol intoxication.

However, the embodied account of alcohol intoxication does not sufficiently explain the construction's pervasive usage—after all, not every user of the phrase *ye-shi-zui-le* has a personal memory of acute alcohol intoxication in which his or her ability to process verbal information was impaired. The metaphorical usage of the phrase is also rooted in the shared cultural memory of *zui* that is deeply embedded in Chinese speakers. In the literary tradition of China, alcohol is a reoccurring motif, and *zui* is sought after by literati for creative and spiritual purposes. In Han, alcohol is linked to the "*carpe diem* theme," and is seen as "an intensifier of sensations amid a deep sense of existential transience and uncertainty" (Kwong, 2013, p. 3). In the Wei-Jin era, drinking "became a tragic way of anaesthetizing the spirit and the senses for sheer survival" (Kwong, 2013, p. 4). For the poet Tao Yuanming, drinking is linked to "spiritual transcendence" (Kwong, 2013, p. 5) that allows him to escape from the world and be unaware of the self. One of the Seven Sages of the Bamboo Grove, Liu Ling, "sought *zui* as a symbol of escape from the miseries of the world and from one's personal emotions" (Liu, 1966, p. 59). The same sentiment is revealed in the wine poet Li Bai's works, eulogizing alcohol because it can "dissolve" one's sorrow (Kwong, 2013, p. 13; see also Smith, 2012, p. 25). For Li, the more important role is that alcohol is a source of inspiration, the "muse," as it is for other literati (see also Benn, 2005, p. 222). J. Liu remarks that it is important to understand that the state alcohol induces in the poets, *zui*, "does not mean quite the same thing as 'drunk,' 'intoxicated,' or 'inebriated,' but rather means being mentally carried away from one's normal preoccupations," and the word *zui*, he believes, should be aptly translated into

"rapt with wine" (p. 59). The state of *zui* is believed to have the power to inhibit sorrow and the conscious self, which makes way for creativity.

The cultural memory of alcohol use to achieve the state of *zui* focuses on how alcohol use affects the body and infuses the mind with alternative ways to perceive the world. The introduction of this other experience temporarily replaces the everyday bodily experience and ways of thinking. This alternative bodily experience, as shown above, is mainly the impairment of cognition. In the context of Chinese, the culture establishes and reinforces the memory of *zui* as a state that entails the impairment of cognition. It could lead to verbal suppression or verbal expression—either of which is not necessarily an accurate and faithful iteration of the mind.

Ye-shi-zui-le thus draws analogy from both the embodied and cultural experience of *zui*. *Zui* is the key to this phrase as it triggers an experience of impaired cognition from alcohol intoxication. This use of this term can be extended to refer to a similar state when one is unable or unwilling to opine and to take responsibility for one's verbal behavior—only it is not induced by alcohol, but induced internally, as an excuse.

4.2.2. Connectivity-enabled Linguistic Change: Internet Neologism and Internet Memes

Ye-shi-zui-le is one of the hundreds of neologisms that emerge in Chinese every year. Following their emergence, they enter into our everyday speech almost instantaneously, thanks to their propagation on the Internet. The Internet has transformed the way language changes, especially with regard to speed. Whereas in the past it could have taken hundreds of years for a new expression to become stabilized, this can now be achieved overnight in the Internet era (Crystal, 2011, p. 57). The essential ingredient the Internet provides is frequency of usage. As the Internet penetrates every aspect and every moment of our lives, a single message can reach almost anyone immediately. When we connect to the Internet to browse online news, to use social media, to chat with friends, or to email a colleague, we also instantly pass on any information in the form of language. This connectivity, supported by connective infrastructure, has achieved large-scale participation, and it has made the transimission of our verbal communication, much more efficient. When a greater number of people are more rapidly exposed to a certain linguistic form, generalizations about the form also emerge more quickly. In this way, the

Internet has sped up the process whereby a new way of saying things gains recognition. And of all media, social media, because of its multilateral channels of communication, becomes the most productive breeding ground of Internet neologisms.

I am not specifically setting this in a Chinese context. On one of the world's largest social media platforms Twitter, waves of memes sweep across the community at a breakneck pace. They quickly become viral and then quickly go out of fashion. The word "Internet meme" is adopted from Zappavigna (2012) in her discussion of the phrasal template that is characterized by the viral spread of formulaic language based on the template on the Internet. On Twitter, for example, the phrasal template is usually embedded in a tweet and wrapped by hashtags, which allow it to be searchable. A highly viral meme that was observed in Zappavignia's Twitter corpus is the "Imma let your finish" meme which follows Kanye West's comment that "Beyoncé is the best of all time" on Taylor Swift's acceptance of a music award in 2009. The meme evolved from his comment and acquired the following template (p. 122):

(34) "Look/Yo [vocative], I'm really happy for you. Imma let you finish. But [noun] is the [superlative] of all time!"

Based on the operational mechanisms of Twitter, Zappavigna believes that this type of expression with Internet memes "seems to center on the way that humor is being used afflictively to connect with an imagined community of users with similar attitudes. We find the template amusing and bond around the value being expressed while also experiencing a Bakhtinian intertextual pleasure when we are able to pick out any popular culture references made" (p. 117).

The Internet neologism that is taken up in this chapter also finds its earliest usage on social media—specifically, on the Chinese social media platform Weibo. It could technically be considered an Internet meme, as it demonstrates characteristics of Internet memes that Zappavigna argues for, evoking connectivity in its popularity and usage. However, the discussion in this chapter will be different from that of the Internet neologisms Zappavigna covers; in particular, I will argue that instead of being ephemeral, this Internet neologism gains the status of a linguistic construction which comes to

develop a stable form–meaning mapping that becomes part of the linguistic knowledge of the users.

Its emergence and constructionalization has its roots in the linguistic system and its particularity in the social context. Social media as a platform will be an important angle when I approach this latter consideration.

4.2.3. Weibo as an Open Public Space

As the birthplace of *ye-shi-zui-le*, Weibo is also the source of the data used in the current study. It is thus crucial to understand the functional mechanism of the platform, at least to the extent that is relevant to the current study.

Weibo (微博) is the Chinese term for a microblogging service. The service offered by Sina, or Sina Weibo, has undoubtedly dominated the Chinese market, garnering 26.1 million monthly active users (MAU)—a number reported in the 2016 first-quarter report from the company.[25] The number of MAU has increased by 32% from last year at the same time, exhibiting a winner-take-all effect. The word *weibo* has almost become the synonym of the microblogging service offered by Sina. For Weibo users, connection to Sina Weibo would mean they are directly connected to the largest online community.

Sina Weibo integrates features that can be found on Facebook and Twitter. According to a service description published by the company, the main functionalities of Sina Weibo is shown in Table 9 (only accessible for registered users).[26]

In the introduction earlier, and in the analysis below, we also encounter the use of emoticons. On the Weibo platform, only emoticons developed by the platform are supported. Thus, it limits emoticon use to a collection that is possibly smaller than that available on a smartphone. The default set that appears in the emoticon menu is presented in Figure 2.

[25] The report: http://weibo.com/ttarticle/p/show?id=2309403974264192658759

[26] http://tw.weibo.com/about/about_qa.html

Table 9. Functionalities of Weibo

Functionality	Description	Translation from Sina's service description
Posting	Users can post a microblog post (*weibo*) up to 140 Chinese characters to their account	
Reposting		'Repost an interesting *weibo* to share it with others.'
Following	Users can subscribe to updates from other users. *Weibo* posted by the subscribed users and the users themselves will appear in chronological order on their wall. By following other users, users become their *fensi* (fans), a transliteration that refers to the singular of 'fans'.	Version 1: 'Whose 140-character *weibo* do you want to read?' Version 2: 'By following others, you keep updated on them. You can find people you are interested in through search, weibo (user) ranking, and the 'Hall of Fame,' and follow them. Their updates will be posted to your wall.' Version 3: 'When following people you are interested in – be it celebrities or your friends, you become their 'fan' (quatation in original text), and their contents that they share will appear on your wall. You will be notified of their updates immediately.'
Commenting	Users can leave comments on posts by any user, as long as they are visible to them through subscription or search.	'Voice your opinions on a *weibo'*
Tagging	When posting a new *weibo*, reposting others' *weibo*, or commenting on a *weibo*, an '@' sign can be use to mention and notify users that the user follows.	'When you post a *weibo,* tag your friend with a '@' sign and they will be notified. This function is equivalent to having a chat with them.'
Hashtagging	Users can indicate the topic of their *weibo* by wrapping it with two '#' signs to make it more searchable. The platform also runs an algorithm to monitor trends on the topic. The results are updated in real-time and appear on user's wall.	Version 1: 'It indicates what people are talking about recently' Version 2: 'You can start a topic by wrapping the text with two '#' signs. For example, #Taiwan is awesome, # Lunar New Year'. A topic can be an event or a trending word. By searching with a topic you can find relevant information more quickly.' Version 3: 'You can start a topic about an event or about a trending keyword. You can invite your friends to participate in the discussion on the topic. And you can participate in topics that other have started. For example, #Golden Music Award#, #Taiwan eats#'

Besides text and emoticons, users can also attach various types of content, including pictures, videos, polls, and all types of links, to a blog post (see Figure 3). Cell phone users also have the option to indicate a geographical location.

The service description also claims: “Post more, comment more, and interact more—that’s the secret to popularity (多分享，多評論，多參與互動，是維持人氣的不二法門。).”

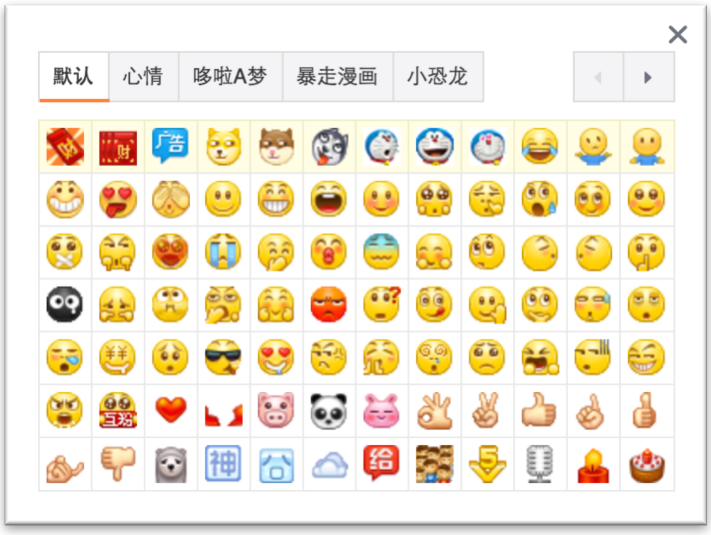

Figure 2. Default emoticons on Weibo

Figure 3. Web interface of Weibo posting section, including various linkable items

It positions itself as "a fun and information-rich communication platform" that serves the purposes of "documenting everyday and life events," "sharing with friends anytime anywhere," "keeping updated with celebrities," and "keeping track of the most up-to-date hot topics."

The open connectivity design (anyone can follow anyone), the encouragement of interaction, and the infrastructure for content sharing allow Weibo to build an online community in which everyone is connected. As users consume content from the platform, they become part of a massive linguistic community. It is massive in comparison even to another popular connective service, Wechat,[27] whose emphasis is on private and direct connection, and whose linguistic community is much smaller in size and also more segregated.

Due to the reasons stated, the platform has also achieved great success within the entertainment business.[28] Its connectivity is capitalized on by celebrities to gather fans and attention, but this also benefits the platform because it draws in a large number of users, forging a win–win situation. Content sharing has been strengthened, especially to accommodate the sharing of entertainment content—shifting the whole tone of the platform into one that is entertainment-oriented. According to a 2015 report from the platform, the most popular types of topics are social topics (25.6%) and celebrity news (25.5%).

Weibo thus becomes a source that provides endless entertainment. Users turn to Weibo when they feel the need to be entertained. Such ferocious consumption of

---

[27] Wechat is a service that is built around mobile devices offered by the Chinese technological magnate Tencent, with 69.7 million MAU globally according to a recent report. Wechat provides connected individual users with services including instant messaging, audio/video chat, subscription to personal updates, and beyond, while also allowing individuals to subscribe to, and potentially interact with, interested public accounts that publish various information. Compared to Weibo, Wechat creates a social environment that is more private, and one that is based on off-line relations. The connectivity channeled by Wechat is thus less far-reaching compared to Weibo.

[28] See an analysis article here: http://www.jianshu.com/p/72bf7ca7e9c6. In the article, it says: "this is mainly because as Weibo improves its user experience, it has also strengthened its partnership with the entertainment business, and it attracts celebrities to broadcast their personal updates on the platform to reel in more users" (Translation from Chinese: "这主要因为微博在提高客户体验的同时，扩大了与娱乐圈的合作伙伴关系，利用明星效应的时时动态吸引更多的客户。").

entertainment has led to the term *shua weibo* (刷微博). It means browsing Weibo, but the verb *shua* (刷) 'to swipe' highlights how users quickly browse through content posts on their wall, particularly on touch-screen devices—people use fingers to slide across the screen to scroll up and down pages. The amount of information (including linguistic information) that people consume has exploded compared to the pre-Internet age. The popularization of smart mobile devices also makes accessing the Internet easier and faster, allowing the Internet to penetrate every waking moment of a modern urban dweller.

The thriving of Weibo benefits greatly from the establishment of the Great Firewall of China at the beginning of the 2010s. The goal of this cyber barrier was mainly to block access to Facebook, Twitter, Google, and YouTube, where uncensored information is free flowing. A direct effect of this policy is that it creates a safe environment for Chinese counterparts of these sites, such as Baidu and Weibo, to bloom uninhibited without competitors. Although Chinese netizens initially resisted the policy, and most have learned countermeasures such as using/purchasing overseas VPNs to gain uncensored access (or 翻墙, literally 'to climb the wall'), most users settle into the more comfortable cyber lifestyle of acquiring information—staying inside the wall. Chinese cyberspace thus has developed into an integral and self-sustaining community that uses its own language and creates its own trends. [29]

The advancement of the cyber era into Web 2.0 (marked by the rise of social media) revolutionized the pattern of information flow. Now, anyone can become the source of information, and all are equipped with the power to radiate information. The philosophy that arises from the cyber infrastructure update naturally goes against a government that exerts a tight ideological control. In addition to the cyber Great Wall, a censorship system is also enforced in these domestic sites. Weibo maintains a filter that

[29] VPN stands for Virtual Private Networks. It allows Internet users to connect to the Internet via a server run by a VPN provider. All data traveling between the users' device and the VPN server is encrypted and hence can bypass the censorship of the Great Fire Wall.

flushes out sensitive political information on its platform, while Baidu also bans the search for certain blacklisted keywords.[30]

In a sense, the censorship imposed by these Chinese websites and directed by the Chinese government is similar to that of their highly commercially-oriented overseas counterparts in terms of how they manipulate information flow. Van Dijck points out that Facebook frequently adds "invisible algorithms" which "control the visibility of friends, news, items or ideas" (2013, p. 49). The aim of promoting certain things and people that cater to the preference of its users turns into a selective filter with the platform's agenda of monetizing its service through ads, and thus users are unknowingly blocked off from certain information. Similarly, starting from 2009, as younger adult users increased on Twitter, the platform started to cater news feeds and celebrity updates to its majority users. Its search results are also filtered by assigning more weights to what the platform deems to be "influential" users, thus sinking the opinions of "uninfluential" users to the bottom (Van Dijck, 2013, pp. 74–75).

The platforms that thrive inside and outside the Great Firewall are thus similar in that none of them are completely neutral; they both cater information flow to powerful parties—in the case of global platforms such as Facebook and Twitter, they yield to capitalistic forces, while Chinese platforms also yield to the government. Besides external censorship, users also exert some kind of self-censorship that regulates their speech and content on the platform so that they can remain popular on it. Research indicates that Twitter users avoid controversial and negative topics to manage their public-facing persona (Marwick & Boyd, 2011a, pp. 125–6).

To sum up, Weibo is an open platform that not only enables, but also encourages, connectivity. The pursuit of popularity, as trained by the platform, has propelled its users to post and share. However, as an open platform, it is also subject to public scrutiny. External censorship prohibits users from posting content that would provoke people by its political nature, and at the same time, self-censorship of users also steers them away from

[30] Blacklisted search terms on Sina Weibo continue to be updated by the China Digital Times <http://chinadigitaltimes.net/2013/06/grass-mud-horse-list/> and New Tang Dynasty Television claimed to have obtained internal documents revealing Baidu's censorship of 13 topics http://www.ntdtv.com/xtr/gb/2009/05/04/a290037.html#video, reported in 2009.

posting content that will be received negatively by other users. Among these strategies, *ye-shi-zui-le* is one of such self-censorship devices.

4.2.4. Discourse Grammar and Construction Grammar

Chinese has long been considered a discourse-oriented language that cannot be fully captured by sentence grammar (Chao, 1968; Li & Thompson, 1976; Tsao, 1979; Chu, 1998; LaPolla, 1995). The discourse features most widely studied in Chinese are anaphoric referencing and its topic-comment structure. On the one hand, it was osbserved that a huge proportion of the language patterns in Chinese does not conform to the grammatical relationship of subject, verb and object, and can only be adequately described as having a pragmatic-driven topic-comment relation, while on the other hand, the information unit of Chinese seems to exceed the boundary of a clause. Chu (1998), for instance, suggests that a sentence (which can span across several clauses, sometimes reaching the length of a paragraph) should be the basic information unit of Chinese, based on evidence from anaphoric referencing.

The endeavor to install an alternative grammar for Chinese finds an echo in the theories of discourse grammar, as espoused in Heine, Kaltenböck, Kuteva and Long (2013). Thetical Grammar (Kaltenböck, Heine, & Kuteva, 2011) is one such theory that deals with linguistic phenomena beyond the scope of sentence grammar, or those that cannot be captured by sentence grammar. In this theory, a thetical is non-syntactic linguistic information that is anchored to a syntactic information unit and usually follows it. A thetical does not have a concrete meaning, though it could relate to the anchor, but more often it is shaped by its function in discourse as it concerns the discourse situation of social interaction. One typical type of thetical is the comment thetical, which is similar to a discourse comment in a Chinese sentence unit. However, describing the structural features of theticals has only been done through very indefinite terms, such as through their prosodic features and relative positional flexibility. The many types of theticals have vastly different forms, and the form-function mapping of theticals as a general category is largely understudied.

But this is not to say that there does not exist form-meaning or form-function pairings in theticals. In the theory of Construction Grammar, a construction can be a meaningful language unit with a linguistic pattern of any length. Tomasello (2003), for

one, has proposed that there are complex constructions dealing with the relation between clauses (coordination and subordination), and these clauses would be considered a construction at the level of a discourse. Several types of constructional theticals have also been identified in English, such as comment clauses (*I guess*), reporting clauses (*according to him*), tag questions (*isn't it*) and others. (Brinton, 2008; Dehé & Kavalova, 2007, p. 2–4; Dixon, 2005, p. 233; Espinal 1991, p. 726–7; Lambrecht, 1996). These theticals are prefabricated chunks that are constructionalized into functional linguistic units that are attached to a sentence. I hypothesize that *ye-shi-zui-le* is thus a comment clause that is found in a discourse construction, whose components have a stable relationship that is governed by the construction. The following section will investigate whether there are form-meaning mappings surrounding content where *ye-shi-zui-le* occurs.

4.3. Data Collection and Manipulation

Driven by the hypothesis that there exists a discourse construction which contains *ye-shi-zui-le*, I examined the language data containing *ye-shi-zui-le* and employed various quantitative and qualitative methods to evaluate the discourse environment where the phrase occurs. Python codes used for data collection and data manipulation are stored in an online repository.[31]

4.3.1. Data Collection

Because the phrase *ye-shi-zui-le* gained popularity on the Weibo platform, this study examines its current usage on the same platform to understand how it has developed, and what it is developing into. Synchronic data of *ye-shi-zui-le* were collected in a semiautomatic fashion by treating the web as corpus. I wrote a computer program to access the Internet and download text data from a designated online location—a method called web scraping.[32] The goal was to scrape as many weibo posts that contain the phrase *ye-shi-zui-le* from the Weibo platform as possible. Typically, this would be done

[31] The online location is <https://github.com/xinjiap/yszl.git>. All the algorithms used in the analysis can be found with the identified file name in this repository.

[32] scrape_weibo.py

by scraping the search results of *ye-shi-zui-le* produced by the platform. But because Weibo is not an open-source platform and limits people's access to its data, I had to reroute and scrape data from a search engine (http://t.zhongsou.com/). One issue with the search engine is that the results it returns are only a small proportion of the actual hits on the Weibo platform, so that data I collected from here is only part of the actual usage.[33] Overall, the data do not seem to be biased and skewed in ways that could make them a bad sample. They are evenly distributed across time, topics, and users, and I deem them appropriate for the current study.

Eventually, I built the WEIBO corpus from data collected by web scraping from Zhongsou search results, and it is used as the main dataset in this study. It consists of 6,602 unique blog posts that contain the target phrase *ye-shi-zui-le*, which were posted over a period of about eight months by users on the Weibo platform during the year 2016. The data were collected over 10 scrapes that span six months, because each Zhongsou search produces only a maximum of 50 pages of the most recent results, which is about 1,000 blog posts. The scrapings thus had to be done periodically. To avoid duplicates, I scraped only the "original" blog posts (excluding reposts, and before admitting new blog posts into the dataset, each of the new entries was compared against existing posts. The corpus contains only text data, and each blog post was given a number as identification. When giving examples in the following sections, I will also include their identification number. The metadata of the blog posts are all preserved externally, so if links, user names, and time stamps need to be retrieved, they are also easily available.

4.3.2. Data Manipulation

As in other chapters, I adopt a usage-based approach to examine the phrase *ye-shi-zui-le*, and possibly the discourse construction that contains it. I work with the theoretical assumption that surface form has the most powerful generalization. The generalizations that language speakers can make from a "surface form" are not restricted

[33] The two searches seem to use different algorithms, and neither of them produce exhaustive results. But by sheer comparison of numbers, Weibo definitely produces more results than Zhongsou. As an example, I searched for posts containing *ye-shi-zui-le* posted during 2017/02/20 17:00-17:59 (UTC +8) on both platforms, and the ratio was 48:6 (Weibo vs. Zhongsou). The results I used, produced by Zhongsou, are probably less than one tenth of the actual *ye-shi-zui-le* productions on the platform. I was never able to find out the daily usage frequency of *ye-shi-zui-le* on the platform, but the number is definitely remarkable.

to the (linguistic) formal and structural patterns enunciated in grammatical terms, but more importantly, they extend to the patterns of occurrence of different semantic categories in a constructional slot. The more frequently a certain lexical item (or items of a certain semantic category) occurs in the constructional slot, the more likely it will influence the meaning of the construction. The preference of certain lexical items for a schematic constructional slot—its idiosyncrasy—is also affected by this prototypical category. My analysis will thus be a combination of quantitative and qualitative analyses: the qualitative looks into emerging semantic categories, and the qualitative looks at the occurrence frequency of a lexical item or a semantic category which reveal the constraints of the construction.

After a preliminary and exploratory analysis of the occurrences of *ye-shi-zui-le* in the WEIBO corpus, I formulated a set of hypotheses and came up with methods to test them. The following is a list of both the hypotheses made and the methods used. The results will be reported in the next section.

Hypothesis I: Is the phrase *ye-shi-zuile* an entrenched phrase? Does it have variational forms?
Method: Annotate blogposts that contain the exact match of the phrase *ye-shi-zui-le*, and identify the phrase used in those blogposts that do not have exact match.
Code: Count_yszl.py

Hypothesis II: Is the subject of *ye-shi-zui-le* the first-person singular *wo*?
Method: Annotate blogposts that contain the exact match of the phrase *ye-shi-zui-le*, and identify the phrase used in those blogposts that do not have exact match.
Code: Count_subject.py

Hypothesis III: Does the phrase *ye-shi-zui-le* always come after a clause (indicating that it is a comment)?

Method: Count the number of words before and after *ye-shi-zui-le* in each blogpost, and determine the relative position of the phrase in the blogpost in ratio to the word count of the entire blogpost.

Code: Position.py

Hypothesis IV: Does the phrase *ye-shi-zui-le* show discourse coherence to the text before and after it?

Method: Examine the characters and punctuations immediately before and after the phrase.

Code: Coherence.py

Hypothesis V: Does the text have discourse coherence with the phrase *ye-shi-zui-le* or exhibit semantic idiosyncrasies?

Method: Exploratory analysis in YSZL Sample Corpus (see Appendix C for this 100-blogpost random sample corpus, extracted with get_sample.py), followed by concordancing of keywords in an observed semantic category.

Code: Concordance.py

Hypothesis VI: Does the phrase *ye-shi-zui-le* have a stance?

Method: Qualitative analysis of the YSZL Sample Corpus

Hypothesis VII: If there is a stance, is it positive or negative?

Method: Sentiment analysis based on lexical items and emoticons.

Code: Sentiment.py

Hypothesis VIII: Are emoticons part of the discourse construction?

Method: Position analysis of emoticons relative to the phrase *ye-shi-zui-le*.

Code: Emoticons.py

For the quantitative analysis, I wrote my own codes (run in Python) to process the data using corpus linguistics methods. The general methods I used for this analysis are frequency counts based on linguistic patterns and collocational searches based on distance. The codes used to test the hypotheses are shown above (and can be found in the online repository). For the qualitative analysis, I closely examined the Sample Corpus (an 100-item random sample corpus generated also by a code).

Two important Python packages I used are a regular expression package and a Chinese segmenter and POS tagger package. Together, they allowed me to mimic searches based on CQL grammar but only more flexible. The regular expression package in Python allows text searches based on patterns. To locate the keyword of each blog post, for instance, I compiled a regular expression pattern ("我?(也是醉了)+"), which basically means "search for a string *ye-shi-zui-le*; if it repeats more than once, consider all the repetitions as part of the pattern, and if it has a preceding *wo*, also consider it part of the pattern." Another important Python package I used is Jieba, a Chinese word segmenter. Its basic function is to segment words in a Chinese sentence and conduct POS tagging based on the segmentation, a dictionary, and some grammatical analysis. The merit of using corpus methods to process textual data, besides giving us a quantitative perspective, is that it counts following objective rules and can thus be more objective than manual coding. Furthermore, all the results produced by these codes are replicable.

Depending on the type of analysis needed, data are cleaned on the run so that unnecessary information is ignored by the program. Symbols that have special functions—such as "@user" or "#topic#" or "http://" links, as well as emoticons that are converted into "[emoticon]," are temporarily removed if they are not a point of interest in the specific analysis. Variants of *ye-shi-zui-le* are normalized for analysis starting from 4.4.2. because regular expressions do not intelligently recognize variants.

4.4. The YSZL Construction

The findings based on the hypotheses and methods above suggest that the phrase *ye-shi-zui-le* is part of a discourse construction, which I call the YSZL construction, and it has the following constructional frame:

(35) The YSZL construction: [[EVENT] [(*wo*) *ye-shi-zui-le*]] [emoticon]

The first component is the EVENT component, which is more schematic (meaning that it is instantiated more often with varying linguistic contents), and the second component is a substantive construction (meaning that it mostly appears as a chunk and has the same wording in every instantiation), which is the phrase *ye-shi-zui-le*; it has a prototypical first-person-singular subject *wo*. The YSZL construction is optionally followed by one (or several) emoticons.

YSZL is recognized as a construction because it exhibits defining the characteristics of a construction as espoused in Construction Grammar (Goldberg, 2013; Hilpert, 2014), mainly in the following aspects:

a. The construction has systematic form–meaning mapping.

a1. The YSZL construction (4.4.3)

FORM: The substantive part of the construction, though sometimes appearing as a stand-alone phrase, more frequently occurs with the schematic part of the construction, and almost always in the shown order

MEANING: An event triggers the experience "*ye-shi-zui-le*"

a2. The EVENT component (4.4.4)

FORM: A description of an event, sometimes giving a context/presupposition relevant to the occurrence of the event

MEANING: The event described is counter to the expectation/presupposition

a3. The *ye-shi-zui-le* construction

FORM: *Ye-shi-zui-le* (4.4.1)

MEANING: A metaphorical extension of the experience evoked by alcohol intoxication

FUNCTION: It is a substitution of emotions unexpressed (4.4.5)

b. The construction has a component with a non-compositional meaning. Specifically, the phrase *ye-shi-zui-le* has a meaning that cannot be deduced by the totality of meaning from its component lexical items. (in 4.1)

c. The construction has idiosyncrasies.

c1. *Wo* (first-person singular) is the prototypical subject of the phrase *ye-shi-zui-le.* (4.4.2)

c2. Of its co-occurring emoticons, the high-frequency tokens show contradictory affective valence, such as "tears of joy" (or "laugh–cry" as labeled by the platform) and the doge face (4.4.6)

d. The construction, including the components of the construction, is in a network of linguistic constructions.

d1. The overall structure of the YSZL construction extends from the information structure construction [[topic] [comment]] which is prominent in Chinese (4.4.3)

d2. The phrase *ye-shi-zui-le* can be used as a stand-alone, or it can be used with an entity construction in place of the EVENT component. It thus suggests that *ye-shi-zui-le* is a node in the network of constructions (example in 4.4.6)

The discussions of these features can be found in the sections indicated above.

4.4.1. The Entrenchment of the Phrase *Ye-shi-zui-le*

The status of the *ye-shi-zui-le* phrase as a substantive rather than a schematic construction is evidenced from its high token frequency. The exact form of the *ye-shi-zui-le* phrase dominates the WEIBO corpus.

The frequency of *ye-shi-zui-le* is calculated against its variants in the corpus. In the 6,602 blog posts, with the exception of only 20 blog posts that use some variations of the *ye-shi-zui-le* phrase, all of the blog posts use the exact phrase of *ye-shi-zui-le*. In terms of percentage, an overwhelming 99.7% of the blog posts have the exact form of *ye-shi-zui-le*. From a technical point of view, such a result seems to be the necessary

consequence of methodological choices for data collection—the WEIBO corpus is built from search results of the phrase *ye-shi-zui-le* (也是醉了). However, the fact that variants exist suggests that the results yielded by the search are not all exact matches of the phrase. Table 10 shows variants of the phrase extracted from these 20 blog posts, with a summary of the variational forms followed. The inclusion of not only the exact forms, but also the nonexact forms, in the results suggests that the search engine has some degree of built-in fuzziness, that it returns not only exact matches but also fuzzy matches. Because of this, I can confidently argue that the dominance of *ye-shi-zui-le* stands.

Table 10. Variational forms of the *ye-shi-zui-le* phrase in WEIBO corpus

也是醉醉了 [362, 5485]	我也是醉醉了 [4585, 5860]
ye-shi-zui-zui-le	*wo-ye-shi-zui-zui-le*
我也是醉的不能再醉了 [721]	我也是醉[dx 数落]了 [5008]
wo-ye-shi-zui-DGR-no-can-more-*zui-le*	*wo-ye-shi-zui*-(emoticon:scold)-*le*
我真的是醉了 [830]	我也是醉[爱爱爱]了 [5512]
wo-really-*shi-zui-le*	*wo-ye-shi-zui*-(emoticon:love)-*le*
我也是醉 了 [937]	我也是醉[冲啊]了 [5625]
wo-ye-shi-zui-(space)-*le*	*wo-ye-shi-zui*-(emoticon:run)-*le*
真是醉了 [2602]	也是是醉了 [5657]
real-*shi-zui-le*	*ye-shi-shi-zui-le*
真的是醉了 [2890, 3497, 5229]	我也是[弱]醉了 [5729]
really-*shi-zui-le*	*wo-ye-shi*-(emoticon:despise)-*zui-le*
虽是醉了也满好吃 [4000]	我也是 醉了 [6176]
although-*shi-zui-le*-also-rather-delicious	*wo-ye-shi*-(space)-zui-le
真滴醉了 [4197]	真的是醉醉了呀 [6449]
really-*zui-le*	really-*shi-zui-zui-le*-MOD:*ya*

Note: The following summarises the different types of variations of *ye-shi-zui-le*

(i) misplaced white space x 2 [937, 6176];
(ii) possible typo [5657];
(iii) emoticon insertion x 4 [5008; 5512; 5625; 5729];
(iv) *zui*-reduplication x 4 [362; 4585; 5485; 5860];
(v) *zui*-degree-modification [721].
(vi) *ye*-replacement by *zhen*(*de*) (meaning *really*; *truly*) x 6 [830; 2602; 2890; 3497; 5229; 4197 also drops *shi*];
(vii) *ye*-replacement and *zui*-reduplication [6449];
(viii) *ye*-dropping by *sui* (*although*) [4000];

The frequency effect of *ye-shi-zui-le* also gives the phrase a status of a "chunk." One the one hand, it is resistant to internal operation such as insertion. Examples in Table 10 show that wrong-word insertion, emoticon insertion, and white-space insertion all occur rarely in the corpus (only seven blog posts in a 6,602-blog-post corpus, see (i) through (iii) below). On the other hand, it is conducive to constructional operations such as reduplication: The reduplication of *ye-shi-zui-le* (*ye-shi-zui-le-ye-shi-zui-le*) occurs in 454 blog posts, making up 6.8% of usage in the entire corpus.

The variations in the corpus are cleaned from here on to match the *ye-shi-zui-le* phrase in the following analysis.

4.4.2. The Subject of *Ye-shi-zui-le*

In the WEIBO corpus, it is observed that *wo* 'I' is the highest frequency subject of the phrase, and hence the prototypical subject of the phrase. Table 11 lists the high-frequency nouns/pronouns/zero anaphora that immediately precedes the phrase *ye-shi-zui-le*:

Table 11. High-frequency noun/pronouns that occur immediately before *ye-shi-zui-le*

noun/pronoun form	translation	occurrence rate
wo (我)	I	45.0%
zero[a]		19.7%
xiaobian (小编)	the editor (self-reference)	0.79%
shushu (蜀黍)	uncle; older man	0.35%
ziji (自己)	reflexive	0.33%
x-*men* (x 们)	pluaral marker	0.2%
an (俺)	I (colloquial)	0.18%
ni (你)	you	0.15%
ta (他，她)	he; she	0.06%

[a]Calculated by counting punctuation and white space

Among these subjects, not only *wo*, but also *xiaobian* (editor: self-reference), *ziji* (reflexive), and *an* (I: colloquial) are all first-person subjects. In my opinion, first-person singular as the entrenched subject of the phrase makes *ye-shi-zui-le* idiosyncratic in the sense that it primes the experience pronounced through *ye-shi-zui-le* as one that is highly

subjective—so subjective that it can only be self-professed, rather than being identified by others.

When compared to the usage pattern of *zui* in modern Chinese (with the meaning 'alcohol intoxication'), we see a clear difference in terms of grammatical subjects. Corpus data show that the state of *zui-le* is more often identified in others (e.g., 3SG-*zui-le*), and hence by others. [34] Of the 524 items that meet the requirements of the search, the second- and third-person singular pronoun occur more often (258 occurrences) than the first-person singular (191 occurrences) in the closely preceding slots. Table 12 shows a clear contrastive pattern of *ye-shi-zui-le* and *zui* in other usages in this respect, and it provies that *ye-shi-zui-le* is a construction independent of the usage of *zui.*

Table 12. Distribution of subject person in *ye-shi-zui-le* and *zui*

	Frequency of occurence			Row total
	1st person	**2nd &3rd person**	**others**	
ye-shi-zui-le	3056 *3008.24* (0.76)	50 *285.35* (194.11)	3496 *3308.41* (10.64)	6602
zui	191 *238.76* (9.55)	258 *22.65* (2445.67)	75 *262.59* (134.01)	524
Column total	3247	308	3571	7126

$\chi^2 = 2794.745$, df = 2, $\chi^2/df = 1397.37$, $P(\chi^2 > 2794.745) = 0.0000$

This is expected, because *zui* (as in 'alcohol intoxication' or 'drunkenness') is regarded negatively in modern Chinese culture, and the self-positivity bias tends to lead people to disassociate themselves from negative qualities (Brown, 1986; H. Zhang, Guan, Qi, &

[34] I used the zhTenTen11 web corpus on SketchEngine and searched for items where a pronoun occurs within the L3 window of the phrase '*zui-le*." This web corpus was built in 2011 prior to the emergence of *ye-shi-zui-le.*

Yang, 2013). Also, the symptoms of alcohol intoxication, including the so-called Asian flush, provide drunkenness with an external dimension that it is highly identifiable.

The disparate collocational patterns with pronouns of the two usages of *zui*—as in intoxication, and in the phrase *ye-shi-zui-le*—can indicate two things. First, the two usages of *zui* have varied connotations, since *zui* in *ye-shi-zui-le* does not have the same negative connotation as *zui* expressing actual intoxication.[35] Second, *ye-shi-zui-le* probably does not have an external dimension that can be identified by others; it is not a physical description, but refers to a mental state.

4.4.3. The Discourse Integration of the EVENT Component and *Ye-shi-zui-le*

I argue that *ye-shi-zui-le* is part of a bigger construction, the YSZL construction. It is preceded by an EVENT component, and together they form the essential components of the YSZL construction: [[EVENT] [(*wo*) *ye-shi-zui-le*]]. This section provides structural evidence not only for their integration, but also for their ordering in the construction. Their relative position indicates an association with the topic-comment construction. Their semantic connection will be further explored in section 4.4.5 after an analysis of the EVENT component is given in 4.4.4.

Quantitative evidence reveals two structural features. First, *ye-shi-zui-le* is more likely to occur at the end of a blog post. In other words, it is more likely to be preceded by other contents instead of leading other contents. Using data in the WEIBO corpus, I treated each blog post as an occurrence of *ye-shi-zui-le* and examined its relative position in

[35] Alcohol use, especially its excessive use to the extent of drunkenness, has never been appreciated in Chinese culture. Bans on alcohol consumption (though not effective) can be traced as far back as Shang. Buddhism, a popular secular religion introduced to China in Tang, in general denounces the use of alcohol; in the monk Daoshi's scriptures, there are discussions of drunkenness as a cause of suffering (Benn, 2005, p. 224). A famous essay imagining a debate between tea and alcohol, written in about the same period, points out the intoxicating effect of alcohol, saying it makes people "muddled" and "garrulous" (Benn, 2005, pp. 220, 228). More recently, in colonial Manchuria, evidence-based research from the developed world warned people against alcoholic consumption due to its impairment of physical and mental capacities, and alcohol use was increasingly depicted as instigating criminal behaviors (Smith, 2012, p. 60–64). In modern-day China, it has been noted that cases of alcohol-related physical diseases and alcohol dependence have increased steadily since the '70s (Cochrane, Chen, Conigrave, & Hao, 2003; Hao et al., 2004), and it has been proposed that alcohol use disorder has become a major public health problem in China (Cheng, Deng, Xiong, & Phillips, 2015). The negative associations with *zui* lie in how it seems to bring out the vile side of people by inhibiting the moral self.

relation to all other characters in the blog post (#topic#, @user, [emoticon], and hyperlinks are removed, and punctuation is also ignored in these calculations). Then I examined the relative position of the *ye-shi-zui-le* phrase in each blogpost in the corpus. The statistics are transposed into Figure 4 below with the first column representing the condition of "no text preceding the phrase," and the last column representing the condition of "all other text preceding the phrase." It shows that while the phrase seems to have wide flexibility in where it occurs in a blogpost (observed normal distribution from the first 10% to the last 10%), its distribution at the sentence-final position is highly significant. In other words, besides the 331 items that have the phrase *ye-shi-zui-le* at the beginning of a post, approximately 94.98% of the items in the corpus follow the constructional frame of [[EVENT] [(*wo*) *ye-shi-zui-le*]]. [36]

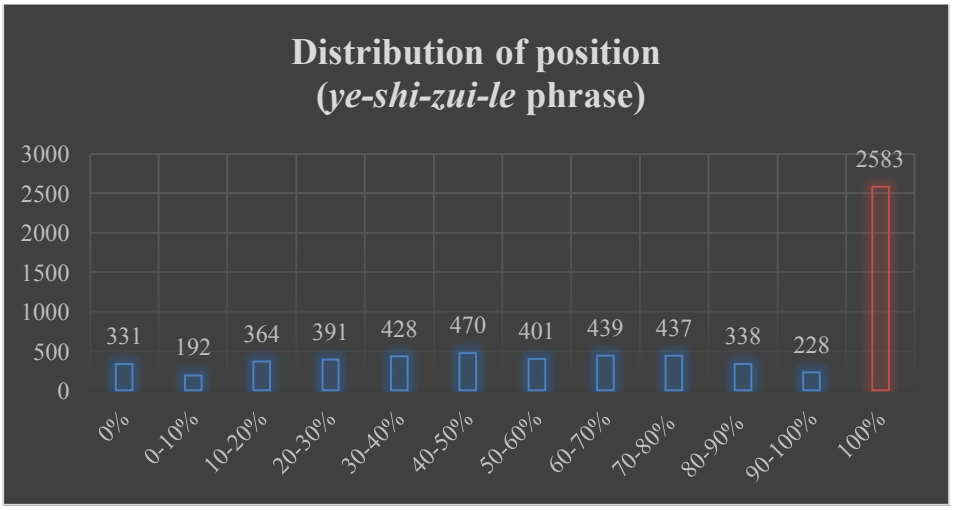

Figure 4. The position of *ye-shi-zui-le* in a blog post

Second, analysis of symbols immediately preceding and following the phrase *ye-shi-zui-le* (including variations of the phrase, such as the use of *wo* as subject and reduplication of *ye-shi-zui-le*) suggests that the prephrasal slot is an open slot and the postphrasal slot is a closed slot. While Internet users do not strictly follow the punctuation conventions of

[36] Because the data are processed by nonintelligent algorithms, variations of the *ye-shi-zui-le* phrase are not recognized, and at least some such cases are included in the 94.98% category. For instance, [125] *wo-nima-ye-shi-zui-le* (*nima* is an emotive) and [411] *jie-ye-shi-zui-le* (*jie* means 'sister,' but here it is also a self-reference).

the written language (in terms of using a comma as a pause and a period to indicate conclusion), similar categorization resembling the use of comma and period in Internet language is observed. Figure 5 shows the occurrence rate of various symbols in both slots. In the prephrasal slot, a character occurs most often, followed by the use of a comma, dwarfing all other punctuation. By comparison, an immediately following character is rare, and the comma does not show statistical advantage compared to the period and the exclamation mark in the postphrasal slot. Based on the data, I would argue that the use of characters and symbols in the prephrasal slot suggests it is a connective slot that builds connection between the phrase *ye-shi-zui-le* and the preceding text. Conversely, their usage in the postphrasal slot indicates a discourse boundary. In terms of discourse structure, the text preceding *ye-shi-zui-le* is more likely to show discourse coherence and integration with phrase. The quantitative findings are confirmed in my qualitative analysis of the YSZL Sample Corpus. In the 100-item corpus, I found 90 items where *ye-shi-zui-le* is preceded by content that triggers the experience.

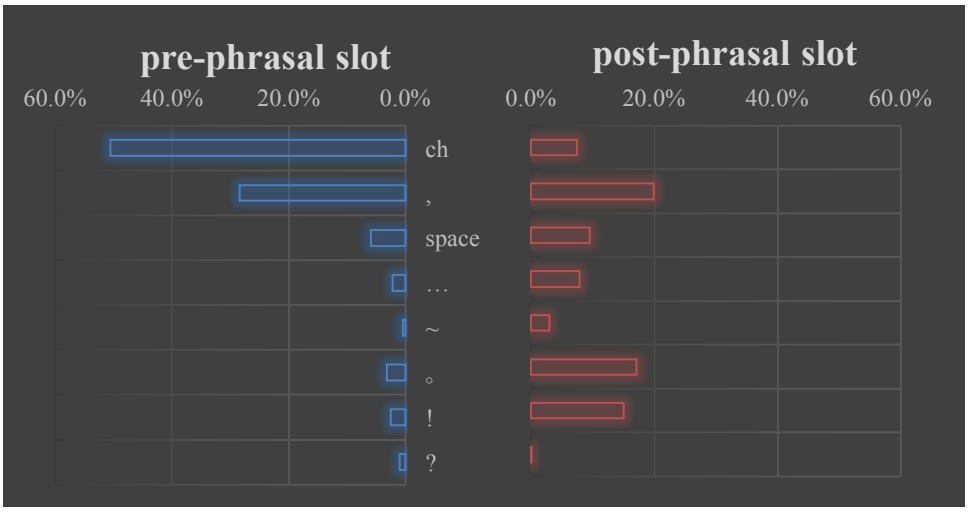

Figure 5. Occurrence pattern of symbols/characters immediately preceding and following *ye-shi-zui-le*

Overall, the data above show that a prototypical YSZL construction (measured by frequency of occurrence) has the order of [[EVENT] [(*wo*) *ye-shi-zui-le*]], and the two components are structurally tied to each other, with the EVENT component preceding the phrase *ye-shi-zui-le*.

The ordering of the components, as well as the discourse structure between the components, resembles that of the topic-comment construction. The topic-comment construction is a pragmatically driven construction that is commonly found across languages (Givón, 1988; Gundel, 1988) The emphasis of the construction is primarily on information structure, where information is given prior to a proposition being made. Seen from this perspective, the EVENT component is thus a topic, and the phrase *ye-shi-zui-le* is a comment on the event:

[[topic] [comment]]

[[EVENT] [*ye-shi-zui-le*]]

A comment can be a factual statement about the topic, an evaluative proposition about the topic, or both. And in this case, since it has been established that *ye-shi-zui-le* does not refer to the factual state of alcohol intoxication, it is more likely that *ye-shi-zui-le* is an evaluative statement towards the event. Combining this finding with the observation from 4.4.2, it becomes clear that the construction is a "subjective" evaluative statement towards the event, which is about what "I" think, and how "I" feel. *Ye-shi-zui-le* as a comment exhibits features similar to that of comment theticals in English. Brinton (2008), for instance, finds a large number of comment clauses employing first- and second-person subjects that become fomulaic. She states that the grammatical subject person is an indication of the comment theticals's concern in speaker-hearer relation. She further comments that first-person theticals, such as *I find*, *I guess* and many others in English, are "inherently subjective" (p. 238). I believe this is also the case of *ye-shi-zui-le.*

In the following section, the focus is on developing a better understanding of the topic (the EVENT component) that is structurally and semantically linked to the comment, so that we are more likely to understand the comment that is directed towards the event. In what immediately follows, I will give a preliminary description of the form–meaning idiosyncrasies of the EVENT component before closely examining its semantic content in 4.4.5.

4.4.4. The EVENT Component

An idiosyncratic pattern that surfaces from the data is the structural marking that suggests an event takes place contrary to the speaker's expectation. Sometimes a context is provided, but more often, the context or presupposition is inferred.

(36) 到哪都下雨_我也是醉了_回家 [5937]
'it rains everywhere I go'_ *wo-ye-shi-zui-le*_ 'Going home now'

(37) 我们居然看见蟑螂还能淡定的洗碗,
'To my surprise we could clean the dishes without being disturbed by the sight of cockroaches around'

并且讨论蟑螂的大小个数,
and talked about the sizes and numbers of those cockroaches,

也是醉了。[4086]
ye-shi-zui-le

(38) 拼了老命赶机,
'I tried so hard to catch the flight,'

飞机既[sic 竟]然延误[擦汗]
'but the flight was unexpectedly delayed (emoticon: wipe the sweats)'

这天气[流汗]延误
'It was delayed in such a weather (emoticon: sweating face)'
我也是醉了... [6204]
wo-ye-shi-zui-le

In (36), the user expects fair weather when traveling, but the reality is that recently he might have encountered more rain than he would like whenever he needs to work outside the office. Here, the construction *nar dou* 'everywhere' implies that the frequency of such incidents is excessive. In (37), the user finds it odd being so at ease with the presence of cockroaches, and even paying attention to their characteristics, while washing dishes. The presupposition is not apparent here, but the unexpectedness is revealed in the use of *juran*, which points the reader to the contradicting cultural stereotype that girls usually freak out at the sight of creepy-crawlies. In (38), the user sets the presupposition that he or she might not be in time for the flight's departure because he

or she is rushing all the way, but contrary to such an assumption, not only is the user on time, he or she is unnecessarily on time because the flight is delayed. The unexpectedness is also in the delay of good weather. The adverb that reveals such "unexpectedness" is *jingran.*

I looked further into the data and identified a list of possible adverbs in Chinese that suggest such "unexpectedness" and conducted a corpus search in the WEIBO corpus. The results show that 21.08% of the items (1,392 items) in the WEIBO corpus used at least one of these adverbs in the L20 window of the phrase (the average distance is L13). The frequency of these adverbs, from high to low, is shown in Table 13.

Table 13. The occurrence frequency of "unexpectedness" adverbs in the WEIBO corpus

hai 还	6.45%	*zhi* 只	1.35%
dou 都	5.76%	*jing* 竟	1.24%
jiu 就	4.33%	*dan* 但	0.82%
juran 居然	1.85%	*que* 却	0.47%
cai 才	1.36%	*dao* 倒	0.26%

It should be noted that because the semantics of "unexpectedness" does not have a compulsory grammatical marking, the actual percentage of items that shows a meaning of "unexpectedness" is higher. Even though the token frequency of each of these modal adverb is not high, the variety of these adverbs indicate an attraction between the EVENT component and the concept of being surprised. The modal adverb is a closed category in Chinese, and this category only has a limited number of items. Table 14 has basically included every member in this category. The type frequency of modal adverbs used in the EVENT component suggests that the meaning of "unexpectedness" is a reoccurring semantic category in the EVENT component, and it indicates that this is a construction that is concerned with subjective stance.

Clary and Tesser (1983) find that compared to expected events, unexpected events are more likely to produce observer involvement and stimulate interpretative activities, because an automatic mode of information processing cannot competently handle the processing of an unexpected event. The EVENT component, where unfamiliarity and novelty is a prominent semantic feature, thus produces the same effect on the recipients

of the information, and these effects are revealed through *ye-shi-zui-le*. On the one hand, *ye-shi-zui-le* expresses speaker involvement, conveying a subjective evaluation. On the other hand, it declares the difficulty of processing this information, which links metaphorically to the experience of *zui* (as I have suggested in section 4.2.1, the metaphorical meaning of *ye-shi-zui-le* might suggest processing difficulty).

Meanwhile, the evaluative stance taken when presented with an unexpected event could also be an important generalization of this construction, and it seems that the affective valence associated with the processing of an unexpected event is neither a positive nor a negative stance. A study investigating the relation between event appraisal and emotion revealed that events perceived as unexpected elicit higher levels of surprise than of other emotions (Roseman, 1996). According to the study, surprise is unlike other emotions in that it is ambivalent in terms of its affective valence. Therefore, I foresee that *ye-shi-zui-le*, if used to indicate an evaluative stance towards the event, is also nonpolar in its affective valence.

In the following section, I will examine the usage data closely to illustrate the kind of stance associated with the use of *ye-shi-zui-le*, which mainly emerges from the usage contexts.

4.4.5. Understanding *Ye-shi-zui-le*

Usage data of the YSZL construction reveal that several types of attitudes emerge from the description of the event, and they somewhat correlate with the referent involved in the event. I suppose it is because the displayed attitudes are sanctioned by participant relations in these events. The following analysis is based on the sampled 100 blog posts from the WEIBO corpus. I retrieved the links for these blog posts and examined the blog posts on their original platform, which allowed me to view picture attachments and video links that were not processed in data collection. The three reference groups I based my categorization on are the distant others, the intimate others, and the self. They vary in the social distance between the speaker and the referent.

Roughly corresponding to the referent group of "distant others" is the content category of "sharing," which constitutes the majority of the sample. This is congruent with the mission statement of the platform and the infrastructure that it enables (see the

discussion in section 4.2.3). In the sample blog posts, I find users share links to funny videos and sensational odd news, sometimes even writing odd events they encounter in their own life to share on the platform. The event that evokes the comment of *ye-shi-zui-le* can be an odd story of a wine-flooded town in France ([4759]), a video of pedestrians who dodge and survive serious car crashes ([2667]), or a picture of the bizarre way people reserve parking spaces by placing water bottles along the space markings ([6047]), as well as complaints towards unhelpful mothers-in-law who make things worse for expecting/new young mothers ([778], [935]) or the experience of being annoyed by serial calls ([5271]).

While most blog posts do not verbally reveal their attitudes towards the event, these events have been construed with a clear orienting attitude. For instance, as a reader, I sense that the first three events provide some kind of amusement/entertainment, and the latter two events seek to strengthen readers' existing disapproval of the given social phenomena. And I assume that what I experience as a reader is the same as the experience of the user who posts the content, because studies have shown news sharing in social media is motivated by the socializing gratification and status seeking when one gains recognition from the like-minded (Lee & Ma, 2012; Quan-Haase & Young, 2010).

However, why the attitude is not revealed verbally, rather than what is the attitude not revealed verbally, is my interpretation of the meaning and function of *ye-shi-zui-le*. In my opinion, *ye-shi-zui-le* operates as an evasive device that allows blog post authors to avoid revealing their responses and attitudes towards the event, and this should be situated in the speech context, which is an open platform. The users need to hide their attitudes, because they would potentially tarnish their online persona if the readers were to be reminded of the self-censorship that social media users impose on themselves (see section 4.2.3). Using the examples above to illustrate, with contents that appear to be funny, the amusement sometimes capitalizes on the misfortune of others. The events that have amused the social media readers have factually been inflicted upon people—the residents of the wine-flooded town, the owner of the broken wine barrels, the drivers who might or might not have survived in the car crash, the car owners who cannot find a parking space because the spaces have been illegally occupied. The schadenfreudian nature of the act of sharing is only justified because the referents of the events are

unknown/unrelated, and other onlookers (other social media users) are kept at a safe distance so as to remain emotionally unattached. But the anonymity does not mean that their misfortunes do not need to be acknowledged. *Ye-shi-zui-le,* in my opinion, in addition to being a response to the "unexpectedness/oddness" of the events, is almost an evasion of the pleasure derived from others' misfortune. In this way, the author of the post appears more humane. Similarly, as regards to the contents that are somewhat critical towards things witnessed or experienced by the post author, most social media users would want to avoid the impression of extreme negativity in this open platform, and *ye-shi-zui-le* is thus used to substitute the worst that might come out of a post author. Furthermore, because the person/organization responsible for instigating the distress can usually be linked to the post author, and hence is less anonymous and more identifiable, the post author avoids a personal attack on whoever is responsible. In this latter scenario, although the referent is not as irrelevant, the referent is kept at a distance because of the existing conflict situation. Referents in both types of situation are thus what I called "the distant others." It seems that the safe distance permits the post author to describe events that can potentially evoke a harsh stance in the audience, but at the same time, post authors substitute their personal stance with *ye-shi-zui-le* so as to protect themselves from judgment in the open cyber platform.

Next most frequently observed in the dataset are blog posts about the self. Some of the blog posters post about themselves, describing their quotidian life events, such as eating, traveling, insomnia, and get-togethers with friends. This is found in more than 30 blog posts in the 100 sampled. These personal updates on social media are what has come to be known as "lifestreaming" on social media. The act is related to status seeking and community building (Barry, Doucette, Loflin, Rivera-Hudson, & Herrington, 2017; Zhao & Rosson, 2009). Those who do this, especially those who seek microcelebrity status, report that the sharing of personal information strengthens their relationship with their audience (Marwick & Boyd, 2011b). Readers of this paper may wonder how Weibo users could be interested in the mundane details of others. Well, although these details could be irrelevant in another context, the middle-class consumerist society that we are situated in has habituated us to pay attention, or even to follow what others are eating, using, and wearing. It is these consumer behaviors that become what others can relate to. In the

following two examples, we read complaints of pricey menus. The examples indicate that consumer information is greatly appreciated, even though it comes in the form of complaint. Looking more closely, both of these complaints are embedded in the users' travel diaries. The excessive details of where he or she stays in Rome, the proximity of a tourist attraction, the great view of the hotel restaurant in (39); the opening of the post, identifying the user's locale in Macau, a casino, the comparison to a specific restaurant in Hong Kong in (40)—the elaboration of the *when* and *where* that are not necessarily relevant to the event reveals the true communicative intent of the users, that this is not really a complaint, but a showing off of their occasional daily life that is not quotidian.

(39) 罗马登陆,住的酒店楼下就是西班牙阶梯,很美!
'Landed in Rome. The downstairs of the hotel he stayed was the Spanish Steps (Scalinata della Trinità dei Monti). So beautiful!'

顺道吃点东西上楼,
'(I) had something to eat on the way back to the hotel.'

这间餐厅就在楼下可以看西班牙阶梯的全景,
'The restaurant was located downstairs to my hotel and had a full view of the Spanish Steps'

东西果然又贵又难吃,那个面咸到 cry 啊!
'The food, not surprisingly, was expensive and not tasty. The noodle was so salty that I wanted to cry.'

饮料倒是非常好喝!
'The drinks were surprisingly very tasty.'

毛四十块钱就吃这点东西也是醉了!
'Forty Euro for food of such quality *wo-ye-shi-zui-le*!'

面条里是大肉块,肥肉块,,,
'One could find big chunks of meat, fatty meat in the noodle…'

不如把材料给我自己烧[流泪] [6029]
'I would rather have the raw food and cook myself (emoticon: cry).'

(40) 其实我不喜欢澳门
'As a matter of fact I do not like Macau'

我也是醉了
wo-ye-shi-zui-le

同一家餐厅 一入赌场东西就是翻倍的贵
'For the same restaurant the price is doubled once you go inside the casino.'

可是它没有香港的海皇好吃。[3089]
'However, it is not as tasty as the Hai Huang Restaurant in Hong Kong.'

Similarly, there are blog posts complaining about bad weather and traffic conditions when traveling (by car or by plane), some attaching a photo of a fancy car that the user is driving ([5937], [6204]). Users also enjoy flaunting their recent purchases online ([48]), and selfies are also a popular add-on to this kind of post about oneself ([315], [5853], [5937], [5945], [6029]). For instance, in a blog post posted by a model, she seemingly complains about the hard work of the day and is happy that the day is over ([315]), but I presume that this user, along with other users who flaunt their possessions, is only displaying the self that is valued in a middle-class consumer society—the decency of personal appearance, material possession, consumption-driven recreation, even professionalism. In this category, I have also encountered blog posts (in addition to [315]) that "complain" about work-related issues. But underneath, I believe the posters are also portraying themselves as competent working professionals. In (41), the complaint targets certain organizational details of the event, but apparently, as the user himself/herself says, for him/her, the experience was "tiring yet rewarding."

(41) 年会盛典,有多少人默默付出默默工作,
'So many people worked silently for the annual gala.'

节目过筛五审，
'The programs needed to go through five rounds of reviews;'

各种会议、各种流程、各种物品、各种表格，一桩桩一件件,
'Many meetings were involved: procedures of all kinds, things of all kinds and tables of all kinds. It is one thing after another.'

累在其中,乐在其中。
'It is tiring yet rewarding.'

整理参会宾客接送机航班、入住信息到这个点也是醉了,
'I have been working to collect the guests' arrival and departure flight information and hotel reservations, *ye-shi-zui-le'*
真心希望任何一件事情都可以提前规划好,
'I sincerely wish everything would be planned well ahead.'

予已方便就是予人方便 [167]
'Convenience comes in both ways.'

The true communicative intent of such blog posts is to proclaim (without specifically saying it) that "I" am truly a member of the middle class, and "I" am someone you can relate to, and someone you should relate to. But again, nothing of this sort can be found in words in any of the blog posts. Here, again, *ye-shi-zui-le* is used as an evasive device that seeks to conceal the user's satisfaction in being able to live up to the middle-class imagination.

The need for entertainment is also a typical middle-class trait. In the "self" group, various users talk about how they feel *ye-shi-zui-le* about insomnia, and how they deal with insomnia with entertainment ([1860], [1971], [5607]). Entertainment has increasingly become something that the public turns to in their spare time, while also acquiring the role of building social connections. This topic also dominates the content of the "intimate others" group.

By the term "intimate others," I try to capture not only the persons who are socially intimate with the user, but also those the user feels emotionally attached to but does not necessarily socialize with on a daily basis, such as idols in the entertainment business.

(42) 我也是醉了[笑 cry]
wo-ye-shi-zui-le (emoticon: tears of joy)

#全都想睡宋仲基# [2207]
'#Everybody wants to sleep with Song Joong-ki#'

(43) 真人秀层出不穷,
'Reality shows come one after another.'

想玩出新意、得到观众认可实属不易,

‘It is really not easy to offer new ideas and win recognition from the audience.’

#我们战斗吧# 在队员选择、环节设置都很有亮点,
‘#Let’s fight# is well designed in the aspects of team selection, settings of each section.’

今晚井柏然与王嘉尔、萧敬腾再次组队共同寻找线索，完成任务,
‘Tonight Jing Boran, Wang Jia’er and Xiao Jingteng will once again form a team, finding clues together to complete the tasks.’

一路上可谓困难重重,
‘There are many difficulties along the way.’

遇 NPC 强势推销党家字画。
‘When Non Player Character (NPC) try to sell him paintings’

此前一直执着于银行密码的井柏然,
‘Jing Boran, who used to dedicate himself to working with bank password,’

变得会过日子起来,
‘starts to know how to live with a budget.’

讨价还价的样子也是醉了。…… [4743]
‘The way he bargains—*ye-shi-zui-le.*’

(44) #绿袍素因虐恋#
‘#The heartbreaking romance between *Lvpao* and *Suyin#’*

被各种虐恋也是醉了。[922]
‘Being forced upon by all kinds of heartbreaking romance—*ye-shi-zui-le’*

As can be observed from the examples above, the intent to connect is most patent through the use of #topic# functionality on the platform. In (42), the triggering event is “everybody wants to sleep with Song”—a Korean male actor who became widely popular through a romantic TV series, in (43), *ye-shi-zui-le* is a comment about a Chinese male singer/actor Jing, who reveals his ordinary-people side when he participates in a reality show, and (44) comments on the abundance of heartbreaking romance between two characters in a TV series. As someone who does not relate to these users’ blog posts, I

cannot comment on the attitudinal stance they take here, except for the fondness, admiration, or frenzy that fans typically have for their idols. A similar stance is found in blog posts about their truly "intimate others"—family members, such as siblings or offspring. For instance:

(45) 响应网友的需求发几张这两天的照片,
'upon the request of our online friends, I am posting some recent photos'

有两张儿子一定要穿姐姐的衣服,
'two of those show my son in his sister's clothes (he insisted on wearing them himself)'

我也是醉了,
wo-ye-shi-zui-le

不过儿子还是很爷们的我不担心。[4718]
'but my boy is a manly man. there is nothing to worry about'

However, I presume it is exactly because of the presence of the majority of outsiders who can read their blog posts but do not identify with them in such fascination (because the referent is only intimate to the user, not necessarily to the readers, and there is a difference of social distance) that the users resort to the nonverbalization of their stance, to leave it open for interpretation by using *ye-shi-zui-le*, which does not have a definite meaning in this context.

Now that I have reviewed some usage examples based on the categorization of the referent of the event and the attitudes users convey and seek to establish through their blog posts, the function of *ye-shi-zui-le* emerges as a phrase that is against the expression of attitudes. And because the semantic content of *ye-shi-zui-le* does not convey a clear stance towards the event, the attitude in which the entire blog post is grounded needs to be anchored somewhere. I suspect that is the function of the emoticons, which are frequent collocates of the construction.

4.4.6. Sentiment Analysis of the Construction

Sentiment analysis is a method used in natural language processing for opinion mining, which helps identify whether people have positive or negative attitudes towards certain objects or events. Sentiment analysis makes use of dictionaries that code lexical

words with affective valence, sometimes with intensity. There also have been studies that use emoticons to fine-tune the results (Pak & Paroubek, 2010; Read, 2005; Zhao, Dong, Wu, & Xu, 2012). In this section, I will show some rudimentary sentiment analyses based on lexical words and emoticons, and how they support the interpretation of *ye-shi-zui-le* above.

First, lexical words with affective valence are largely absent in the construction. This is consistent with the view that the use of *ye-shi-zui-le* and the bigger YSZL construction, is against the expression of attitudinal stance.

I use a sentiment dictionary (Xu, Lin, Pan, Ren, & Chen, 2008) built for Natural Language Processing (NLP) sentiment analysis for this study. Inaccuracy can come from two sources. First, because it is very difficult to automatically extract only text in the EVENT component, I will instead use all text preceding the phrase *ye-shi-zui-le* (excluding @user, #topic#, [emoticons] and hyperlinks) in a blog post. This could potentially increase hits. Second, because the sentiment dictionary is mostly built from a standard dictionary, it is bound to miss some colloquial expressions and slang. This could potentially create some misses. Still, overall, I anticipate that these two factors will somewhat balance each other out and produce results that can be at least suggestive to the argument.

After several tests, I decided to remove several high-frequency monosyllabic verbs (often used with a neutral tone) that could skew the results.[37] The results based on my algorithm suggest that a majority of items are empty of sentiment words in the prephrasal text. The results show that 81.05% of the items do not contain words of either positive or negative sentiments as listed in the dictionaries, 11.42% of the items contain only positive sentiments, 5.19% contain only negative sentiments, and a mere 2.33% contain both (see Figure 6).

[37] These include *shuo* 'to speak,' and *wan* 'to play' in the negative sentiment dictionary, and *yao* 'to want,' *yong* 'to use,' *xiang* 'to think,' *guo* 'to pass,' *gan* 'to feel,' *cheng* 'to call,' and *jiang* 'to talk' in the positive sentiment dictionary.

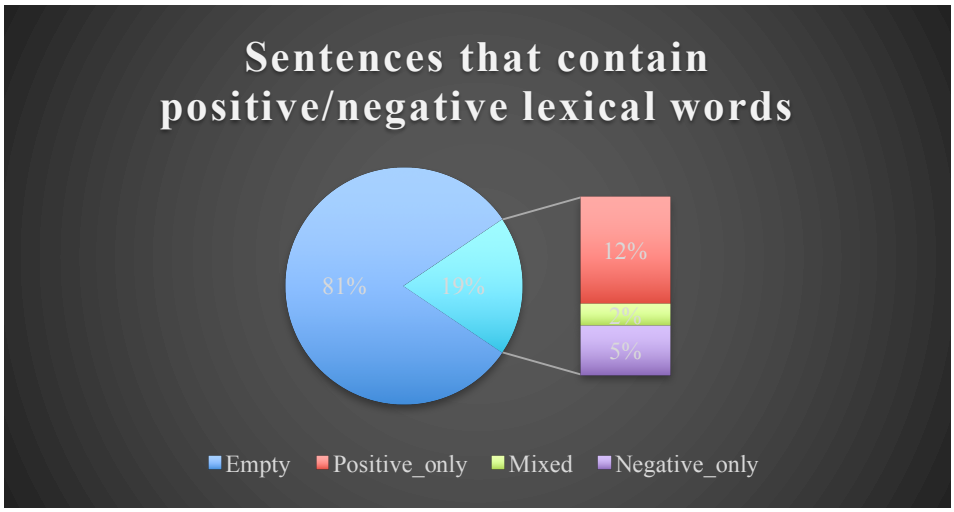

Figure 6. Occurrence of sentiment words in the WEIBO corpus

Even the high-frequency sentiment words (top 20) occur with extremely low frequency in the corpus (see Table 14). In order to further understand how well or poorly the algorithm works, let us look at how the first 10 items in the corpus are coded. The text in the EVENT component has been underlined (examples 46 through 55).

Table 5. Top 20 sentiment words in WEIBO corpus (prephrasal)

好	‘good’ (positive): 268	感谢	‘thank’ (positive): 17
喜欢	‘like’ (positive): 89	怕	‘be afraid’ (negative): 17
爱	‘love’ (positive): 55	愿意	‘willing’ (positive): 16
美	‘beautiful’ (positive): 41	呆	‘stay’ (negative): 15
希望	‘hope’ (positive): 27	赞	‘thumbs up’ (positive): 15
开心	‘happy’ (positive): 26	支持	‘support’ (positive): 15
骂	‘scold’ (negative): 26	谢谢	‘thanks’ (positive): 15
需要	‘need’ (positive): 25	哈	‘Ha!’ (positive): 15
关注	‘focus’ (positive): 24	通过	‘pass’ (positive): 15
堵	‘block’ (negative): 20	舒服	‘comfortable’ (positive): 14

(46) (empty)
开始记录的日子,那就应该把过程都分享给大家。
‘Since I have started documenting, I should share the whole process with everyone.’

上次就提到的排毒茶。Morning 那包是减轻食欲,燃烧脂肪。
'I mentioned a detox tea last time. The "Morning" one is for burning fat.'

上班时候的我保持精神充沛。
'It keeps me in good spirit at work.'

Night 那包是效果是排出体内毒素,还可以提高睡眠质量。
'The "Night" one not only detoxes but also improves the quality of sleep'

这次呢我把购买地址贴在评论第一条吧
'This time I will put the purchase link on the first floor in the comment section'

之前好多妹子找不到也是醉了。[0]
'many gals couldn't find it'—*ye-shi-zui-le.*

(47) (empty)
说的开会,也没有人告诉我是面对面的座谈会啊!
'No one ever told me the meeting is a face-to-face symposium!'

蒙圈了
'I am flabbergasted.'

逼着写微电影剧本,一个月拍出来,要不然通报, 我也是醉了 [1]
'I am now forced to finish a script for a short film, which needs to be completed in a month. Otherwise, I will be reported and criticized publicly.' *wo-ye-shi-zui-le.*

(48) (empty)
在马代最后一天这条微博才发了出去
'I finally managed to send out this weibo post on my last day in Maldives.'

这网络也是醉了 [2]
'The Internet connection here'—*ye-shi-zui-le*.

(49) (negative)
失恋妹子伤不起啊,
'A woman out of love is scary'

公交车边涂指甲油边骂,
'riding on a bus, while doing her nails and scolding'

看的我也是醉了[3]
'I saw this and' *wo-ye-shi-zui-le*

(50) (negative)
我要投诉快递、
'I want to file a complain about the delivery company.'

昨天就到的快递,说昨天就送的,到现在都还没给送,也是醉了。 [4]
'The package had arrived yesterday. It is supposed to deliver yesterday but it has not been delivered even now.' *ye-shi-zui-le*

(51) (empty)
有个这样的女友也是醉了[5]
'Having a girlfriend like this'—*ye-shi-zui-le*

(52) (empty)
这是办公室,不是奇石展览馆,
'This is an office, not an exhibition hall for exotic stones.'

多摆石头就能增旺风水?也是醉了。[6]
'More stones can boost the feng shui?' *ye-shi-zuile*

(53) (empty)
大兄弟,你这么铁公鸡是怎么能把的到妹纸
'Brother, you are so cheap. How come you can get a girl for youself?

交往 3 年才花不到 100 也是醉了[7]
'You guys have been together for three years and in total you spent less than 100'—*ye-shi-zui-le.*

(54) (empty)
【这样的课外读物,也是醉了!】[8]
'Such extracurricular reading materials'—*ye-shi-zui-le*!

(55) (empty)
【济南遇上""修路季""能绕就绕吧】
【Ji Nan encounters "season for road construction": better to detour】

""转了五里山转七里山,转了七里山转英雄山,太添堵了。
'I have to detour from the wuli shan to qili shan, from qili shan to yingxiong shan. So ridiculous.'

""近日,济南的大修路让平时开私家车上下班的市民宋先生十分心塞,
'Recently, the massive road construction in Jinan makes Mr. Song very troubled, who needs to drive to and off work on a daily basis.'

而最让他闹心的是""快看到家门口了,路封了,
'the most disconcerting part is that when he is almost home, the road is blocked.'

5 分钟的路绕了 1 个小时硬是没到家,也是醉了。 [9]
'the five-minute back home takes more than an hour'—*ye-shi-zui-le*.

In (46), the user gives a positive comment about a product ("保持精神充沛," "提高睡眠质量"), but does not use any words that show strong emotions. In (47), "逼" (force) actually carries some adversative connotations, but is missing in the dictionary. "蒙圈了" is a very colloquial expression to say one is shocked and confused, and it is also missing in the dictionary. (48) gives a truthful account of what happened, so the coding is accurate. (49) and (51) are coded negative because of the hits in the negative dictionary, "伤" (to hurt), "骂" (to scold), and "投诉" (to complain), but "投诉" is actually not in the EVENT component. (51) and (54) do not describe an event, but refer to an entity, and they are empty of words of sentiments. (53) actually contains a slang term, "铁公鸡," that the negative dictionary does not catch, but again, it is not technically a sentiment, but since it does not occur in the EVENT component, manual coding will ignore this occurrence. (55) also has several colloquial expressions that have not been included in the negative dictionary: "添堵," "心塞," and "闹心" (all express annoyance), but again, they do not occur in the EVENT component. Overall, it shows that the automatic coding based on dictionaries is prone to miss colloquial expressions, the items that are missed mostly do not occur in the EVENT component, and the overall counting is offset. Again, the results here are only suggestive.

Contrary to the absence of affective words, emoticons occur rather frequently in the WEIBO corpus. Emoticons are found in nearly half of the corpus (41.57%), and 254 unique emoticons are found in total. Danesi proposes that emoticon allow users to "control the emotional features of discourse" (2016, p. 23), and the use of facial emoticons (as well as emoticons) is frequently compared to facial expressions, adding

paralinguistic cues that are lost in computer-mediated communication (Danesi, 2016, p. 21; Schnoebelen, 2012). In my understanding, exactly because the attitudinal stance has been concealed, in both the description of the event and the phrase *ye-shi-zui-le*, emoticons serve the role of anchoring and guiding the reader when interpreting a blog post using the YSZL construction. Figure 7 shows the more frequent emoticons in the order of their occurrence frequency.

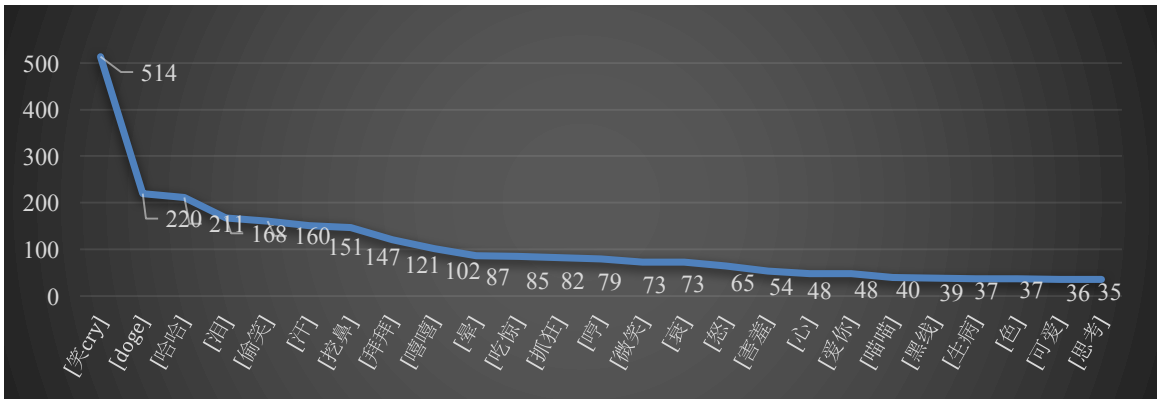

Figure 7. Occurrence counts of emoticons corpus-wide

The most frequent emoticon, found in 7.78% of the items in the corpus, is the "tears of joy" face, as it is coded in English. In the Weibo platform, it is coded as "笑 cry"—or "laughter cry," The same emoticon was elected Oxford Dictionaries' "Word of the Year" in 2015. It is uncertain if its coding in a different language would affect usage, but it seems that in the Chinese cyber community, the emoticon has steered away from the emotion of "joy," and is used to disguise emotion that could be too strong and would seem offensive. The runner-up is the "doge" face, which is frequently used to indicate ambiguity. The face of "surprise," however, is not a particularly high-frequency item. These emoticons can be found in Figure 8 below.

An interesting phenomenon with the usage of emoticons in the WEIBO corpus is that the emoticons tend to occur immediately after the phrase. Figure 8 below shows the relative position of emoticons in the phrase. Most occurrences of emoticons are found unanimously in the R0-R5 position of the phrase (center of the phrase and center of the emoticons are compared), with the L5-L0 window almost completely empty. Their

occurrences at various other distances from the phrase are also astonishingly similar—the higher the frequency, the more similar the pattern they show—almost as if these emoticons have some kind of syntactic rules as collocates of the phrase *ye-shi-zui-le.* While the mechanism of their systematicity is yet to be understood, this could amount to evidence that emoticons act as a repair device to fill in information that *ye-shi-zui-le* intentionally conceals. By revealing their attitudinal stance through emoticons, users can avoid revealing their stance in verbal forms entirely.

4.5. Interim Summary

In the preceding sections, I show how the form–meaning pair of *ye-shi-zui-le* emerges from the YSZL construction. A usage-based approach to language means that frequency of occurrence affects our perception of linguistic patterns, and I thus use quantitative data to construct a prototype of the construction to show what are the central senses of this construction, and how different subconstructions help shape other form–meaning pairs in the same construction. I was able to gain insights into the YSZL construction through the prototype and the constructions to which it is connected. As the quantitative data show, the substantive *ye-shi-zui-le* is embedded in a bigger YSZL construction, whose prototypical form can be represented as (56):

(56) [[EVENT] [(wo) *ye-shi-zui-le*]][emoticon]

The YSZL construction is an instantiation of the topic-comment construction. The first part of the construction (the EVENT component) describes an event, and the second part of the construction is a comment towards the event. More specifically, the EVENT component presents an "unexpected" event, and the phrase *ye-shi-zui-le* serves a pragmatic function accompanying the presentation of the event. The construction also has a frequent collocate—the emoticons, which often occur immediately following the construction. The YSZL construction thus has a function of presenting both information and personal opinion. However, as the usage data show, speakers rarely reveal their attitude towards the event verbally, and even if they do, sometimes through the use of emoticons following the YSZL construction, the attitude is unclear and ambiguous. In my opinion, the ambiguity of personal opinion and attitudes in this construction is the

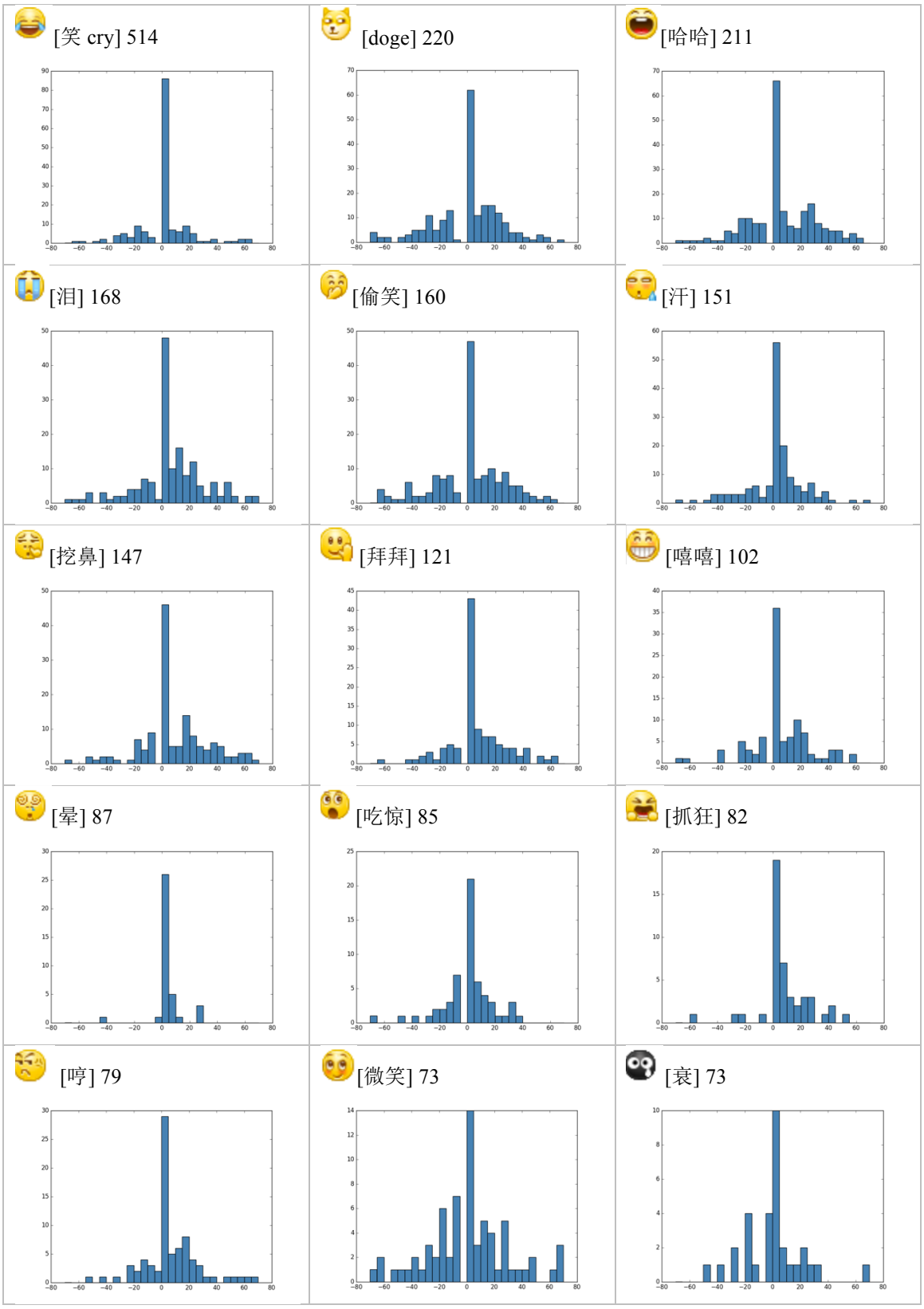

Figure 8. Position of occurrence of emoticons in relation to *ye-shi-zui-le*

function of *ye-shi-zui-le*: it serves as an evasive device to cover the actual attitude towards the event, and it is used to express resistance to verbal expression. Such a usage of *ye-shi-zui-le* cannot be derived from its literal meaning, but is associated with the experience of alcohol intoxication metaphorically. The embodied experience of *zui*, transmitted through the Chinese cultural memory, entails a state of impaired cognition. This is similar to the experience evoked by the instantiations of the YSZL construction. In a state of alcohol intoxication, a person might be unable to properly process an event due to impaired cognitive functions; in a YSZL instantiation, the speaker presents an odd event and is not able to show a clear and definite attitude—due to personal choice or external pressure. The verbal presence of *ye-shi-zui-le* is a metaphor, used as an excuse that exempts the speaker from expression, and it replaces the attitudes that should have been uttered.

4.6. The Social Basis of the YSZL Construction

This construction thus seems full of paradoxes. While there is the desire for expression (posting a Weibo blog post), there seems to be a tendency to conceal some of its content. While there is the presence of emotions (revealed by sentiment words and emoticons), one also puts up the appearance that one is really just uncertain of one's feelings (the lexical meaning of *zui*). All these seeming paradoxes are direct results of the context where this construction emerges—cyberspace, specifically a social media platform. These paradoxes are exactly what Internet users experience daily situated in the social environment and in the communication setting. The emergence of such a construction is a necessary outlet to purge the inner struggles of the Internet users. The structural features of the YSZL constructions have their functional motivations, and the semantic content is congruent with the social reality where the linguistic construction emerges.

4.6.1. The Instantiation of the Topic-comment Construction

The topic-comment construction is an information-heavy construction, meaning that it perhaps carries more information than other constructions. In a topic-comment construction, the topic tends to introduce a new entity to the universe of discourse, or activate a dormant entity created from a previous introduction, and the comment is a

predication on the newly introduced topic. Other constructions, by comparison, fulfill only one of the two tasks: The existential construction typically introduces a new entity, whilst a subject-predicate construction provides information on an existing topic.

The use of the topic-comment construction is thus perfect for Weibo blog posts. A Weibo blog post is a relatively short utterance, and it is created without the presence of a specific interlocutor or a pragmatic environment, let alone a mutual universe of discourse. Having the burden of creating and possibly sustaining a self-sufficient universe of discourse efficiently within the allowed length of 140 characters, the information-heavy topic-comment construction seems to be most suitable for the discourse environment.

The specificity of the YSZL construction, having a schematic component and a substantive component, also allows it to maximally thrive in this particular connective platform. The substantive component, the phrase *ye-shi-zui-le*, allows users to participate in the community event of Internet meming—which, as Zappavigna (2012) observes, is achieved through hashtagging in the Twitter community. According to her, Internet meming creates a sense of bonding and affiliation among Internet users. I believe this is true of most Internet neologisms in the Chinese cyber community as well. The schematic component of the construction, the EVENT component, provides an open slot that encourages personal creativity. It is observed in the data that the event described is personal, or at least immediate to one's surroundings. The YSZL construction thus seems to possess the essential qualities of a good social media blog post: having original content, while appealing to the interests of the masses.

The content enabled by the construction also helps the users to achieve connectivity on the platform. While blog posts are published in the form of monologues, users have anticipated that they will invite other users into their universe of discourse. They present an event, and then a personal opinion, resembling the elements of modern-day media, in the hope of creating controversy and drawing in discussion. The YSZL construction also tends to be used to describe an unexpected event (or an uncommon entity, usually with an attachment of a picture or video links). I believe this seeks to create a novelty effect and draw attention. Attention is important, because as Goldhaber (1997) proposes, it has become a valuable asset, a currency in a sense, that can be converted into money. Celebrity is a good example of those who monetize attention.

Goldhaber also remarks that "since it is hard to get new attention by repeating exactly what you or someone else has done before, this new economy is based on endless originality, or at least attempts at originality." The description of an unexpected event is thus an act of convenient formulaic creativity enabled by the construction. It can be linked to one form of the economic base of our society. And since the advent of connective technology has created an opportunity (possibly illusory) for everyone to have a channel to gain and monetize public attention, it is only natural that social media users seek novelty when publishing content on the platform.

4.6.2. The Expression/Suppression Paradox

The most puzzling part about the YSZL construction is probably in the paradox of what the user really desires to express. I believe the construction is inherently paradoxical and self-conflicting. This is also probably the most interesting aspect of this construction, and I suspect such a distinct an apparent conflict has a social basis.

First, while social media is about self-expression, it is at odds with the suppression of emotion and opinions, which is part cultural and part structural in China. To a certain extent, such a value has been enforced in cyberspace in language use. For instance, profanity is largely censored on major (if not all) local platforms. The Chinese netizens have developed strategies to bypass censorship by introducing euphemisms, such as using homophones (e.g., 尼玛 in place of 你妈) and acronyms (e.g., 'TM' in place of 他妈). At one point, when netizens realize that whatever intensive statement they make, their language will be censored, they stop enunciating these strong emotions and strong opinions at all. And *ye-shi-zui-le*, conveying an unwillingness to speak, is definitely associated with the dilemma of language censorship that they encounter daily. Meanwhile, the integration of emoticons on the Weibo platform provides a legitimate channel of expression of emotion. I assume this is why emoticons are often appended to *ye-shi-zui-le*, because the users are actually expressing their emotion in ways that are not permitted via language.

One would assume that the strong emotion that is concealed in the construction is mostly negative, but that is probably not the case. The sentiment analysis did not find a lot of blog posts revealing sentiments, although there are more blog posts with negative

sentiment lexical words. It is understandable if people are hiding negative sentiments. I notice that postings on the topics of traffic and work (though only about 3% respectively in the corpus) in general contain negative sentiments.[38] The suppression of negative sentiments is also understandable in work-related posts: As working professional, one's public image can be harmed by openly saying negative things about one's work. For traffic, there is simply no point in saying anything about it, because it is a pain of urbanization that most cities in China are going through right now. But what about positive sentiments? A quick perusal of the blog posts reveals that a lot of blog posts are engaged in the unvalued activity *shai*. *Shai* (晒), originally 'to air the laundry in the sun,' does not have the meaning of 'air the dirty laundry' in the Chinese context; instead, it refers to the act of 'showing off the personal treasure.' In our data, Weibo users show off their fine dining experiences, vacation trips, shopping hauls; all are associated with a demonstration of wealth, something that is still not much appreciated in the Chinese culture. If one is tempted to do so, then one can at least try to conceal one's excitement and pride associated with these spending behaviors. In the WEIBO corpus, I find 12.57% of the blog posts are on food, 9.27% on travel, and 7.82% on shopping.[39]

In essence, I attribute the inherent paradox of this construction to the disorientation induced by the blending of public and private spheres taking place on the social media platform (and again, this sense of disorientation can be associated with the embodied experience of alcohol intoxication). The connectivity of cyberspace poses it as a public sphere, and the open social media platform Weibo is even more so. However, the platform also encourages users to post personal content. They are suddenly compelled to expose their personal life for the scrutiny of hundreds of millions of viewers. The platform thus creates a completely new space, in the sense that it is pseudopublic and pseudoprivate, where conventions of communication are to be defined. Internet regulations are stipulated to guide some aspects of these public communications, such as the enforcement of profanity censorship mechanisms on various platforms. However,

[38] The topic search was guided by lists of keywords I created for each topic. See code in topic.py

[39] Topics were identified by dictionaries of topic keywords. For instance, key words for the topic of "food" would be "eat," "restaurant," "rice," "dinner," etc.

with communications at a more personal level, it is up to the users to decide to what extent they are comfortable being public or staying private. I believe the contest between the desire to express and the urge to suppress is a product of the process of seeking a comfort zone in social media communication.

4.7. A Linguistically Correct View of Internet Neologisms

Every year, an unofficial election is held to select the top 10 Internet neologisms of the year, because it is believed that they can reflect the social issues and important social events of the year. Whenever a new Internet neologism emerges, Chinese netizens fervently take it up and participate in the carnivalesque event of Internet meming. Meanwhile, there are concerns that Internet neologisms will tarnish the beauty of the Chinese language, and hold the view that they should be "crashed," and be forbidden in the written language.[40]

I might not be speaking of all Internet neologisms, but at least in the case of *ye-shi-zui-le*, this chapter has shown that not only does it have linguistic roots in our common linguistic knowledge, but also it has a cultural and social basis for its emergence. Due to this convergence, not only has it stayed in the Chinese language two years after its emergence, it is transforming other linguistic constructions. To me, a linguistically correct view of Internet neologisms is that they should not be treated with bias on the grounds of vulgarity. After all, languages evolve in their own ways. An Internet neologism might go out of fashion sooner than there is any plan to ban it, or it might thrive despite all the adversity.

4.8. Summary

In this chapter, I have examined the usage data of the Internet neologism *ye-shi-zui-le* and proposed that it does not emerge as an individual phrase, but in a discourse construction, the YSZL construction, which has the form [[EVENT] [(*wo*) *ye-shi-zui-le*]. The meaning of the phrase, to denote a state of cognitive impairment, is a metaphorical extension of *zui*, or the experience of alcohol intoxication. Yet its function as an evasive

[40] http://www.qstheory.cn/culture/2016-09/20/c_1119592688.htm; http://www.cac.gov.cn/ztzl/xzt/39/zt/index.htm

device, to purposefully not express a personal attitude, emerges in its relation to the other component, the EVENT component. In the data, I found "unexpectedness" as a reoccurring semantic category of the EVENT component, which triggers the metaphorical meaning of *zui*. The events described in the EVENT component, however, show clusters of usage contexts that encourage the verbal suppression of a subjective evaluation towards the event, which is at least partially caused by the structure of the communication platform Weibo. A majority of my data shows that *ye-shi-zui-le* occurs in context, anchoring to certain types of event description. But there are also cases where it occurs in isolation, which can only mean that this is a more recent development as it gains independence in the discourse as its meaning and function stabilize through its usage in a construction (yet it is still anchored to an implied context that is not verbalized).

Because the data is not diachronic (in the sense that it does not include data predating the constructionalization of YSZL), I cannot say definitely whether the linguistic innovation is truly catastrophic. But by looking at the overwhelming uniformity of data instantiating the YSZL construction, and the consistent form-meaning pairs of the phrase *ye-shi-zui-le* (not having even one case to denote actual alcohol intoxication in the sample corpus), I doubt there were any in-between stages, or constructional changes that are widespread enough to say the constructionalization was a gradual formation. For most language users who learn the construction, they learn it after they have only seen a few usage instances on social media platforms; and because most usage instances follow the constructional frame, each of them could make good exemplars to enable generalizations. n individual speakers, and likewise in the whole linguistic community, the knowledge of this construction comes almost instantenously. In this sense, the linguistic innovation of the YSZL construction is abrupt.

In the case of the YSZL construction, I see the external factors having a bigger role in its emergence, and in particular the social basis of its emergence I discussed in section 4.6, which can be summed up as the emergence of an open social media platform. The experience as a user of such a brand new type of communication creates new needs of expression, hence new means to express. The employment of the topic-comment construction, the "unexpectedness" in the EVENT component, the use of *zui* to not

express personal opinions, are all material internal to the linguistic system, that are evoked by external factors to serve a new function in a new construction.

.

CHAPTER V
EPILOGUE

In this dissertation, I have analyzed three cases of language change in Chinese. Chapter II discussed the emergence of the new grammatical category of numeral classifiers within the QUANTIFICATION construction in oracle-bone inscriptions and bronze inscriptions. Chapter III examined the new *bei* construction that became popular in cyberspace and posited that it is a development from the *bei* passive construction. Chapter IV introduced the YSZL discourse construction in connection with the Internet neologism *ye-shi-zui-le*. These three case studies have common implications that could further our understanding of language change.

First, regarding the unit of change, all of these cases point to constructions as the basic units of change. In Chapter II, I proposed that the emergence of the numeral classifier as a new grammatical category did not occur in isolation; instead, it is a new type of quantification unit that emerges in the QUANTIFICATION construction. The emergence of numeral classifiers happened in conjunction with the emergence of other quantification units. The category 'numeral classifer' emerged in conjunction with certain quantification units and in relation to the noun which it quantifies. Individual units of a noun within the [Noun Numeral $\text{Unit}_{\text{numeral classifer}}$] were the basis for the generalization that lead to the emergence of the numeral classifier category, indicating that the unit of change is not a single classifier but the entire NUMERAL CLASSIFIER construction. In Chapter III, similarly, the emergence of the new *bei* construction occurs not only within the unit of "*bei* X," but it occurs in the clausal construction [N_{person} "*bei* X"], which has idiosyncrasies both within the subject slot and X. In Chapter IV, I examined the emergence of the Internet neologism *ye-shi-zui-le* and identified its emergence in a discourse construction, within which a preceding component of the phrase provides context and anchors its meaning.

Second, regarding the manner of change, all of these cases provide evidence that the emergence of new linguistic constructions can be abrupt rather than gradual. In Chapter II where diachronic data of a certain capacity were available, the usage data tended to show this language change was driven by high-frequency exemplars. In

particular, I presumed the counting of captive Qiang people in the form of '*qiang* (numeral) *ren*,' and the counting of chariot-drawing horses in the form of '*ma* (numeral) *pi*,' were two exemplars that form the conceptual basis for numeral classifiers. Both of these two concepts have cultural importance that prompted their high-frequency of occurrence. Although neither of these numeral classifiers can be considered prototypical by contemporary standards, they were quite different from any other quantification units in their time period. In this sense, their abrupt appearance gives rise to a catastrophic language change—a new quantification unit, and a new grammatical category with its own pattern of usage.

Synchronic data from the two case studies in modern Chinese also indicate that these new linguistic constructions were introduced into the language through high-frequency exemplars, which orient language users' perception towards the new linguistic pattern. In the case of the new *bei* construction in Chapter III, it is possible that a single innovation *bei jiuye* 'being employment-ed' set off the creative interpretation and reproduction of the pattern, and its meaning of MISREPRESENTATION was extended to different kinds of IMPOSITION through its instantiations in a variety of usage contexts. The emergence of the MISREPRESENTATION construction, as the extended period of synchronic data shows, is not comprised of small-step gradual changes. As a development of the *bei* passive, it exhibits specialized meaning and a function that cannot be found in the generic meaning of *bei* passive. The new *bei* construction is an abrupt innovation, even as a subconstruction of the *bei* passive. In the case of the YSZL construction in Chapter IV, because it is a discourse construction and has uneven degrees of schematicity within its components, its entrenchment was not accomplished by the propagation of a single usage instance. Having both a schematic component and a substantive component gives the YSZL construction an advantage, because it has a fillable template allowing it to achieve maximal productivity (type frequency) through the schematic component, while ensuring formal stability (token frequency) through its substantive component. In this case, the exemplar of the construction is not a single instance, but the combination of substantive and schematic components—the phrase *ye-shi-zui-le,* which makes the construction identifiable and comprehensible, along with a loosely schematic component that describes an unexpected event. Eventually, the

instantiations of YSZL plague Chinese social media, and the resulting high frequency of usage allowed it to quickly gain currency among Chinese Internet users. Overall, the three case studies in this dissertation lend support to an exemplar-based model of catastrophic language change. By catastrophic, I am not identifying the term as proposed by Lightfoot (1991, 1997), which argues for a catastrophic change occurring as a result of the resetting of parameters in the Universal Grammar (UG) when children pick up a robust cue that points to an alternative structure. Instead, my data indicate that a new construction can emerge even without the presence of robust cues (given by sufficient type frequency), and it can occur within a single generation of language users—as Chapter III and IV indicate. Whether the next generation will process the innovations differently and whether they may endure and persist in the language is yet to be determined.

Third, considering the causes of change, external factors are crucial to language change. All three case studies in this dissertation have shown that external factors help reconstruct language change in progress, especially in the emergence of exemplars as a response to changing material reality. In Chapter II, archeological evidence of the change of chariot structure suggested that there is a material basis that calls for the development of an individuated quantification unit for horses and the Qiang people. In Chapter III, being a witness of the language change in progress, I enjoyed a vantage point to understand the imposed events of "*bei* X" in its entirety, and came to the conclusion that these events have a large-scale presence. In Chapter IV, similarly, I also observed that several of the typical usages of the YSZL construction serve to deliver content that are promoted by a social media platform, for instance, content sharing to amuse and bond with the virtual community.

To view it from a different angle, one might say that social reality and the bigger social context have at least partially motivated language change in these instances. In Chapter II, the emergence of numeral classifiers could be associated with the commonality of warfare amidst the formation of nation states, a process that is accompanied by a burgeoning awareness of the accumulation of wealth and the value of transferable objects; there was hence an increased need for accurate counting. In Chapter III, new *bei* came into wide circulation because of the prevalence of social issues that

appeal to public attention, and the general presence of institutional discourse and institutional structures that evoke a sense of "being imposed upon" among the general public. In Chapter IV, I proposed that social media users go through the experience of disorientation in a metaphorical way, including the inability to expresss, impaired judgment, and cognitive function, because the the intrusion of social media platforms into private life has blended their public and private lives, and this blending upsets their sense of orientation in ways that they have never experienced before—except in a state of alcohol intoxication. In short, from the usage instances, I was able to identify the motivation of using, or even creating, the new construction as embedded in the existing or emerging social reality.

Additionally, the medium that carries language and text is another external factor that affects or, in these cases, promotes and consolidates the development of new linguistic constructions. In Chapter II, oracle-bone inscriptions and bronze inscriptions represent some form of official discourse. The promulgation of the use of numeral classifiers in this specific medium might have had a trickle-down effect that permeated throughout different genres and registers of language use. In Chapters III and IV, the language change in focus took place in cyberspace. The speed and connectivity of the Internet ensures that not only are new linguistics constructions delivered to every connected Internet user instantly, but users are also exposed to their instantiations much more frequently. The heightened frequency of usage as experienced by every individual user has greatly shortened the time span it takes for a new linguistic construction to stabilize, and it accelerates the process in general.

In this dissertation, besides engaging with the theoretical discussion on the process of language change (whether it is gradual or catastrophic), I also have tried to propose a methodological consideration when investigating language change—that is, the consideration of external factors. In linguistic research, the concern for external factors can rarely be found beyond work by variationists and work on creoles and pidgins, but it is my belief that it should be more widely employed. I examined emerging linguistic constructions in a scenario where language contact is not relevant, and the external consideration is not about the influence of other languages. Instead, I showed that the integration of external factors in the sense of the social context of language use is highly

informative. As I have argued previously, one finds in external factors not only evidence that helps to construct language change in progress, but also sheds light on the social motivations that drive the emergence of new linguistic constructions. Such a connection between language structure and language use is sanctioned by the usage-based approach to language, found in at least some types of Construction Grammar, which posit that linguistic patterns (the systematic mappings of form and function) emerge in usage contexts. In other words, generalizations over many instances of usage map certain linguistic functions onto certain linguistic forms. Therefore, in the case of language change led by exemplars, the context of occurrence of these exemplars became the context of emergence for new linguistic constructions. The emphasis on context as a factor of language change is something that has been validated by previous scholarship (see, for example, Traugott & Trousdale, 2013), but consideration of the broader context of occurrence as an external factor that motivates language change is something the current study seeks to advocate.

To conclude, the current study adopts a constructional approach in the investigation of language change. Concentrating on three case studies of language change in Chinese, I show that language change can be led by exemplars and can be catastrophic rather than gradual. Catastrophic change is achieved due to both internal and external factors. In the emergence of a new linguistic construction, a network of constructions grants language users access to the constructional knowledge of existing constructions that are related to the new construction—this is the conceptual basis for the comprehension of new constructions internal to the language system. Additionally, their usage contexts allow users to form generalizations towards the meaning and function of the new construction—the content of is provided and enabled only when incorporating the external dimension of the language system, a.k.a. language use. With a constructional approach, language change can thus be evaluated both with internal factors and external factors, giving us a comprehensive understanding of the process.

APPENDIX A

X IN MISREPRESENTATION "BEI X" INSTANCES

Instance		Definition	Token Frequency
jiuye	就业	'to be employed; employment'	25
jingshenbing	精神病	'mental illness'	13
zengzhang	增长	'to increase'	12
zisha	自杀	'(to commit) suicide'	12
daibiao	代表	'to represent'	11
juankuan	捐款	'to donate money'	7
xingfu	幸福	'to be happy; happiness'	7
jiujia	酒驾	'to DUI'	3
shizong	失踪	'(for a person) to go missing'	3
xiaokang	小康	'to be moderately prosperous'	3
daiyan	代言	'to endorse a brand'	2
fumian	负面	'negative (reviews)'	2
huaijiu	怀旧	'to be nostalgic'	2
jianqiang	坚强	'to be strong'	2
kaixin	开心	'to be happy'	2
lihun	离婚	'to divorce'	2
manyi	满意	'to be satisfied'	2
piaochang	嫖娼	'to pay for prostitution'	2
qushi	去世	'to pass away'	2
shishi	逝世	'to pass away'	2
shouru	收入	'income'	2
beitong	悲痛	'to grieve'	1
cuangai	篡改	'to alter'	1
diyi	第一	'to be the top one'	1
gaoshang	高尚	'to be noble'	1
gaoxin	高薪	'decent pay'	1
hanguo	韩国	'Korea'	1
ji'e	饥饿	'hunger; to be hungry'	1
jinu	激怒	'to infuriate'	1
kuanrong	宽容	'to tolerate'	1
lingshi	零食	'snack'	1
qidai	期待	'to expect'	1
tigao	提高	'to improve'	1
yiliuhua	一流化	'to make … the best of its kind'	1
youxiu	优秀	'excellence; to be excellent'	1

zhangxin	涨薪	‘to raise pay’	1
zhenjing	震惊	‘to shock’	1
zhongchan	中产	‘middle class’	1

APPENDIX B

X IN IMPOSITION "BEI X" INSTANCES

Instance		Definition	Token Frequency
gaotie	高铁	'high-speed rail road'	4
baohu	保护	'to protect'	2
fuwu	服务	'to serve'	2
xiaofei	消费	'to consumer; to spend'	2
xiaosan	小三	'mistress'	2
xiguan	习惯	'to get used to; habit'	2
zhangjia	涨价	'to raise prices'	2
aizi	艾滋	'AIDS'	1
baotuan	抱团	'to form a league'	1
baoyang	包养	'to keep a mistress'	1
biaoyan	表演	'to perform'	1
biaoyu	标语	'slogan'	1
buke	补课	'to make up lessons'	1
chaiqian	拆迁	'to relocate'	1
chengming	成名	'to become famous'	1
chuming	出名	'to become famous'	1
fangjia	放假	'to observe holidays'	1
fangnu	房奴	'a person that carries a mortgage'	1
fangui	犯规	'to foul'	1
fu	富	'to be rich'	1
gaige	改革	'to reform'	1
gongli	功利	'to be practical'	1
guanxin	关心	'to care'	1
hedao	喝到	'to knock out someone'	1
hengfu	横幅	'banner'	1
hexie	和谐	'to be harmonious'	1
jiaban	加班	'to work overtime'	1
jiasu	加速	'to accelerate'	1
juanzeng	捐赠	'to donate'	1
juhui	聚会	'reunion'	1
kuasheng	跨省	'to cross provincial boarders'	1
lianmeng	联盟	'to form an alliance'	1
maifang	买房	'to buy a house'	1
pingyong	平庸	'to be ordinary'	1
qiangjiu	抢救	'to resuscitate'	1

qianyue	签约	‘to sign a contract’	1
shangke	上课	‘to attend classes’	1
shiyan	实验	‘to experiment’	1
shiye	失业	‘to lose one’s job’	1
shoufei	收费	‘to charge a fee’	1
si	死	‘to die’	1
tianjia	天价	‘to be overpriced’	1
tongchou	统筹	‘to coordinate’	1
weibo	微博	‘Weibo’	1
wupan	捂盘	‘to withhold from market’	1
xiangqin	相亲	‘to go on blind-dates’	1
xinyang	信仰	‘to have faith in’	1
xinyongka	信用卡	‘credit card’	1
xiujia	休假	‘to take a break’	1
yinmou	阴谋	‘conspiracy’	1
Zanzhu	赞助	‘to sponsor’	1
Zaofu	造富	‘to make someone rich’	1
Zhengce	政策	‘policy’	1
Zhijing	致敬	‘to pay tribute to’	1

APPENDIX C

YSZL SAMPLE CORPUS

Index	Blog post
48	Play Boy X Joyrich SS16 自己自留了一大堆也是醉了![色][色][色]
167	年会盛典,有多少人默默付出默默工作,节目过筛五审,各种会议、各种流程、各种物品、各种表格,一桩桩一件件,累在其中,乐在其中。整理参会宾客接送机航班、入住信息到这个点也是醉了,真心希望任何一件事情都可以提前规划好,予已方便就是予人方便
315	收工什么的咀开心啦[噢耶] 今天造型换的我也是醉了[笑 cry]
375	也是醉了,那么有钱 http://t.cn/RyhYH5S
428	【如果说"扒光衣服"、"围殴脚踢"、"下跪掌嘴"等霸凌行为都可以被容忍,那么我也是醉了!!!】 校园霸凌愈演愈烈的根源性问题在于仅把"辱骂"和"伤害"等暴力行径视为"恶作剧之类的小事",这也是此类事件始终无法得到遏制的真正原因。http://t.cn/RGVN1w5
469	上次回家是大雪,这次回家是大雾[笑 cry]高速公路全部封路了!我也是醉了。 http://t.cn/R2Wx6nM
559	大年初三老姐看韩剧 我也是醉了[微笑]
575	#越界有理#这画风也是醉了,话说回来从关爱女性的角度看待安全套,看来ABC 是真的要越界出安全套啦?@ABC 官方微博
650	刷了一天微博,也八卦了一天,对于昨晚的几个吐槽点,最终得一个结论,年度心机婊非他莫属,我不知道他叫啥名,竟然假唱都可以把自己唱哭了,还感动无数的猪!我也是醉了……看他微博还不让批评,全是机器人刷评[吐]还自我陶醉.....真是个奇葩.....哎...这个世界充满的谎言,很可怕!晚安
698	假唱对不准也是醉了,我的尴尬症[doge]
729	充电速度慢成这样也是醉了,充了 10 小时的电[撇嘴] http://t.cn/RU1yYYv
776	尼哥的袜子也是醉了!真是骚气
778	我堂妹 3 月要生了,超出来估计是个女孩儿,还没生呢婆婆就说过几年再生一个吧,也是醉了,就好像再生一个肯定是儿子似的,生男生女孩子发展怎样性格如何都在自己培养,这种老旧思想觉得生了女的一定吃亏,生了儿子就占便宜似的…就好像姓氏多珍贵似的还得传下去,不都是普通人吗姓啥有神马区别[拜拜]
922	#绿袍素因虐恋# 被各种虐恋也是醉了。
935	宝宝出生三个月,月嫂走后,婆婆来做中饭,拖下地,就回家了,也不怎么照顾宝宝,晚饭都是公公来做的,她说我应该付她工资 3000 一个月,说再给我们当保姆,也是醉了也是醉了
956	饿的不要不要的了,起来煮碗面,然后看傻狍子看到这个点儿,我也是醉了!哈哈哈哈太可爱了[笑 cry][笑 cry][笑 cry][笑 cry][笑 cry]

966	有人要红了![偷笑]《快乐星挑战》2015 年度总决赛直播,导演在后台边切换边自嗨,看完视频我也是醉了~不愧是根正苗红的文艺综艺队伍![给力][给力][给力][good][good][good]@广西综艺快乐星挑战 http://t.cn/RbdPgi5 .
1010	也是醉了!飞机上做一个小时还不起飞。
1077	#路飞小猴看世界#纽约一个小哥把化妆这事玩出花了。。这画风也是醉了[晕]
1143	#吃是治愈之王#一个人干一只我也是醉了减肥这件事留到年后吧
1196	生理期都被弄乱掉了,碰上墨迹的人也是醉了也是醉了
1298	【春晚吐槽】前不久刚刚过去的猴年春晚可谓是槽点满天飞啊,小编也为大家整理了一下小伙伴的吐槽,也是醉了。[doge][doge][doge]
1466	哈哈,遇到这种同学也是醉了!【【彩虹之音】合租同学看见我跟男人钻在一个被窝里,然后打电话告诉了我妈】 http://t.cn/RGQunsQ
1520	醉不得。。。。这个老砍疯不得。我也是醉了 http://t.cn/R2dbAsp
1584	一个大度大气的局长遇到一个私欲膨胀的党组书记,这组合也是醉了。
1660	印度为了防作弊 测验时所有脱光 所有赤裸裸应对作弊也是醉了(组图)为了能让本身的孩子顺遂测验通关,印度家长们"外挂"讲授楼墙壁向科场递纸条的新闻一度成为热门话题。若何治理测验作弊,印度军方则亮出了本身的狠招。-悠悠网 http://t.cn/RGlpyrG
1667	死活不推送也是醉了[泪][泪][泪]
1699	好疼的一个游戏,到最后才懵逼也是醉了。。 http://t.cn/RGmGpRn
1860	夜晚一个人看最强大脑心算视频我也是醉了,呵呵。
1971	越疲惫,反倒是早早就醒来睡不着,四点多爬起来写字,我也是醉了。。[喵喵][doge]
1990	#山海经蓝盈莹逆袭# 演的挺好,神烦黄俐啊!但是居然有辣么多出不了戏的人也是醉了难道不比 30 几岁只会演傻白甜的女演员好多了
2207	我也是醉了[笑 cry]#全都想睡宋仲基#
2346	@龚文祥 10w 抽奖是真的,抽奖结果也是公平公正的[doge]看下评论一堆人说内幕什么的,我也是醉了…这个不存在内幕,设置条件都摆在那里了,用的是平台抽奖,要是不公平,那也是新浪抽奖平台的问题。想起我一个朋友,拿钱做抽奖,每次都拿平台开,结果每次没抽到别人就有人说假抽奖,然后他就不玩了。
2425	看了你们装逼的颜值,我也是醉了.。。。这个视频真的有毒 http://t.cn/R4cFdWg
2652	这年头弱智的浓度你是没有办法理解的,我某次演讲说了一句非常自谦的话,居然被人砍掉一半,当作我装自己牛逼的话了,唉。我也是醉了。
2664	[#沫友吐槽#]一个多小时才叫了两个号,这个办事速度也是醉了…… by@_阳光下的露珠
2667	高手神闪避~这走位,神乎其技,看的我也是醉了。

2674	怎么请一个假都那么的难呢?都说了回家得搭 5、6 个钟才可以回到市区。好朋友结个婚就请那么一节课就那么难吗?我也是醉了。 上午上完课再出天河搭车,不塞车也是要三更半夜回到市区,人生地不熟,又没有人来接,发生什么事是不是你可以负责的,难道一张请假条纸就那么的贵重吗?
2733	这首歌的歌词我也是醉了[doge]
2745	这一晚上把人焦虑的,也是醉了,别来招惹我,我烦着呢。
2798	男朋友情商低是什么体验? 那叫一活宝,哈哈!遇见也是醉了[嘻嘻] http://t.cn/RqpkVqw
2877	成宿做梦我也是醉了 4.22 梦境:我被下药了,类似大麻之类的,然后就被朝阳群众举报了,然后我被拉去化验,遇到了陈乔恩! 然后陈乔恩跟我说 别害怕,巴拉巴拉 ,然后我有惊无险的被放出来了,陈乔恩说我小时候救过她所以送我一辆玛莎拉蒂! 然后我开着玛莎拉蒂去接我...全文: http://m.weibo.cn/2105321423/3967352078657845
2961	今天带着小宝贝儿逛了一天,也真的够累的[悲催][悲催]如此逗逼的三个麻麻也是醉了也是醉了
2983	这么大的雨我也是醉了[围脖] http://t.cn/RUHN7dX
2986	我也是醉了,我咋吃不胖?[哈哈][哈哈][哈哈]@蓝雨范范 [哈哈]@陵前创想电脑科技 @罗玉凤 http://t.cn/RyCyY7v
3027	#蜀黍说安全# 看看国外的路怒,也是醉了[衰]http://t.cn/RqoZc0O
3089	其实我不喜欢澳门 我也是醉了 同一家餐厅 一入赌场东西就是翻倍的贵 可是它没有香港的海皇好吃。 http://t.cn/8kkzh7L http://t.cn/RqSdxCG .
3109	是喔,认真的人自带光芒。比如看球赛的我哥,刷微博的安某人。 早饭都还没吃。[晕]一大早被借钱转了账才怕被骗,也是醉了。哎.....
3211	喝咖啡撒葱花的年代,丈母娘也是醉了。 http://t.cn/R2dy8eB
3224	真的也是醉了,他的每本专辑我都会去听 [拜拜][拜拜][拜拜]
3339	把我拍成这样,也是醉了,#论男友会拍照的重要性##今日贴纸打卡#
3368	最近爱上素颜臭美,皮肤太好 [害羞][害羞]最后一张笑死我了,有逗逼朋友我也是醉了 http://t.cn/R2WigYP
3563	周末凌晨就开始跟人撕逼,我也是醉了……不发表预设立场的观点,但也实在忍受不了智商低过下限太多的脑残。
3942	我也是醉了[哼]这鱼被我弄的粉粉碎 成鱼酱了[笑 cry][笑 cry]
3948	#王凯征服栗坤#我也是醉了,帅气死了。 [din 癫当阿赞圣诞奔][din 癫当阿赞圣诞奔][din 癫当阿赞圣诞奔]
4058	#出租额头广告位#我也是醉了,怎么看不懂?[睡]这样的真的是太有挑战性了。
4086	我们居然看见蟑螂还能淡定的洗碗,并且讨论蟑螂的大小个数,也是醉了。
4121	PS:拿六年前的概念车图片也能上头条,我也是醉了。 //【隔壁村村长要买辆劳斯拉斯 村们纷纷围观 小王笑而不语】http://t.cn/R5GBqrp (分享自 #今日头条#)

4179	#那样的颜这样的歌#郑凯,你确定你是在唱歌的吗,我也是醉了。我完全被黄教主的歌声惊到了,告诉我这不是真的
4397	坐地铁被检查身份证,过检查站还被检查身份证,我也是醉了也是醉了
4454	#欧洲杯#足球改规则了?战术创新?也是醉了!两只软脚虎,除了撕扯没给这场比赛留下其他"精彩"。 http://t.cn/R2dUf7C
4477	征文:真的有德隆球迷 (1) @yuimokin 理由也是醉了。 http://t.cn/R5I9DTX
4718	响应网友的需求发几张这两天的照片,有两张儿子一定要穿姐姐的衣服,我也是醉了,不过儿子还是很爷们的我不担心。
4743	真人秀层出不穷,想玩出新意、得到观众认可实属不易,#我们战斗吧# 在队员选择、环节设置都很有亮点,今晚井柏然与王嘉尔、萧敬腾再次组队共同寻找线索,完成任务,一路上可谓困难重重,到遇 NPC 强势推销党家字画。此前一直执着于银行密码的井柏然,变得会过日子起来,讨价还价的样子也是醉了。答题环节...全文: ...
4751	林大叔愤怒吐槽:一大早,林大叔就被十个八个电话吵醒,几乎全部都是骚扰电话或是推销电话,唯一一个不算的是快递员的送货电话。现如今家人朋友客户熟人联系你,基本上都是用微信了,语音通话只能大部分都是骚扰推销甚至是诈骗电话。这年头,经常打电话联系你的只有快递员外卖员,也是醉了,也算悲哀。
4759	【也是醉了, 街道被葡萄酒淹了】 当地时间 8 月 3 日夜间,一伙不明分子在法国东南部离马赛不远的小镇赛特,将装在五个酒桶里的数千升的葡萄酒打开倒掉,造成市内街道葡萄酒泛滥。洪水般的葡萄酒涌入街道,同时渗入大楼地下室和地下停车场。消防员们不得不冒着醉酒的危险在充满酒香的街道上清理淤积的葡萄...全文: ...
4776	对于加班到现在的我也是醉了,最主要的是还没收获。狗屁制度。
4975	#超女套路#超级女声真是套路太深,竟然现场改赛制,评委选不出人让队长自己选人进终极 PK,搞得蓝队队长黄汐源被迫选自己,换谁谁都不敢选队友吧,赛制那么容易改威信何在,评委不选人是来打酱油的么[微笑]也是醉了[拜拜]
5002	#来马栏山抓皮卡丘#皮卡丘就是我的最爱,超级喜欢他们啊哈哈,所以风油成精,我也是醉了。
5039	我也是醉了,今早 5 点为止一眼没合,。祝我小表妹 12 岁生日快乐[爱你][爱你][爱你]。 http://t.cn/RyhHbzL
5054	#贵阳日爆#哈哈 ,贵阳惊现妖娆哥,这烤鱿鱼的姿势也是醉了![哈哈][哈哈][哈哈]http://t.cn/Rt5JJdj
5110	一滴都不能少,结局也是醉了[笑 cry]#笑弯了别怪我# http://t.cn/Rt4nSMP
5239	CHLOE 深梅紅豬包, 中號....(歐洲叫 SMALL) **聽 MM 說有國內 SALES 叫這大號, 我也是醉了** 這個超優惠 9380 到手~~~ [色]
5267	【【大韶短评】】台湾媒体的逻辑也是醉了。 http://t.cn/Rt0J8Qt (分享自@凤凰新闻客户端)

5271	真是服了,大半夜的貌似别人用了"呼死你"软件,也不显示号码?都不知道为什么 也没得罪人啊?我也是醉了[怒][怒]这种情况怎么办?报警管用吗?感觉自己被盯梢了,挺吓人的!@齐鲁晚报 @青岛交通广播 FM897 @半岛都市报 @QTV-2 生活服务频道 @青岛公安 @青岛公安 110 @青岛身边事 @青岛凡事 @青岛新闻...全文: ...
5366	#陛下该吃药了#哈哈,真是脑洞够大的也是醉了呀!赶紧呼叫皇帝,给你看病吧!
5430	#石家庄分享# 如果人民币会说话是什么样的?哈哈哈这配音也是醉了~[笑cry][笑 cry]http://t.cn/RyV11Ey
5582	和我一样同为巨蟹座的完美女人,阿娟的 30 岁,伴手礼也是醉了,搞得跟她姑娘百露一样!我今天的半丸子头和阿娟的贵妇造型美不美[哈哈][哈哈]
5598	寒战 2 很有意思,不过影院中竟然有人鼾声如雷,引发全场关注,也是醉了[哼][汗]
5607	一夜未眠,居然不困,也是醉了…有点饿,想吃一碗热腾腾的拉面! http://t.cn/8F4NUIk
5627	【悟评论】两市截止 11 点 10 分成交 2500 亿,缩量明显。沪市成交仅有 800 亿,如此下去全天成交金额沪市仅有 1600 亿,实在太少,蓝筹没资金推动是最大的牛皮理由。当然市场希望如此的牛皮,痛炒个股实在快活,这个想法也是醉了。
5683	有网友曝光称:这样的牛奶还能喝吗?小编仔细看了一下,也是醉了。牛奶外包装的质量如此堪忧,真不知道牛奶还能不能喝。[吃惊][http://t.cn/Rt7jwvj
5712	一大早抽了四大管血,也是醉了!
5834	长沙这些街巷名的奇葩由来你造吗?看到最后我也是醉了!!!via:长沙潮生活
5853	收到粉丝发来的图片也是醉了,逗 B 的我~[抠鼻]
5860	谢谢你们私信的表情包,这样的表情我也是醉了[偷笑][害羞][害羞]再次感谢大家的厚爱[害羞][害羞],就不一一回复了。有你们的地方就是天堂[愉快][愉快]… 周末愉快! http://t.cn/R2WxdDX
5865	#贵阳身边事#【这样的老人[弱]】网友爆料:今天贵阳火车北站售票厅,售票窗口前排着长长的队伍前突然窜进一老人,手持贵州省老年人优待证就往窗口递并挤掉了旁边的人,在售票员呵斥与众人的指责中,该老人还拿着"优待证"在众人面前晃了晃以示自己特殊身份,随后又转到了左边售票窗口…也是醉了![怒]
5937	到哪都下雨 我也是醉了 回家
5945	铁锅铁锅炖炖炖炖炖.ps.摄影师也是醉了也是醉了
6029	罗马登陆,住的酒店楼下就是西班牙阶梯,很美!顺道吃点东西上楼,这间餐厅就在楼下可以看西班牙阶梯的全景,东西果然又贵又难吃,那个面咸到 cry 啊!饮料倒是非常好喝!毛四十块钱就吃这点东西也是醉了!面条里是大肉块,肥肉块,,,不如把材料给我自己烧[流泪] http://t.cn/RU1yE2j

6047	#青岛身边事# 【青岛街头现奇葩占车位方式 油瓶"圈地"】青岛田家花园内道路上,一处餐饮商铺前用 20 多个装着水的油桶将门前"地盘"圈了起来,防止外来车辆停放。一般占车位和阻挡车辆停车,会想到用凳子、假模特、隔离墩及锥形帽等道具。[doge]这阵势看着也是醉了,你身边还有那些奇葩占车位方法?
6050	世界这么大,什么人都会有!遇到这种人我也是醉了!买了 2 斤膏蟹,吃后审请退款,说蟹膏是人工添加的,因为蟹的肚脐有两个洞,说是打针留下的,还说央视有爆光这种情况!我能为母蟹申冤吗?母蟹说那是我的生殖器,我的洞,我的阴道口啊,图 5 是公蟹图片,公蟹说:母蟹没两个洞的话我两个"鸡巴"怎么交配
6080	#英雄联盟视频# 这犀利的操作,我也是醉了。。。。。#英雄联盟 5 周年#
6204	拼了老命赶机,飞机既然延误[擦汗]这天气[流汗]延误我也是醉了... http://t.cn/R2WJzW3
6212	倒时差的第 5 天(去年的白色 v 领 tee 在这种桑拿天真是 rio 实穿啊) #泡之流水账# 家里有个能折腾的男朋友,我不在的日子 [奥特曼][奥特曼][奥特曼]把我的熊仔们都速配成功了[挖鼻][挖鼻][挖鼻]我也是醉了也是醉了
6314	说了番真心话到这年龄却要带上面具才说出来,为说些真心话带上面具也是醉了,但我希望我们每个人都能好,不忘初心,谢谢关心和爱我的人,还有下次过生日就 30 了,送蛋糕就别弄数字蜡烛,生怕我不知自己要奔三老了是不是。。。
6315	@淮安机场官方微博 @东方航空 就在这转个机,一本书留在座位上也能被你们给收走。我也是醉了!太特么差劲了! http://t.cn/R2dLSse
6438	这个小区的马蹄美女天使花坛 我也是醉了 http://t.cn/R2WxYao
6536	#100 天阅读计划# p15 mins20.《启动大脑》p1-28:大脑的潜能和各种举例我的总结也是醉了也是醉了
6584	也是醉了,大家都在安静的招手,就你俩蹦蹦跳跳[挖鼻]

APPENDIX D
LIST OF ABBREVIATIONS

ASSOC	Associative
BCE	before the Common Era
COP	Copula
CQL	Corpus Query Language
DEM	Demonstrative
DIR	Directional
LOC	Locative
N	Noun
NC	numeral classifier
NOM	Nominative
NP	noun phrase
O_d	direct object
O_i	indirect object
PFV	Perfective
POS	part of speech
POSS	Possessive
PSV	Passive
SUBJ	Subject
V	Verb
VP	verb phrase
X	open category
YSZL	*ye-shi-zui-le*

REFERENCES CITED

Abraham, W., & Leisiö, L. (2006). *Passivization and typology: Form and function* (Vol. 68). Amsterdam: John Benjamins Publishing.

Aikhenvald, A. Y. (2000). *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.

Aitchison, J. (2001). *Language change: Progress or decay?* Cambridge: Cambridge University Press.

An, F. (2009). Cong liangci de yufahua guocheng kan yuyan jiegou de neibu tiaozheng. *Hanyu Xuexi*, (4), 56–60.

Barry, C. T., Doucette, H., Loflin, D. C., Rivera-Hudson, N., & Herrington, L. L. (2017). "Let me take a selfie": Associations between self-photography, narcissism, and self-esteem. *Psychology of Popular Media Culture*, *6*(1), 48–60.

Benn, J. A. (2005). Buddhism, alcohol, and tea in medieval China. In R. Sterckx (Ed.), *Of Tripod and Palate: Food, Politics, and Religion in Traditional China* (pp. 213–236). New York: Springer.

Birnbaum, I. M., Hartley, J. T., Johnson, M. K., & Taylor, T. H. (1980). Alcohol and elaborative schemas for sentences. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(3), 293.

Birnbaum, I. M., & Parker, E. S. (Eds.) (1977). *Alcohol and human memory.* New Jersey: Lawrence Erlbaum Assoc.

Bisang, W. (1999). Classifiers in East and Southeast Asian languages: Counting and beyond. In J. Gvozadanovic (Ed.), *Numeral types and changes worldwide* (pp. 113–185). Berlin: Mouton de Gruyter.

Brinton, L. J. (2008). *The comment clause in English: Syntactic origins and pragmatic development*. Cambridge: University Press Cambridge.

Brown, J. D. (1986). Evaluations of Self and Others: Self-Enhancement Biases in Social Judgments. *Social Cognition; New York*, *4*(4), 353–376.

Bybee, J. L. (2006). From usage to grammar: The mind's response to repetition. *Language*, (82), 711–733.

Bybee, J. L. (2013). Usage-based Theory and Exemplar Representations of Constructions. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 49–69). Oxford: Oxford University Press.

Bybee, J. L., Perkins, R., & Pagliuca, W. (1994). *The evolution of grammar*. Chicago: University of Chicago Press.

Chao, Y. R. (1968). *A grammar of spoken Chinese.* Oakland: University of California Press.

Chappell, H. (2009). Formal and colloquial adversity passives in standard Chinese. *Linguistics*, *24*(6), 1025–1052.

Chen, F. (2002). Cong "mei" yu "ge" kan hanyu fanzhixing liangci de yanbian. *Yuwen Yanjiu*, (1), 33–35.

Chen, M. (1956). *Yinxu buci zongshu*. Beijing: Kexue Chubanshe.

Cheng, H. G., Deng, F., Xiong, W., & Phillips, M. R. (2015). Prevalence of alcohol use disorders in mainland China: A systematic review. *Addiction*, *110*(5), 761–774.

Chou, F. (1959). *Zhongguo gudai yufa*. Taipei: Institute of History and Philology, Taiwan Academia Sinica.

Chu, C. (1998). *A Discourse Grammar of Mandarin Chinese*. New York: Peter Lang.

Clary, E. G., & Tesser, A. (1983). Reactions to Unexpected Events: The Naive Scientist and Interpretive Activity. *Personality and Social Psychology Bulletin*, *9*(4), 609–620.

Cochrane, J., Chen, H., Conigrave, K. M., & Hao, W. (2003). Alcohol Use in China. *Alcohol and Alcoholism*, *38*(6), 537–542.

Croft, W. (2001). *Radical construction grammar*. Oxford: Oxford University Press.

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

Crystal, D. (2011). *Internet linguistics: a student guide*. Milton Park, Abingdon; New York, N.Y: Routledge.

Dai, Q., & Jiang, Y. (2005). Lun Zangmianyu de fanxiangxing mingliangci. *Zhongyang Minzu Daxue Xuebao*, *32*(2), 124–129.

Danesi, M. (2016). *The semiotics of emoji: The rise of visual language in the age of the Internet*. London: Bloomsbury Publishing.

Dehé, Nicole and Yordanka Kavalova (Eds.). (2007). *Parentheticals.* (Linguistics Today, 106.) Amsterdam, Philadelphia: John Benjamins Publishing.

Ding, S., Lü, S., Li, R., Sun, D., Guan, X., Fu, J., & Chen, Z. (1953). *Xiandai Hanyu Yufa Jianghua*. Beijing: Shangwu Yinshuguan.

Dixon, R.M.W. (2005). *A Semantic Approach to English Grammar*. Oxford: Oxford University Press.

Drocourt, Z. (1993). Analyse syntaxique des expressions quantitatives en chinois archaïque. *Cahiers de Linguistique-Asie Orientale*, *22*(2), 217–237.

Espinal, M. 1991. The representation of disjunct constituents. *Language* 67: 726–62.

Fillmore, C. J. (1989). Grammatical construction theory and the familiar dichotomies. In R. Dietrich & C. F. Graumann (Eds.), *Language processing in social context* (pp. 17–38). Amsterdam: Elsevier.

Fujii, N. (1992). Changes in subject marking in Japanese. In M. Gerritsen & D. Stein (Eds.), *Internal and External Factors in Syntactic Change* (Vol. 61, pp. 257–294). Berlin: Walter de Gruyter.

Gerritsen, M. (1992). Internal and external factors in the stabilization of verb-last order in Dutch infinitive clauses. In M. Gerritsen & D. Stein (Eds.), *Internal and External Factors in Syntactic Change*, *61*, 355–394. Berlin: Walter de Gruyter.

Gerritsen, M., & Stein, D. (1992). *Internal and external factors in syntactic change* (Vol. 61). Berlin: Walter de Gruyter.

Givón, T. (1988). The pragmatics of word order: predictability, importance and attention. In E. M. M. Hammond &. J. Wirth (Ed.), *Studies in Syntactic Typology* (pp. 243–284). Amsterdam: John Benjamins.

Goldberg, A. E. (1995). *Constructions: A construction gramamr approach to argument structure*. Chicago: Chicago University Press.

Goldberg, A. E. (2003). Constructions: a new theoretical approach to language. *Trends in Cognitive Science*, *7*(5), 219–224.

Goldberg, A. E. (2006). *Constructions at work : the nature of generalization in language*. Oxford; New York: Oxford University Press.

Goldberg, A. E. (2013). Constructionist approaches. *The Oxford Handbook of Construction Grammar*, 15–31. Oxford: Oxford University Press.

Goldhaber, M. H. (1997). The attention economy and the Net. *First Monday*, *2*(4). Retrieved from http://journals.uic.edu/ojs/index.php/fm/article/view/519.

Gong, Y. (2007). Hanyu "shu-liang-ming" jiegou chansheng chuyi. *Hanyu Zhishi* (4), 83–85.

Greenberg, J. (1972). Numeral classifiers and substantival number: Problems in the genesis of a linguistic type. *Working Papers on Language Universals, No. 9.*, 1–39.

Guan, X. (1953). *Yinxu jiagu keci de yufa yanjiu*. Beijing: Zhongguo Kexueyuan.

Gundel, J. K. (1988). Universals of topic-comment structure. *Studies in Syntactic Typology*, *17*, 209–239.

Hao, W., Su, Z., Liu, B., Zhang, K., Yang, H., Chen, S.,Cui, C. (2004). Drinking and drinking patterns and health status in the general population of five areas of china. *Alcohol and Alcoholism*, *39*(1), 43–52.

Heine, B., Claudi, U., & Hünnemeyer, F. (1991). *Grammaticalization: A conceptual framework*. Chicago: University of Chicago Press.

Heine, B., Kaltenböck, G., Kuteva, T., & Long, H. (2013). An outline of discourse grammar. In Shannon T. Bischoff & Carmen Jany (Eds.), *Functional approaches to language* (pp. 155-206). Walter de Gruyter.

Himmelmann, N. P. (2004). Lexicalization and grammaticization: Opposite or orthogonal. In W. Bisang, N. P. Himmelmann, & B. Wiemer (Eds.), *What makes grammaticalization?: A look from its fringes and its components* (pp. 21–42). Berlin: Walter de Gruyter.

Hilpert, M. (2014). *Construction grammar and its application to English*. Edinburgh: Edinburgh University Press.

Hopper, P. J. (1991). On some principles of grammaticization. *Approaches to Grammaticalization*, *1*, 17–35.

Hopper, P. J. (1998). Emergent grammar. *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure*, *1*, 155–176.

Hopper, P. J., & Thompson, S. A. (1980). Transitivity in grammar and discourse. *Language*, *56*(2), 251-299.

Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization*. Cambridge: Cambridge University Press.

Huang, S. (1961). Lianghan shidai de liangci. *Zhongguo Yuwen*, 21–28.

Huang, Z. (1964). Cong jiawen, jinwen liangci de yingyong kaocha hanyu liangci de qiyuan yu fazhan. *Zhongguo Yuwen*, (6), 432–441.

Israel, M. (1996). The way constructions grow. In A. Goldberg (Ed.), *Conceptual structure, discourse and language*, 217–230.

Jing-Schmidt, Z. (2005). *Dramatized discourse: The Mandarin Chinese Ba-construction*. Amsterdam: John Benjamins.

Jing-Schmidt, Z., & Jing, T. (2011). Embodied semantics and pragmatics: Empathy, sympathy and two passive constructions in Chinese media discourse. *Journal of Pragmatics*, *43*(11), 2826–2844.

Jing-Scmidt, Z., & Kapatsinski, V. (2012). The apprehensive: Fear as endophoric evidence and its pragmatics in English, Mandarin, and Russian. *Journal of Pragmatics 44*(4), 346-373.

Jing-Schmidt, Z., & Peng, X. (2016). The emergence of disjunction: A history of constructionalization in Chinese. *Cognitive Linguistics*, *27*(1), 101–136.

Kaltenböck, G., Heine, B., & Kuteva, T. (2011). On thetical grammar. *Studies in Language*, *35*(4), 852–897.

Kwong, C. (2013). Making Poetry with alcohol: Wine consumption in Tao Qian, Li Bai and Su Shi. In I. Yue & S. Tang (Eds.), *Scribes of Gastronomy: Representations of Food and Drink in Imperial Chinese Literature*. Hong Kong: Hong Kong University Press.

Labov, W. (1963). The social motivation of a sound change. *Word*, *19*(3), 273–309.

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

Lambrecht, Knud. (1996). On the formal and functional relationship between topics and vocatives: Evidence from French. In Goldberg, A. (Ed.), *Conceptual Structure, discourse and language.* (pp. 267-288) Stanford: Center for the Study of Languge and Information.

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

Langacker, R. W. (2008). *Cognitive grammar: A basic introduction*. Oxford: Oxford University Press.

Langacker, R. W. (2012). *Essentials of cognitive grammar*. Oxford: Oxford University Press.

LaPolla, R. J. (1995). Pragmatic relations and word order in Chinese. *Word Order in Discourse*, *30*, 297.

Lee, C. S., & Ma, L. (2012). News sharing in social media: The effect of gratifications and prior experience. *Computers in Human Behavior*, *28*(2), 331–339.

Lehmann, C. (1995). Thoughts on grammaticalization. München: LINCOM Europa.

Li, C., & Thompson, S. (1976). Subject and topic: A new typology of language. In C. Li (Ed.), *Subject and Topic* (pp. 457–489). New York: Academic Press.

Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.

Li, Y. (2000). Kaobeixing liangci ji qi zai hanzangyuxi liangci fazhan zhong de diwei. *Zhongguo Yuwen*, (1), 27–34.

Liu, J. J. (1966). *The art of Chinese poetry*. Chicago: University of Chicago Press.

Liu, S. (1965). *Weijin nanbeichao liangci yanjiu*. Beijing: Zhonghua Shuju.

Ma, C. (1984). *Di yu Qiang.* Shanghai: Shanghai Renmin Chubanshe.

Marwick, A. E., & Boyd, D. (2011a). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, *13*(1), 114–133.

Marwick, A. E., & Boyd, D. (2011b). To see and be seen: Celebrity practice on twitter. *Convergence: The International Journal of Research into New Media Technologies*, *17*(2), 139–158.

Michaelis, L. (2004). Type shifting in construction grammar: An integrated approach to aspectual coercion. *Cognitive Linguistics*, *15*(1), 1–68.

Mithun, M. (1992). External triggers and internal guidance in syntactic development: Coordinating conjunction. In *Internal and External Factors in Syntactic Change* (Vol. 61, pp. 89–130).

Modern Chinese Dictionary (6th ed). (2010). Beijing: Shangwu Yinshuguan.

Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 54.

Pak, A., & Paroubek P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of the International Conference on Language Resources and Evaluation Conference, Malta, 17-23 May 2010* (pp. 1320-1326). European Language Resources Assocation.

Panther, K. U., Thornburg, L. L., & Barcelona, A. (Eds.). (2009). *Metonymy and metaphor in grammar.* Amsterdam: John Benjamins Publishing.

Peterson, J. B., Rothfleisch, J., Zelazo, P. D., & Pihl, R. O. (1990). Acute alcohol intoxication and cognitive functioning. *Journal of Studies on Alcohol*, *51*(2), 114–122.

Peyraube, A. (1998). On the history of classifiers in Archaic and Medieval Chinese. In *Yuyanxue Luncun (21)* (pp. 99–122). Beijing: Shangwu Yinshuguan.

Quan-Haase, A., & Young, A. L. (2010). Uses and gratifications of social media: A comparison of Facebook and instant messaging. *Bulletin of Science, Technology & Society*, *30*(5), 350–361.

Read , J. (2005). Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *Proceedings of the ACL Student Research Workshop* (pp. 43-48). Association for Computational Linguistics.

Roseman, I. J. (1996). Appraisal determinants of emotions: Constructing a more accurate and comprehensive theory. *Cognition and Emotion*, *10*(3), 241–278.

Ryan, C., & Butters, N. (1983). Cognitive deficits in alcoholics. In B. Kissin & H. Begleiter (Eds.), *The Biology of Alcoholism* (pp. 485–538). New York: Springer.

Sanches, M. (1971). *Numeral classifiers and plural marking: An implicational universal*. (Unpublished manuscript).

Schnoebelen, T. (2012). Do you smile with your nose? Stylistic variation in Twitter emoticons. *University of Pennsylvania Working Papers in Linguistics*, *18*(2), 14.

Shi, C. (2013). Xin "bei" zishi de shengcheng jizhi, yuyi lijie ji yuyong xiaoying. *Dangdai Xiucixue*, (1), 12-28.

Shibatani, M. (1985). Passives and related constructions: A prototype analysis. *Language*, *61*(4), 821–848.

Smith, N. (2012). *Intoxicating Manchuria: Alcohol, opium, and culture in China's northeast*. Vancouver: University of British Columbia Press.

Söderlund, H., Parker, E. S., Schwartz, B. L., & Tulving, E. (2005). Memory encoding and retrieval on the ascending and descending limbs of the blood alcohol concentration curve. *Psychopharmacology*, *182*(2), 305–317.

Steele, C. M., & Josephs, R. A. (1990). Alcohol myopia: Its prized and dangerous effects. *American Psychologist*, *45*(8), 921.

Stefanowitsch, A., & Gries, S. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics, 8*(2), 209–243.

Sweetser, E. (1990). *From etymology to pragmatics: The mind-body metaphor in semantic structure and semantic change*. Cambridge: Cambridge University Press.

Thomason, S. G., & Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Boston: Harvand University Press.

Traugott, E. C., & Trousdale, G. (2010). *Gradience, gradualness and grammaticalization*. Amsterdam: John Benjamins Publishing.

Traugott, E. C., & Trousdale, G. (2013). *Constructionalization and constructional changes* (1st ed.). Oxford: Oxford University Press.

Tsao, F. (1979). *A functional study of topic in Chinese.* Taipei: Student Book Co.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352.

Van Dijck, J. (2013). *The culture of connectivity: A critical history of social media*. Oxford: Oxford University Press.

Wang, C. (2009). "bei" zi de linglei yongfa: Cong "bei zisha" tanqi. *Yuwen Jianshe*, (4), 65-66.

Wang, L. (1943). *Chinese modern grammar*. Beijing: Commercial.

Wang, L. (1958). *Hanyu shigao.* Beijing: Zhonghua Shuju.

Wang, L. (1985). *Zhongguo Xiandai Yufa.* Beijing: Shangwu Yinshuguan.

Wang, M. (2008). *Qiang zai Han Zang zhijian.* Beijing: Zhonghua Shuju.

Wang, Y. (1980). *Early Chinese Coinage*. New York: Sanford J. Durst Numismatic Publications.

Wang, Y. (2011). “Xin bei zi goushi” de cihui yazhi jiexi: Dui "bei ziyuan" yilei xin biaoda de renzhi goushi yufa yanjiu. *Waiguoyu,* (3), 13-20.

Weinreich, U., Labov, W., & Herzog, M. I. (1968). *Empirical foundations for a theory of language change*. Austin: University of Texas Press.

Xiao, R., McEnery, T., & Qian, Y. (2006). Passive constructions in English and Chinese: A corpus-based contrastive study. *Languages in Contrast*, *6*(1), 109–149.

Xiao, Y. (2016). The rise of catchword “被 *bei* -XX”: Grammaticalization and reanalysis. *Chinese as a Second Language (漢語教學研究—美國中文教師學會學報)*, *51*(1), 79–97.

Xing, J. (2013). Semantic reanalysis in grammaticalization in Chinese. In Z. Jing-Schmidt (Ed.), *Increased empiricism: Recent advances in Chinese linguistics* (pp. 223–246). Amsterdam/Philadelphia: John Benjamins.

Xu, L., Lin, H., Pan, Y., Ren, H., & Chen, J. (2008). Qinggan cihui benti de gouzao. *Qingbao Xuebao,* 2.

Yang-Drocourt, Z. (2004). *Evolution syntaxique du classificateur en chinois du XIIIe siècle av. J. -C. au XVIIe siècle.* Paris: EHESS, CRLAO.

Yuan, H. & Liang, J. (2016). "Bei + X" goushi yi de gainian zhenghe fenxi. Waiyu Yanjiu, (1), 33–39.

Zappavigna, M. (2012). *Discourse of Twitter and social media: How we use language to create affiliation on the web*. London/New York: Continuum International Publishing Group.

Zhang, H., Guan, L., Qi, M., & Yang, J. (2013). Self-esteem modulates the time course of self-positivity bias in explicit self-evaluation. *PLOS ONE*, *8*(12), e81169.

Zhang, Y. (2001). *Yinzhou Jinwen Jicheng Yinde*. Beijing: Zhonghua Shuju.

Zhang, Y. (2002). Ye lun hanyu "shu-liang-ming" xingshi de chansheng. *Guhanyu Yanjiu*, (2), 26–29.

Zhao, D., & Rosson, M. B. (2009). How and why people twitter: The role that micro-blogging plays in informal communication at work. In *Proceedings of the ACM 2009 International Conference on Supporting Group Work* (pp. 243–252). New York: ACM Publications.

Zhao, J., Dong, L., Wu, J., & Xu, K. (2012). Moodlens: an emoticon-based sentiment analysis system for Chinese Tweets. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1528-1531). New York: ACM Publications.