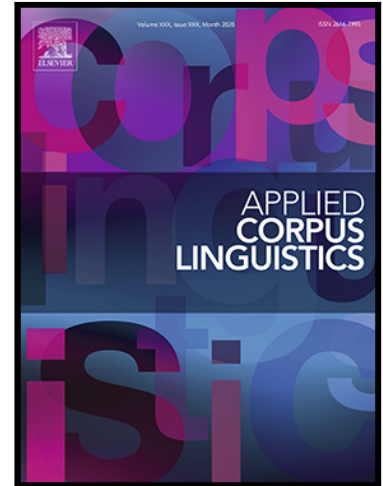


Journal Pre-proof

Examining the impact of growing household over-indebtedness in China: A corpus linguistics analysis of a popular online debt support forum



Dr. Jiaqi Feng Guo , John Marshall

PII: S2666-7991(25)00024-3
DOI: <https://doi.org/10.1016/j.acorp.2025.100141>
Reference: ACORP 100141

To appear in: *Applied Corpus Linguistics*

Received date: 2 December 2024
Revised date: 6 May 2025
Accepted date: 2 August 2025

Please cite this article as: Dr. Jiaqi Feng Guo , John Marshall , Examining the impact of growing household over-indebtedness in China: A corpus linguistics analysis of a popular online debt support forum, *Applied Corpus Linguistics* (2025), doi: <https://doi.org/10.1016/j.acorp.2025.100141>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2025 Published by Elsevier Ltd.

Examining the impact of growing household over-indebtedness in China: A corpus linguistics analysis of a popular online debt support forum

Authors:

Dr. Jiaqi Feng Guo¹

Chinese Department, School of Languages and Translation Studies, University of Turku, Turku, 20014, Finland.

Email address: jiaqi.guo@utu.fi

John Marshall²

Independent Researcher

Email address: john.marshall1981@icloud.com

Corresponding author:

Dr Jiaqi Feng Guo

Declaration of interest statement:

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Abstract: In recent years, Chinese household finances have been reshaped by a massive surge in borrowing. This has seen various financial products including mortgages, consumer loans and credit cards shift from being relatively rare to become central to economic life. Inevitably this is creating ‘over-indebtedness’, where many borrowers cannot repay their obligations. However, in contrast to its scale, there has been little sociological examination of the phenomenon’s impact. This is problematic

because, unlike the limited investigation in China, numerous international studies show that over-indebtedness can create significant social harm. To address this research deficit, we undertook corpus analysis of the largest debt support forum on the Chinese internet. While we acknowledge the limitations of online data in this context, the scope of our dataset and the analytical techniques employed, mitigate them to a meaningful extent. Our results suggest borrowers are most concerned about debt collection practices. This includes ‘contact bombing’, where debt collectors attempt to weaponise their social networks by repeatedly calling relatives, friends, and colleagues to pressurise them to repay or face ‘social death’. Discursive patterns indicate the consequences permeate forum members’ lives, distorting their relationships, time horizons and decision-making. Forum discussions also shed light on attitudes to China’s consumer rights framework, with borrowers often reporting a perceived lack of regulatory protection. Alongside its empirical value, the research has important methodological implications. Specifically, few studies have undertaken sociological inquiry in China via corpus analysis of large natural language Mandarin datasets. We provide a showcase for this approach.

Keywords:

Over-indebtedness; Debt collection; China; Corpus linguistics; Online forums; Contact bombing

1. Introduction

In the last 15 years, household borrowing in China has surged. The scale of the change is evident from data which show that household debt has leapt from 11 percent of GDP in 2006 to 62 percent in 2023 (IMF, 2024). This trend has been driven by explosive growth in various lending products. For example, the number of credit cards in use has increased from 50 million in 2006 to nearly 800 million today (Slotta, 2024). However, while the quantitative parameters of China's household financialisation process are well-understood thanks to abundant macro-economic analysis, the question of how accumulating debt is manifesting at the micro-level in everyday life has received significantly less attention.

This is problematic because, in contrast to China, numerous international studies suggest quickly growing household borrowing often precipitates substantial 'over-indebtedness' – where borrowers cannot repay what they owe without severely compromising their life quality – and this can have far-reaching consequences. Specifically, research demonstrates that excess household debt harms mental and physical health (Sweet et al., 2013; Turunen & Hiilamo, 2014), exacerbates poverty and precariousness (Achtziger, 2022), and can weaken communities by rupturing families and driving social isolation (Montgomerie & Tepe-Belfrage, 2017). However, such studies also highlight the danger of assuming that the findings from one setting apply elsewhere. This is because how indebtedness intertwines with different countries' social, institutional, and moral frameworks is contingent on specific contextual circumstances. In the words of Harker (2017, p.600), there are unique 'debt ecologies'.

Given this, there is a pressing need for research that expands our understanding of the everyday implications of over-indebtedness in the distinct circumstances of contemporary China. To generate insights, we conducted corpus linguistics analysis of an online forum for those dealing with excess debt - 债务者联盟 *zhàiwùzhě liánméng* 'Debtor Alliance'. To the best of our knowledge, this is the largest such forum on the Chinese internet. The forum's orientation guidance states it is intended to be a resource for those 'burdened by debt' to 'exchange experiences' and offer 'mutual assistance'. The entire forum content has been collated for analysis, resulting in a dataset of 6.45 million words.

The data were investigated using three corpus linguistics techniques: keyword, collocation and concordance analyses.

Although we acknowledge the constraints of both forum data in general, and specifically online data in the Chinese context (e.g., censorship), which we fully discuss later, the size of our dataset, and the fact that the study's analytical techniques excel at identifying subtle patterns in language that are not readily apparent to manual observation help to mitigate them. It is also crucial to note that due to the lack of public guidance regarding over-indebtedness in China, those facing this issue have few other resources to turn to apart from what is available online. Therefore, investigating such an extensively used forum, and the intersubjective understanding that is being formed through it, is worthwhile despite the potential limitations.

The study is guided by an exploratory research question: What issues do over-indebted individuals in China choose to discuss online, and what do their linguistic choices suggest about how they conceive the impact of these issues? Through this approach, we not only deepen knowledge of an important unfolding phenomenon in China, but methodologically we also demonstrate how corpus linguistics techniques can support sociological research in this context via the analysis of a large natural language Mandarin dataset. This is a method that remains relatively under-utilised in the context of Chinese sociology.

2. The impact of over-indebtedness

As noted above, in the last decade and a half, Chinese household borrowing has grown extremely rapidly. While this has undoubtedly benefitted many by providing a means to smooth income and expenditure, it is also inevitably generating over-indebtedness. Although no precise figures exist for the number of people who are experiencing this situation, research suggests that once a country amasses high household debt (generally considered to be above 60 percent of GDP), approximately one fifth of households will suffer debt distress at any one time (Leandro & Botelho, 2022; D'Alessio & Iezzi, 2013; Lombardi et al., 2017).

However, beyond a body of macro-economic analysis which focuses on linkages to other financial metrics like GDP, there has been little sociological examination of this household

financialisation process. Indeed, a review of scholarly databases reveals only eight studies that assess Chinese household debt from this perspective, seven of which adopt the same methodology – correlational analysis of indicators of debt, income, and well-being arising from general social surveys (e.g., China Family Panel Study). These suggest that high debt increases the likelihood of depression and anxiety (Liu et al., 2021; Hu et al., 2023), worsens physical health (Song et al., 2023), erodes life satisfaction (Liu et al., 2020; Xiao et al., 2021), and that these negative effects occur across groups, including those over 60 (Song et al., 2020) and children in highly-indebted households (Liu et al., 2024). However, while these studies are useful, given their methodology, they only produce a surface-level representation of the effects of rising debt. They are also generally based on relatively old data, gathered when household borrowing was just starting to accelerate.

Besides such research, only one more in-depth investigation has sought to unpack how over-indebtedness is manifesting in everyday life by probing the experiences of those dealing with it. Specifically, Rao and McDonald (2023) examined how over-indebted individuals coped during the early stages of the COVID-19 pandemic. To generate data, they analysed 24 videos made by defaulting borrowers, shared on the short-video platform Xigua. They found that the video creators were most concerned about debt collection agencies and their aggressive pursuit of repayment. They stated a key feature of debtor-debt collector interactions was the use of collective imaginaries concerning the state and family, and the moral contest they reflect about default. They observed debtors often referenced government statements about the need for collective sacrifice in the face of challenges like the pandemic to resist collection pressure. In contrast, the agencies reportedly invoked traditional moral narratives, particularly of behaving in a way that avoids shaming one's family. They concluded that debtors are attempting to 'redefine the moral contours' of indebtedness by embracing a 'defaulter identity' (p.263). However, while the study reveals intriguing aspects of how over-indebtedness is manifesting in China and interacting with specific cultural discourses, it has certain limitations. First and foremost, is its small sample. Moreover, the video creators share similarly narrow characteristics – male migrant workers in unstable employment.

The outcome of this limited research is we are left with only a rudimentary understanding of how over-indebtedness is affecting both individual borrowers and Chinese society more generally. A

large body of international research underscores why this is a problem and the need for more data-rich, methodologically-diverse analysis. In particular, the extensive array of studies conducted in other settings reveals a fundamental point, namely, although over-indebtedness generates negative outcomes wherever it occurs, the depth and severity of these outcomes are strongly modulated by the distinct cultural, economic and social conditions that exist between countries.

These include: the contrasting legal and regulatory frameworks designed to protect borrowers from predatory lending and collection practices (Sparkes et al., 2018), divergent discourses about the morality of debt (Graeber, 2011), the maturity and competitiveness of the financial sector (Finney et al., 2007), varying financial education and literacy (Disney & Gathergood, 2013), historically contingent family structures and living arrangements (Harker, 2017), and the differing levels of communal support arising from distinct welfare systems and civil societies (Clark et al., 2022). These not only affect how debt arises and operates in a society, but also how individuals understand themselves and their ability to cope when facing over-indebtedness.

Take for example the legal frameworks governing debt collection. As Stănescu et al. (2022) highlight, borrowers in some countries are shielded from abusive measures by robustly enforced regulation. In contrast, weaker safeguards in other contexts allow debt collectors to employ a range of coercive tactics. The significance of such disparities is underscored by Deville (2015), who demonstrates that aggressive debt collection can disrupt fundamental aspects of life. For instance, persistent harassment can lead individuals to internalise their debt as a defining part of their identity and this often triggers a cascade of negative outcomes, including the corrosion of relationships as personal concerns are overwhelmed by economic ones. Moreover, research by Lea (2021) illuminates the psychological mechanisms of how intimidatory debt collection can induce temporal disorientation. Specifically, under constant pressure, individuals may become fixated on past decisions and future obligations, leaving them unable to focus on the present and impairing their capacity to make healthy decisions.

Therefore, given the impact of such institutionally and culturally contingent factors, there is a need for context-specific investigation of over-indebtedness. The relative lack of such research in China can be explained by two factors. The first is although Chinese household financialisation began

accelerating 15 years ago, it is only with COVID-19 and the subsequent economic slowdown that many families' finances have come under sustained pressure. Specifically, while both wages and GDP were increasing by 10 percent per annum at the start of the 2010s, which provided many households high financial resilience, growth is now significantly lower (World Bank, 2024). Therefore, substantial over-indebtedness in China remains an emerging phenomenon. The second reason relates to access. In particular, conducting research in China is becoming increasingly challenging due to the country's tightening political environment (Mazzocco & Kennedy, 2022). Consequently, it is more difficult for scholars to undertake the sort of detailed ethnographic fieldwork which has helped to unpack how over-indebtedness is intertwining with other countries' societies. Furthermore, studies in many of those settings have located research participants via debt-focussed charities. However, as far as we are aware, no such organisations exist in China.

To deepen our understanding of how over-indebtedness is unfolding in the Chinese context, this study takes its methodological lead from one of the key strands in the international research that does not rely on fieldwork. In particular, several studies (e.g., Stanley et al., 2015; Stout, 2016), have demonstrated that digital ethnography via the investigation of online forums is also a fruitful means of examining over-indebtedness. This research shows that debtors often seek emotional support and practical advice via forums, which leads them to provide a 'rich tapestry of qualitative detail' about their situations (Montgomerie & Tepe-Belfrage, 2017, p.11).

One of the main reasons that debt forums are so valued seemingly relates to the anonymity they offer in the context of an issue that frequently evokes shame. However, counter-intuitively this anonymity also works to facilitate intimacy and camaraderie with bonds formed via the sharing of experiences. This is what Deville (2016, p.39) characterises as the emergence of 'debtor publics'. Furthermore, the meaning-making that takes place within these forums extends beyond immediate participants. As Deville (2015) notes, such spaces also serve as information archives for future users, enabling a form of asynchronous engagement that amplifies the influence of the content over time. In this way, debt forums can enable 'online ambient affiliation' (Zappavigna, 2011, p.15), where individuals, often strangers and frequently without direct interaction, can come to form shared values and evaluative stances by engaging with common topics in a digital environment.

From a methodological perspective, nearly all forum-based studies on over-indebtedness have employed qualitative thematic analysis. However, while this approach undoubtedly offers valuable insights, it is constrained by the amount of data that can be examined, and has been criticised for cherry-picking and researcher bias (Kogen, 2024). The only exception comes from Lawson et al. (2023) who used corpus methods to investigate two UK debt forums. Like the earlier qualitative research, they found that forums facilitate collective understanding and support, but they showed that corpus linguistics techniques can identify subtle patterns concerning debt discourse inaccessible to traditional methods due to their ability to reveal latent patterns in large datasets.

However, while these forum-based studies are extremely useful in investigating how over-indebtedness materialises, it is vital to recognise that their settings, e.g., the UK and the USA, differ significantly to China, particularly regarding freedom of speech and the circumstances of the internet. This is relevant as the divergences in these areas in China have important implications for the ability to undertake digital ethnography and corpus linguistics research. In the next section, these issues will be elaborated alongside the study's data source, analytical techniques and ethical considerations.

3. Methodology

3.1 Data source

To expand understanding of over-indebtedness in China, we conducted corpus linguistics analysis of the 'Debtor Alliance' forum. As noted in the introduction, this was created for those struggling with debt to share advice and support. In May 2024, when we collected our data, it had 58,119 members. These are individuals who have applied to the forum moderator to be able to start or contribute to discussion threads. However, the forum's full content is also publicly accessible to view by non-members.

Using a custom script created in Python, we collated the complete forum content since its inception in 2019. This resulted in a dataset of 6.45 million words, spread across 26,990 separate threads. At the time of data collection, approximately 40 new threads were being created each day, with the most popular receiving several hundred replies. Before describing how we analysed this data, we briefly review two issues. First, we provide more details about where the forum is hosted as this

sheds light on its likely users. Then, we discuss the limitations to Chinese online data and why our analysis remains worthwhile despite them.

The Debtor Alliance forum is hosted on a platform called Douban. Douban originated as a venue to share reviews of books and films, and while this remains a key function, it has evolved into a more general discussion site, particularly via its ‘Small Groups’. The forum sits within the ‘Life’ section of the Small Groups. In terms of understanding who uses the forum, as for much research of online communities, because users are not required to display any demographic information, it is not possible to determine a precise user profile and then reliably generalise findings. However, several market research reports help contextualise Douban users. Specifically, these suggest a typical user was born after 1980, lives in a major city (i.e., a more economically developed area), and holds a university degree (China-Digital, 2023). Reflecting this relatively young, urban and well-educated user-base, Douban has gained a reputation as a site for discussion of hot-button social issues (Wu, 2021). In terms of Douban’s overall reach, it is estimated to have 60 million registered users (SDG, 2024).

Alongside Douban’s user demographics, another crucial issue that requires careful consideration relates to the circumstances of the internet in China and freedom of speech. In particular, it is vital to acknowledge that China’s internet is subject to heavy monitoring, control and censorship (Liu, 2024). Therefore, while individuals can post anonymously in terms of their displayed usernames, which limits others from knowing who they are, there is general awareness that such anonymity is superficial and that both Douban and governmental authorities are able to track anyone who posts material. This is because all forums in China require registration with a mobile phone linked to one’s unique personal ID. However, while this inevitably promotes self-censorship regarding certain topics, there are three offsetting factors which mean analysing the forum remains a valuable endeavour.

Firstly, as noted earlier, there is minimal public guidance in China about dealing with over-indebtedness and no debt-focussed charities. This means over-indebted individuals have few support resources beyond what they can find online. Consequently, the forum, which we believe is the largest concerning this topic on the Chinese internet, is vital to mediating wider understanding of over-indebtedness. Given this, there is undoubtedly value in examining what users do share and the

intersubjective conceptualisation they are forming and propagating, even if they may not post as freely as individuals elsewhere. Secondly, the harmful outcomes of over-indebtedness mainly materialise in individuals' private lives. This is important as there is greater leeway to discuss such issues in China compared to more overtly political topics. Finally, the corpus linguistics techniques employed in this study excel at identifying hidden layers in language which are invisible to manual observation or even to the individuals who generate it. Instead, these only emerge as patterns from large-scale data-driven analysis. This makes them less conducive to being filtered out by moderation or censorship.

3.2 Data analysis

We analysed the forum dataset in Sketch Engine, a commercial corpus tool. It enables various statistical tests to be conducted to identify patterns within a text relating to word frequency, word sequences, and words that habitually appear in close proximity. We used three corpus linguistics techniques to unpack our data – keyword, collocation, and concordance analyses. These combine quantitative and qualitative elements.

Firstly, keywords were determined by comparing the focus corpus, i.e., the Debtor Alliance dataset, with a reference corpus. A reference corpus serves as a representative baseline of 'standard' language usage (Baker & McEnery, 2015, p.2). We employed the 13.5 billion-word zhTenTen17 corpus of simplified Chinese available via Sketch Engine. This is an acceptable reference as it also comprises data drawn solely from the Chinese internet (compiled in 2017), helping to ensure that we compared similar types of language. To validate its suitability, we built a small sub-corpus from the main zhTenTen17 corpus comprising only micro-blog data from Sina Weibo (which is often equated to Twitter/X) and is likely to contain language of a similar register to a discussion forum. The sub-corpus contained 2.5 million words. When we used it as a reference, it resulted in a nearly identical selection of keywords as the overall zhTenTen17 corpus.

The purpose of the focus/reference corpus comparison is to reveal words that occur at a statistically higher rate in the focus corpus compared to the baseline. The resulting keywords point to 'distinctive linguistic features' of the data and therefore to the issues foregrounded by its creators

(Coltman-Patel et al., 2022, p.3). Sketch Engine describe keywords as a text's 'DNA'. To determine keywords, Sketch Engine calculates a 'keyness score'. This uses a simple maths approach and indicates how many more times a word appears in the focus corpus versus the reference corpus, having been normalised to account for the differing corpora sizes.

The second technique, collocation analysis, involves identifying patterns of word co-occurrence and the strength of the association. The significance is through their habitual co-occurrence, words, including keywords, can develop evaluative prosodies and reveal attitudes and ideologies (Lustig & Brookes, 2022). As Baker and McEnery (2015, p.2) note, 'collocates are extremely interesting, as if two words are repetitiously associated with each other, then their relationship can become reified and unquestioned'. We focussed on the words that typically co-occur with the keywords in the dataset, limiting our analysis to those among the 10 strongest collocates determined by logdice score.

A further important issue to note regarding collocation concerns how this concept manifests in Chinese relative to English. While the fundamental principle remains the same in both languages, i.e., certain words are habitually used in combination and thus form linguistic conventions among the speech community, Chinese collocation exhibits some specific characteristics. For instance, Chinese possesses several unique structural features, including the frequent use of multi-character lexical items, four-character idiomatic set phrases, and tightly bound verb-object constructions. Sketch Engine accounts for these linguistic specifics through tailored NLP processes, particularly Chinese word segmentation and grammatical pattern identification. These techniques ensure the robustness of collocational extraction from Chinese text data (Huang et al., 2005).

Finally, we employed concordance analysis, which allows detailed follow up into keywords and collocations via examining their use in context. Specifically, concordance lines show all instances of the keywords and their collocates with their surrounding text. We provide concordance line excerpts throughout the results section to exemplify our findings.

A key challenge of corpus linguistics is selecting a manageable amount of data to analyse. This is because hundreds of keywords can emerge from a large dataset. This requires decisions about cut-off points. To start, we decided to focus on keywords with a keyness score of more than 25. This

was designed to result in an amount of data that was sufficient to capture the forum's key themes but not overwhelm us. This resulted in 181 keywords. We built our understanding of these through inspection of their collocation patterns using Sketch Engine's 'Word Sketch' function, and how they are typically used via concordance lines. As many keywords appeared thousands or tens of thousands of times on the forum, we examined a random sample of 100 concordance lines for each (Brookes & Baker, 2022). Both researchers undertook this process separately with the aim of grouping the keywords into thematic categories to help organise our analysis. While our literature review work gave us awareness of potential categories, we did not seek to impose these on the data. Instead, we followed an inductive approach, allowing themes to emerge via our analysis. We initially achieved excellent agreement, but where we disagreed, we jointly reexamined the collocation patterns and concordance examples and agreed a final categorisation scheme. This resulted in 11 categories (Table 1).

Table 1. Keyword thematic categories

Thematic category	Description	Example	Number of keywords
Financial institutions and products	Names of banks, online lenders and financial products	微粒贷 (<i>wēilìdài</i> , 'WeChat micro-loans')	26
Financial terminology	Technical terminology relating to credit, debt and transactions	罚息 (<i>fāxī</i> , 'interest penalty')	46
Debt collection	Debt collection practices, and the responses of forum members to them	上门 (<i>shàngmén</i> , 'doorstep visit')	33
Judicial system enforcement	Formal legal proceedings/prosecution	传票 (<i>chuánpiào</i> , 'court summons')	7
Relationships with family, friends and colleagues	How over-indebtedness affects existing social ties	社死 (<i>shèsi</i> , 'social death')	5
Relationships with forum members	Development of relationships among forum members	共勉 (<i>gòngmiǎn</i> , 'let's do this together!')	16
Relationships with regulatory agencies	Development of relationships with regulators	信访 (<i>xìnfǎng</i> , 'petitioning system')	8
Time horizons	Perceptions of time	一开始 (<i>yìkāishǐ</i> , 'at the beginning')	5
Rationalities	Forum members' goals and decision-making	上岸 (<i>shàngàn</i> , 'to come ashore')	16

Discourse markers	Words that help structure discourse	反正 (<i>fǎnzhèng</i> , 'anyway')	4
Miscellaneous	Words that do not fit any other category	疫情 (<i>yìqíng</i> , 'pandemic situation')	15

As Table 1 shows, 'Financial Terminology' is the largest keyword category and 'Financial Institutions and Products' the third largest. However, detailed analysis revealed they are not the most significant for understanding themes within the dataset. This is because their keywords are generally not the focus of discussions. Instead, they usually feature in the background information provided by forum members that contextualises their primary concerns. For instance, in a typical thread, a forum member will start by explaining which lender/s they borrowed from and via what product/s, before proceeding to their main topic, e.g., their concerns about the impact of their debt on their family relations. In the results section, we therefore concentrate on the categories: Debt collection; Judicial system enforcement; Relationships with family, friends and colleagues; Relationships with forum members; Relationships with regulatory agencies; Time horizons; and Rationalities, as these emerged as the most significant for understanding forum discussions. All keywords in these categories are presented in Table 2. For all 181 keywords, see Appendix A.

3.3 Ethics

Although our method relies on big-data analysis, it is vital to recognise the underlying contributions come from thousands of individuals. Given this, considering issues relating to research participants is a prerequisite. In this study, we followed the guidelines of the British Psychological Society for investigating online communities. In particular, these state that the observation of behaviour should only take place where there is no reasonable expectation of privacy (Kaye et al., 2021). In this case, as described, the Debtor Alliance forum is freely available, and contributors have knowingly shared their posts on a publicly accessible site. However, we also believe given the sensitive nature of some of the topics involved, safeguards are needed. Therefore, we do not include usernames or personally

identifying information with the concordance excerpts provided in the results section. Excerpts have also been translated from Chinese to English, providing another layer of anonymity.

4. Results

Table 2. Keywords within selected thematic categories

Thematic category	Keywords (Keyness score)
Debt collection	催收 <i>cūshōu</i> , 'pressurised debt collection' (1188); 通讯录 <i>tōngxùnlù</i> , 'contact list' (260); 催狗 <i>cūgǒu</i> , 'chasing dog' (159); 停催 <i>tíngcuī</i> , 'stop the chasing' (157); 催 <i>cuī</i> , 'to chase/to urge' (114); 打给 <i>dǎgěi</i> , 'to call' (83); 骚扰 <i>sāorǎo</i> , 'harass' (67); 回电 <i>huídian</i> , 'call back' (60); 工号 <i>gōnghào</i> , 'company number' (59); 爆 <i>bào</i> , 'explode' (57); 拉黑 <i>lāhēi</i> , 'block' (54); 怼 <i>duì</i> , 'verbally fight/rebut' (45); 接 <i>jiē</i> , 'answer (the phone)' (38); 烦死 <i>fánsǐ</i> , 'extremely annoying' (38); 录音 <i>lùyīn</i> , 'record' (38); 恶心 <i>ěxīn</i> , 'disgusting' (37); 座机 <i>zuòjī</i> , 'landline' (34); 慌 <i>huāng</i> , 'panic' (32); 松口 <i>sōngkǒu</i> , 'relent' (31); 上门 <i>shàngmén</i> , 'doorstep visit' (31); 回 <i>T huíT</i> , 'reply to a text' (31); 催债 <i>cūzhài</i> , 'pressurised debt collection' (29); 短信 <i>duǎnxìn</i> , 'text message' (28); 失联 <i>shīlián</i> , 'lose contact' (27); 威胁恐吓 <i>wēixiékǒnghè</i> , 'threats and intimidation' (27); 消停 <i>xiāotíng</i> , 'quieten down' (27); 恐吓 <i>kǒnghè</i> , 'threaten' (26); 网商
Judicial system enforcement	起诉 <i>qǐsù</i> , 'to be sued' (122); 律师函 <i>lǜshīhán</i> , 'legal letter' (116); 诉前 <i>sùqián</i> , 'before being sued' (86); 吓唬 <i>xiàhu</i> , 'scare' (58); 传票 <i>chuánpiào</i> , 'court summons' (57); 法务 <i>fǎwù</i> , 'law' (51); 撤诉 <i>chèsù</i> , 'withdraw the lawsuit' (41); 坐牢 <i>zuòláo</i> , 'be sent to prison' (29); 套路 <i>tàolù</i> , 'trick' (28)
Relationships with family, friends and colleagues	坦白 <i>tǎnbái</i> , 'confess' (139); 家里人 <i>jiālǐrén</i> , 'family members' (63); 爸妈 <i>bànmā</i> , 'parents' (46); 社死 <i>shèsi</i> , 'social death' (40); 瞒 <i>mán</i> , 'to hide' (26)
Relationships with forum members	有没有 <i>yǒuméiyǒu</i> , 'do you have or not?' (273); 协商 <i>xiéshāng</i> , 'negotiate' (186); 嗯嗯 <i>ēnēn</i> , 'exclamation of acknowledgement' (154); 楼主 <i>lóuzhǔ</i> , 'thread creator' (85); 加油 <i>jiāyóu</i> , 'come on!' (83); 进组 <i>jìnzǔ</i> , 'join the group' (74); 组里 <i>zǔlǐ</i> , 'in the group' (72); 投诉 <i>tóusù</i> , 'complain' (56); 恭喜 <i>gōngxǐ</i> , 'congratulations!' (53); 咋 <i>zá</i> , 'what?' (44); 老哥 <i>lǎogē</i> , 'brother' (31); 共勉 <i>gòngmiǎn</i> , 'let's do this

together!' (29); 哦哦 *ōō*, 'exclamation of understanding' (29); 组规 *zǔguī*, 'group rules' (28); 多久 *duōjiǔ*, 'how long?' (26); 咋样 *záyàng*, 'what do you think?' (26)

Relationships with regulatory agencies

信访 *xìnfǎng*, 'petitioning system' (76); 没用 *méiyòng*, 'useless' (43); 保监 *bǎojiān*, 'Insurance Regulatory Commission' (40); 银监局 *yínjiānjú*, 'Banking Regulatory Bureau' (39); 银监会 *yínjiānhuì*, 'Banking Regulatory Commission' (35); 信访 *xìnfǎng*, 'petitioning system' (30); 信访办 *xìnfǎngbàn*, 'Petition Bureau' (27)

Time horizons

逾期 *yúqī*, 'overdue' (383); 贷养贷 *dàiyǎngdài*, 'a loan to care for a loan' (346); 一开始 *yīkāishǐ*, 'from the start' (44); 拆东墙补西墙 *chāidōngqiángbǔxīqiáng*, 'pull down the east wall to repair the west wall' (27); 拖 *tuō*, 'delay' (26)

Rationalities

还完 *huánwán*, 'completely repay' (428); 结清 *jiéqīng*, 'settle/repay' (274); 上岸 *shàngàn*, 'come ashore' (217); 还款 *huánkuǎn*, 'repay' (115); 还清 *huánqīng*, 'completely repay' (111); 还上 *huánshàng*, 'repay' (92); 攒钱 *zǎnqián*, 'earn money' (85); 攒够 *zǎngòu*, 'earn enough' (63); 还钱 *huánqián*, 'repay' (57); 攒 *zǎn*, 'earn' (46); 还债 *huánzhài*, 'repay' (43); 老赖 *lǎolài*, 'deadbeat defaulter' (41); 躺平 *tǎngpíng*, 'to lie flat' (38); 摆烂 *bǎilàn*, 'let things fall apart' (36); 副业 *fùyè*, 'second job' (34); 节流 *jiéliú*, 'reduce spending' (25)

4.1 Debt collection

Our analysis starts with the thematic category of debt collection. It contains the keyword, 催收 *cūishōu*, with the highest keyness score in the dataset, nearly double that of any other. This is a compound word comprising two verbs, 催 'to urge/chase' and 收 'to collect', and can be translated as 'pressurised debt collection'. Its strongest collocate is 暴力 *bàolì* 'violent' and this habitual combination, alongside their prominence in forum discussions, reveals that one of the most significant implications of over-indebtedness for many forum members is the relationship and interactions it has fostered with China's debt collection industry. This is typified in the following concordance excerpts:

1: I've recently become overdue and I've been experiencing violent pressurised debt collection ever since. I'm on the edge of a complete meltdown.

2: I feel like I am being destroyed by this violent pressurised debt collection. I owe money not my life.

3: I have been suffering violent pressurised debt collection for several months now. There's been several times where I thought I can't go on.

Another keyword from this category, 催狗 *cuīgǒu*, underscores how many forum members' conceive the nature of this new relationship. Also a compound word, it comprises the verb 催 'to urge/chase' with the noun 狗 'dog', and is the terminology employed on the forum to describe debt collectors. Conceptualising them as 'chasing dogs' has a strongly negative resonance as derogatory terms featuring the modifier 'dog' have long been used in China to imply moral degeneracy. For example, in the Mao-era those accused of betraying China were labelled 走狗 *zǒugǒu* 'running dogs'. Given this, it is unsurprising 'chasing dogs' has several collocates connected to unpleasant actions/emotions such as 骚扰 *sāorǎo* 'to harass' (which is also a keyword) and 疯 *fēng* 'insane'. The strength of feeling that exists among forum members is exemplified in the following concordances:

4: These chasing dogs all have black hearts. They are just thugs who drive people into their graves.

5: The chasing dogs are all shameless scum.

The reasons for these powerful feelings are also revealed by the keywords in this category. Specifically, 通讯录 *tōngxùnlù* 'contact list/contacts' and its strongest collocate 爆 *bào* 'explode' (also a keyword) describe a debt collection practice that we translate as 'contact bombing'. This is one of the most commonly discussed topics on the forum and it refers to the situation whereby when borrowers take out loans in China, they are often required to provide lenders access to their phone and WeChat (the country's most popular social media app) contacts. Then, if they become overdue, this information is provided to a collection agency who not only repeatedly calls and messages the debtor, but also their contacts, typically multiple times a day, to urge them to pressurise their acquaintance to repay. Forum members report that alongside phoning them (打给 *dǎgěi*) and sending text messages (短信 *duǎnxìn*), debt collectors also conduct doorstep visits (上门 *shàngmén*), with all these forms of contact characterised by threats and intimidation (威胁恐吓 *wēixiékǒnghè*):

6: My contacts have been bombed. My father receives more than ten calls each day, and all of my friends are getting text messages.

7: I've had arrears of 65,000 for more than 4 months. I've been hounded with text messages and doorstep visits.

8: After being overdue for a week, I was called tens of times every day...the threatening text messages were endless. They also bombed my contacts, calling my friends and relatives tens of times a day. It's all threats and intimidation.

9: The chasing dog is insane. She called my workplace yesterday and kept saying "tell this deadbeat to pay back the loan." Then she hung up, called again from another phone number, and hung up again. She also called other departments and did the same thing.

4.2 Judicial system enforcement

Alongside debt collection by third-party companies, the category of keywords relating to the judicial system show that the prospect of being formally prosecuted is also frequently discussed. Keywords in this category include 传票 *cháuipiào* 'court summons', 律师函 *lǜshīhán* 'legal letter' and 坐牢 *zuòláo* 'be put in prison'. However, it is relevant to note that 吓唬 *xiàhu* (itself a keyword) collocates with many of these keywords. While 吓唬 means 'to scare', colloquially it implies bluffing, i.e., an intimidating threat which will not be implemented. Indeed, our concordance analysis shows a frequent response to concerns about prosecution is to state that it is simply a scare tactic used by debt collectors:

10: A legal letter is just a load of crap. They also produce fake court summons just to scare people.

4.3 Relationships with family, friends and colleagues

Given the practice of contact bombing, it is unsurprising that many forum members express high levels of anxiety about the impact of over-indebtedness on their social ties. Specifically, the keywords 爸妈 *bà mā* 'parents' and 家里人 *jiālǐ rén* 'family members' reflect key spheres of concern. These words have numerous negative collocates such as 连累 *liánlěi* 'to implicate' and 骚扰 *sāorǎo* 'harass'. A particular worry is revealed by the strong collocation between 'parents' and 身体 *shēntǐ* which literally

means ‘body’, but actually refers to ‘health’. Specifically, many forum members are afraid the distress caused by debt collection will irrevocably damage their parent’s mental and physical well-being:

11: My parents are nearly at breaking point, and I’m really worried about their health. The bank interest and penalties are terrifyingly high, and they can’t handle any more stress...They said they can’t sleep because of the constant collection calls. I feel so guilty.

Another noteworthy collocate of both ‘parents’ and ‘family members’ is 帮 *bāng* ‘help’, with many forum members recommending that turning to one’s parents for assistance is an advisable means of escaping indebtedness. The keyword 坦白 *tǎnbái* ‘to confess’ also collocates with both ‘parents’ and ‘family members’, with this act of openness and contrition often reported as necessary in securing such support:

12: I confessed today and my parents have said they will help me repay.

13: You should confess to your parents. Only they will help you unconditionally.

Alongside their families, forum members also discuss worries relating to their wider social networks. This is revealed by some of the words that collocate with ‘contact list’ including 单位 *dānwèi* ‘workplace’ and 好友 *hǎoyǒu* ‘friends’:

14: My contact list contains all of the leaders at my workplace. I am so worried I can’t sleep.

15: The collection company has my contact list and warned me that if I don’t repay by 1pm tomorrow, they will message all my friends, sending them the video of when I took out the loan. They told me that if I don’t want to be known as a deadbeat defaulter, I need to pay off all my debt immediately.

In terms of the perceived cumulative effect of the collection practices, forum members frequently conceptualise this via the keyword 社死 *shèsǐ* ‘social death’. This reflects deep anxiety that the damage which is being caused to their social networks will result in their complete isolation:

16: My family and some of my friends know about my debt, but I'm afraid it'll spread to my colleagues as well. That'll be true social death.

17: I'm numb from the crazy collection dogs constantly calling my family and colleagues. I'm dealing with absolute social death...I see no hope in life. With the New Year approaching, I'm supposed to see my parents and relatives, but I just can't face it. Even thinking about it makes my chest feel tight. I've self-harmed many times and have severe depression. At my worst, I've thought about how to end my life.

4.4 Relationships with forum members

Given the vulnerability often displayed, the forum has become an important site for the provision of emotional support. This is reflected in keywords such as 嗯嗯 *ēnēn*, the written form of a colloquial vocalisation to indicate encouragement and acknowledgement, and exhortations such as 共勉 *gòngmiǎn*, 'let's do this together!', 加油 *jiāyóu* 'come on!', and 恭喜 *gōngxǐ* 'congratulations!', with members routinely encouraging each other for various actions, e.g., making repayments, confessing, etc. Another significant aspect of interactions is the creation of neologisms. As reported earlier, forum members label debt collectors as 催狗 *cūigǒu* 'chasing dogs'. Notably, this word is absent from the reference corpus, indicating it may have developed via forum discussions.

As well as emotional support, the provision of practical advice is another important forum function. In particular, one of the main issues that feature in such discussions is indicated by the keyword 协商 *xiéshāng* 'to negotiate' and its strongest collocate 怎么 *zěnmě* 'how', with forum members discussing how to agree terms with lenders to extend repayment deadlines and avoid debt collection. This collocation reflects the fact that there is a thirst for information about how to navigate over-indebtedness and that in the absence of alternative resources, debtors are turning to each other to create a joint repository of knowledge. The keywords 有没有 *yǒuméiyǒu* 'do you have or not' and 多久 *duōjiǔ* 'how long' reflect a similar dynamic of forum members requesting and providing information. Additional collocates of 'negotiate' include the names of various banks and internet lenders. However, there is no collocational relationship between 'negotiate' and 'chasing dogs'. Instead, when forum members discuss how to respond to the actions of collection agencies, they most frequently use the

keyword 投诉 *tóusù* ‘make a complaint’. This reflects many forum members’ belief that the coercive practices which collectors employ violate their rights:

18: We live in a society ruled by laws. If you are being violently harassed, you should make a complaint immediately.

19: This sort of violent pressurised debt collection infringes your privacy and rights. Make a complaint!

4.5 Relationships with regulatory agencies

In terms of who members recommend complaining to, this is revealed by the collocates of ‘make a complaint’. In particular, 银监会 *yínjiānhuì* ‘Banking Regulatory Commission’, 保监 *bǎojiān* ‘Insurance Regulatory Commission’ and 信访 *xìnfǎng* ‘Petitions Bureau’ are three of the strongest collocates of ‘make a complaint’. All are also keywords. The first two are the main financial regulatory agencies, while the Petitions Bureau is the formal channel that exists in China for citizens to bring grievances to the attention of the government. These keywords highlight that another highly significant aspect of over-indebtedness concerns the interactions it fosters with such bodies.

However, while it is common to recommend complaining, doing so is frequently reported as ineffective. Specifically, this is shown by the collocation between agencies such as the Petitions Bureau and the Banking Regulatory Commission and ‘没用’ *méiyòng* ‘useless’ (itself a keyword). The frustration often expressed is captured in the following excerpts:

20: I'm so annoyed! I've already complained to the Banking Regulatory Commission twice, and I've filed two petitions, but its useless, nothing's worked.

21: Last week, I experienced violent pressurised debt collection from Huabei. I filed a complaint through the national petition system...The very next day, someone from the petition bureau came to my village. They said they'd follow up, but nothing has come of it. Just a couple of days ago, the collector called again. So, even filing complaints with the national petition system is useless.

4.6 Time horizons

Several keywords relate to time. Of these, the most significant is 逾期 *yúqī* ‘to exceed the deadline/become overdue’. In particular, for many forum members becoming overdue marks the crossing of a crucial temporal boundary. This is unsurprising because it is often immediately afterwards that they are contacted by collection agencies. It is also relevant that ‘become overdue’ collocates with several other time markers such as 今天 *jīntiān* ‘today’, 明天 *míngtiān* ‘tomorrow’, 开始 *kāishǐ* ‘start’, and 就 *jiù* ‘just’, suggesting that it is frequently the shock of being exposed to debt collection that drives many of them to join the forum:

22: I’ve just become overdue. JD’s debt collection agency called me. They were very arrogant and told me that if I didn’t pay back what I owe immediately they’ll contact my family.

23: I want to ask, how long will it take for my contacts to be bombed when I’m overdue? It’s just I’m going to become overdue this month.

Another important time-related issue created by over-indebtedness, discussed in the literature review, concerns the narrowing of temporal horizons which can lead to harmful short-term decisions. Two keywords suggest that this phenomenon is occurring here. The first is 贷养贷 *dàiyǎngdài* ‘a loan to care for a loan’. The second is 拆东墙补西墙 *chāidōngqiángbǔxīqiáng*. This means ‘tear down the east wall to repair the west wall’, and metaphorically indicates a temporary solution that leaves the cause of a problem unaddressed. Specifically, both indicate that many forum members are borrowing to cover other debt repayments. As strong collocates of ‘a loan to care for a loan’ include 雪球 *xuěqiú* ‘snowball’, 死路 *sǐlù* ‘dead end’ and 挖坑 *wākēng* ‘dig a hole’, this indicates that it is often precipitating a debt spiral:

24: I am still in a vicious cycle of borrowing to take care of my loans. The money I have to pay back this month far exceeds my salary income. I can’t make ends meet.

25: I have 20,000 – 30,000 in loans that I have taken to take care of other loans. It’s a dead end.

26: I used to run a business and relied on loans to keep it going. My income was stable, but I had already started tearing down the west wall to repair the east wall. Later, I switched to live streaming, but that failed too. Without a stable income, everything just collapsed.

4.7 Rationalities

Several keywords link to how the experience of over-indebtedness is affecting forum members' rationality, i.e., their goals and the decisions taken to pursue them. One of the most prominent is 上岸 *shàngàn*. This translates to 'come ashore' and is the shared terminology that members use to describe the goal of paying off their debts and returning to a more secure life. Several other keywords express a similar meaning, e.g., 还完 *huánwán* and 还清 *jiéqīng*. Nearly all of the strong collocates of these keywords relate to one of two sentiments, urgency, e.g., 赶紧 *gǎnjǐn* 'as quickly as possible', or difficulty, e.g., 争取 *zhēngqǔ* 'to struggle/strive'.

27: I am striving to come ashore as soon as possible.

28: Make good plans, work hard to make money, and strive to come ashore as soon as possible. Let's do this together.

Another goal that the keywords suggest forum members are keen to achieve relates to managing how they are perceived. This is revealed by the keyword 老赖 *lǎolài* which is a slang term for someone who deliberately does not pay their debts. It can be translated as 'deadbeat defaulter'. As concordance excerpt 15 indicates, it is evident that collection agents use it in their interactions with debtors as a means of applying psychological pressure ('if you don't want to be known as a deadbeat defaulter...') and concordance excerpts reveal that many forum members are particularly keen to underline that this term is unfair and does not apply to them:

29: Actually, as most of our friends in the group have said, we're not deadbeat defaulters who refuse to pay. The truth is we've run into some real financial problems and can't make the payments right now.

30: I don't think I am a deadbeat defaulter. A deadbeat defaulter is someone who has money but is unwilling to pay it back. I didn't say I won't pay it back. I'm just paying off my credit card now and don't have that much money to repay the other instalments.

31: Being overdue and being a deadbeat defaulter are two completely different things. We're not deadbeat defaulters. This point must be clearly understood.

However, while attitudes relating to repaying and maintaining one's reputation dominate this category's keywords, two other keywords suggest they are not universal. In particular, 躺平 *tǎngpíng* 'to lie flat' and 摆烂 *bǎilàn* 'to let things fall apart' indicate that for some their over-indebtedness has created resignation and demotivation:

32: I've chosen to lie flat. I've given up on myself and returned to the countryside.

33: Don't worry, brother. I used to be like you, afraid of losing face. But now I'm letting everything fall apart. They can call whoever they want. I just don't answer when they call me. I've been overdue for more than half a year now.

34: I've completely given up and am letting everything fall apart. Whatever happens, happens.

5. Discussion

We set out to examine what issues over-indebted individuals in China choose to discuss online and what their linguistic choices suggest about how they conceive the impact. As our results demonstrate, forum members are deeply troubled by the actions of debt collectors, which they frame as a form of systematic violence. Of particular concern is contact bombing. Through our analysis, we have uncovered how the distress caused by this practice infiltrates multiple dimensions of borrowers' lives, spanning their relationships, time horizons, and decision-making processes. In contrast to debt collection in many other jurisdictions, contact bombing appears unusually intrusive, with collection agencies seemingly able to access and exploit a broader range of personal information than is typical. Given the extensive international literature on the harmful consequences of aggressive debt collection, our findings of numerous negative outcomes among Chinese borrowers are, unfortunately, unsurprising.

However, although the consequences may be predictable, that does not mean they are straightforward. For instance, a noteworthy aspect of these collection practices revealed by our

analysis is they generate a complex social paradox, namely, while they corrode forum members' existing relationships, they also demonstrate a socially generative potential. Specifically, it seems it is often becoming overdue and the shock of being subjected to intimidatory debt collection that leads individuals to the forum where they engage in behaviours reflective of community building. These dynamics materialise through two primary mechanisms - the widespread provision of practical advice and emotional support, as well as collective linguistic innovation, such as the neologism 'chasing dogs'. These observations suggest the emergence of a jointly constructed social space where individuals transform personal trauma into a form of collective meaning-making. This echoes Deville's (2016, p.16) conclusion concerning the potential for online forums to facilitate 'debtor publics'.

Furthermore, our results indicate that the issue of debt collection also has saliency due to the relations it generates with regulators. Specifically, forum members often express anger at collection practices and the perceived violation of the privacy and consumer rights many feel entitled to as citizens of an increasingly developed and prosperous society (Duan & Zhang, 2021). This frequently leads them to turn to regulatory agencies for recourse. However, the collocation patterns relating to these bodies indicate that forum members are often dissatisfied with the support they receive. Although friction arising from an increasingly rights conscious public and the enforcement of those rights has been documented in several areas across Chinese society (e.g., environmental protection, labour regulation, etc.), if the current economic slowdown continues and over-indebtedness increases, the actions of the finance and debt collection industries may become an increasingly prominent source of tension in this regard.

Given such insights, our study clearly helps to illuminate this important unfolding social phenomenon in China, while also enriching the international literature on over-indebtedness by generating detailed findings from an understudied setting. However, we believe its contribution goes beyond its empirical results and the topic of household debt. Specifically, from a methodological standpoint, it also shows the potential of corpus linguistics research to support sociological inquiry in China. This is significant because until now this method has been relatively under-utilised, particularly in relation to large natural language Mandarin datasets. Instead, when corpus linguistics

techniques have been employed, they have generally either been done so in a learner corpus context to investigate the second language acquisition of Mandarin (e.g., Xu et al., 2019), or to investigate discourse in the English-language versions of Chinese state media relative to international outlets (e.g., Liu & Yu, 2023).

Looking ahead, we believe that corpus linguistics holds promise for Chinese sociological research for two key reasons. First, as noted earlier, the country's political environment is constraining the ability to conduct fieldwork. By facilitating remote data collection without direct participant involvement, corpus linguistics offers a valuable alternative. Second, corpus techniques excel at uncovering subtle themes and evaluative stances embedded in language. This is important because given the political climate, people are ever more wary of directly discussing potentially sensitive issues. Thus, probing attitudes via natural language data is one way of partially navigating this challenge. While we do not claim that this means corpus methods can overcome the constraints of censored data and curtailed public discourse, they nevertheless offer a potentially more revealing approach compared to traditional textual analysis in such a challenging context.

However, while we advocate the value of corpus linguistics to support sociological research in China, our results also show that to be done successfully, this method must be combined with strong subject-area knowledge. This is because while corpus techniques can signpost potentially salient issues arising from natural language data, they require careful interpretation. Two examples from our results underscore this.

The first relates to contact bombing. While this practice would be harmful anywhere, the intensity of distress our analysis reveals it generates arguably can only be adequately understood in relation to specific aspects of Chinese society. In particular, because abstract social trust in China is relatively weak (Yao et al., 2017), individuals must often rely on personal trust as a form of social insurance and a tool to access resources. This means social networks are an indispensable type of capital. As such, the potential destruction of these networks by debt collectors is deeply disturbing, not only due to shame and inconvenience, but also because it threatens potentially ruinous isolation.

The second example concerns interpretation of conflicting findings. For instance, we detected contrasting attitudes to being a defaulter relative to Rao and McDonald (2023). Specifically, as

reported, they found that the Chinese debtors they investigated accepted and sought to ‘remoralize’ being a defaulter (p.269). However, our results suggest that many forum members strongly refute this identity. Here contextual knowledge is necessary to understand why this divergence may arise due to the studies’ samples. In particular, Rao and McDonald investigated extremely disadvantaged borrowers (migrant workers in informal factory employment), whereas the individuals who use the Debtor Alliance forum are likely to occupy higher socio-economic positions. This is significant as the disparity may affect the value individuals place on their social networks. For instance, while the most disadvantaged in China are unlikely to have been able to construct high-utility networks due to their marginalisation, meaning they are less concerned about attempts to weaponise them, more-advantaged individuals may feel they have more to lose, especially during a period of heightened economic uncertainty. Therefore, while corpus linguistics has significant promise for sociological research in China, it will only be effective when it integrates methodological skills with subject-area expertise to enable valid interpretation of the results.

Finally, although our study has deepened knowledge of this important topic, it is important to restate its main limitations. First and foremost are the circumstances of the Chinese internet. Specifically, while it is clear we have generated valuable insights from the information that individuals have chosen to share, it is nonetheless likely that the constraints to public discourse that exist in China will have influenced what forum users are willing to discuss. This must be kept in mind when interpreting the study. Secondly, given what we know of a typical Douban user, we must reiterate that we make no claims to generalisability. Instead, our findings remain indicative of the types of issues those struggling with over-indebtedness in China face. They therefore require validation across different regions and groups. Finally, it is important to note that China’s financial regulation continues to progress rapidly and if the actions of debt collectors can be successfully curtailed, this will significantly improve the landscape for debtors. Therefore, further research will be needed to understand the impact of ongoing regulatory change.

6. Conclusion

Our study helps to illuminate how growing household over-indebtedness is intertwining with China's unique social, institutional and cultural framework. In particular, our results indicate that Chinese debtors are often subjected to aggressive debt collection practices and the resulting distress infiltrates multiple aspects of their lives. We have also revealed that these practices are generating a complex paradox, namely, they are simultaneously both socially destructive and generative. Finally, alongside our empirical results, our study has important methodological implications. In particular, we have demonstrated how corpus linguistics can be used to support meaningful sociological research in China, which to date has rarely been accomplished. Therefore, our study lays the ground not only for further detailed investigation of the micro-level impact of rising household debt in China, but also for a more methodologically-diverse approach to Chinese sociology.

References

- Achtziger, A. (2022). Overspending, debt, and poverty. *Current Opinion in Psychology*, 46, 101342.
<https://doi.org/10.1016/j.copsyc.2022.101342>
- Baker, P., & McEnery, T. (2015). Introduction. In P. Baker & T. McEnery (Eds.), *Corpora and Discourse Studies: Integrating Discourse and Corpora* (pp. 1–19). Palgrave Macmillan UK.
https://doi.org/10.1057/9781137431738_1
- Brookes, G., & Baker, P. (2022). Fear and responsibility: Discourses of obesity and risk in the UK press. *Journal of Risk Research*, 25(3), 363–378.
<https://doi.org/10.1080/13669877.2020.1863849>
- Cesar Leandro, J., & Botelho, D. (2022). Consumer over-indebtedness: A review and future research agenda. *Journal of Business Research*, 145, 535–551.
<https://doi.org/10.1016/j.jbusres.2022.03.023>
- China-Digital. (2023). *DOUBAN: The ecosystem and opportunities of the Chinese social network*.
<https://china-digital.com/blogs/overview-douban/>
- Clark, A. A., Collard, S., Holme, C., Jenkins, D., Duucale, E., & Osman, I. (2022). *Community-centred services support residents to build financial resilience*. University of Bristol.
- Coltman-Patel, T., Dance, W., Demjen, Z., Gatherer, D., Hardaker, C., & Semino, E. (2022). ‘Am I being unreasonable to vaccinate my kids against my ex’s wishes?’ – A corpus linguistic exploration of conflict in vaccination discussions on Mumsnet Talk’s AIBU forum. *Discourse, Context & Media*, 48, 100624. <https://doi.org/10.1016/j.dcm.2022.100624>
- D’Alessio, G., & Iezzi, S. (2013). *Household Over-Indebtedness: Definition and Measurement with Italian Data* (SSRN Scholarly Paper 2243578). Social Science Research Network.
<https://doi.org/10.2139/ssrn.2243578>
- Deville, J. (2015). *Lived Economies of Default: Consumer Credit, Debt Collection and the Capture of Affect*. Routledge. <https://doi.org/10.4324/9780203383254>

- Deville, J. (2016). Debtor publics: Tracking the participatory politics of consumer credit. *Consumption Markets & Culture*, 19(1), 38–55.
<https://doi.org/10.1080/10253866.2015.1068169>
- Disney, R., & Gathergood, J. (2013). Financial literacy and consumer credit portfolios. *Journal of Banking & Finance*, 37(7), 2246–2254.
- Duan, Y., & Zhang, G. (2021). The Emerging Consciousness of Citizens' Rights in China: Interest Recognition and Generational Variation. *Journal of Contemporary China*, 30(127), 152–167.
<https://doi.org/10.1080/10670564.2020.1766905>
- Finney, A., Collard, S., & Kempson, E. (2007). *Easy come, easy go: Borrowing over the life cycle*. University of Bristol. <https://www.bristol.ac.uk/geography/research/pfrc/themes/credit-debt/easy-come.html>
- Graeber, D. (2011). *Debt: The First 5,000 Years*. Melville House. <https://davidgraeber.org/books/debt-the-first-5000-years/>
- Harker, C. (2017). Debt space: Topologies, ecologies and Ramallah, Palestine. *Environment and Planning D: Society and Space*, 35(4), 600–619. <https://doi.org/10.1177/0263775816686973>
- Hu, M., Nie, W., Song, J., Wang, T., & Ye, X. (2023). Relationship between household financial debt and depressive symptoms: A longitudinal study in China. *BMJ Open*, 13(9), e074024.
<https://doi.org/10.1136/bmjopen-2023-074024>
- Huang, C.-R., Kilgarriff, A., Wu, Y., Chiu, C.-M., Smith, S., Rychlý, P., Bai, M.-H., & Chen, K.-J. (2005). Chinese Sketch Engine and the extraction of grammatical collocations. *Proceedings of the Fourth SIGHAN Workshop on Chinese Language Processing*.
<https://aclanthology.org/I05-3007.pdf>
- IMF. (2024). *Global Debt Database—Household debt, loans and debt securities*.
https://www.imf.org/external/datamapper/HH_LS@GDD
- Kaye, L., Hewson, D. C., & Buchanan, T. (2021). *Ethics guidelines for internet-mediated research*. BPS. <https://www.bps.org.uk/guideline/ethics-guidelines-internet-mediated-research>
- Kogen, L. (2024). Qualitative Thematic Analysis of Transcripts in Social Change Research: Reflections on Common Misconceptions and Recommendations for Reporting Results.

International Journal of Qualitative Methods, 23, 16094069231225919.

<https://doi.org/10.1177/16094069231225919>

Lawson, R., Lutzky, U., Kehoe, A., & Gee, M. (2023). “Sorry to hear you’re going through a difficult time”: Investigating online discussions of consumer debt. *Applied Corpus Linguistics*, 3(2), 100056. <https://doi.org/10.1016/j.acorp.2023.100056>

Lea, S. E. G. (2021). Debt and Overindebtedness: Psychological Evidence and its Policy Implications. *Social Issues and Policy Review*, 15(1), 146–179. <https://doi.org/10.1111/sipr.12074>

Liu, J. (2024). Internet Censorship in China: Looking Through the Lens of Categorisation. *Journal of Current Chinese Affairs*, 18681026231220948. <https://doi.org/10.1177/18681026231220948>

Liu, J., He, X., & Dong, Y. (2024). Household debt and children’s psychological well-being in China: The mediating role of parent–child relations. *Children and Youth Services Review*, 157, 107387. <https://doi.org/10.1016/j.childyouth.2023.107387>

Liu, P., Zhou, L., Tian, Y., & Nie, W. (2021). Association between household debt and depressive mood among Chinese residents. *Public Health*, 194, 202–207. <https://doi.org/10.1016/j.puhe.2021.03.015>

Liu, S., & Yu, H. (2023). What is newsworthy about Covid-19? A corpus linguistic analysis of news values in reports by China Daily and The New York Times. *Humanities and Social Sciences Communications*, 10(1), 1–11. <https://doi.org/10.1057/s41599-023-02241-5>

Liu, Z., Zhong, X., Zhang, T., & Li, W. (2020). Household debt and happiness: Evidence from the China Household Finance Survey. *Applied Economics Letters*, 27(3), 199–205. <https://doi.org/10.1080/13504851.2019.1610706>

Lombardi, M. J., Mohanty, M., & Shim, I. (2017). *The real effects of household debt in the short and long run*. <https://www.bis.org/publ/work607.htm>

Lustig, A., & Brookes, G. (2022). Construction of group norms in a radical acceptanceonline forum for heavy alcohol users: A corpus-based discourse analysis. *The International Journal on Drug Policy*, 109, 103862. <https://doi.org/10.1016/j.drugpo.2022.103862>

Mazzocco, I., & Kennedy, S. (2022, August 29). *Doing Fieldwork in China: Evolution and Prospects*. Big Data China. <https://bigdatachina.csis.org/doing-fieldwork-in-china/>

- Montgomerie, J., & Tepe-Belfrage, D. (2017). Caring for Debts: How the Household Economy Exposes the Limits of Financialisation. *Critical Sociology*, 43(4–5), 653–668. <https://doi.org/10.1177/0896920516664962>
- Rao, Y., & McDonald, T. (2023). Debt at a distance: Counter-collection strategies and financial subjectivities of China's working-class defaulters during COVID-19. *Economy and Society*, 52(2), 250–273. <https://doi.org/10.1080/03085147.2022.2154501>
- SDG. (2024, October 24). *Top Chinese Social Media Platforms in 2024: A Comprehensive Guide for Marketers*. Sekkei Digital Group. <https://sekkeidigitalgroup.com/top-chinese-social-media-platforms-2025/>
- Slotta, D. (2024). *China: Credit Card Number*. Statista. <https://www.statista.com/statistics/450048/total-number-of-credit-cards-in-china/>
- Song, H., Wang, R., Bishwajit, G., Xiong, J., Feng, Z., & Fu, H. (2020). Household debt, hypertension and depressive symptoms for older adults. *International Journal of Geriatric Psychiatry*, 35(7), 779–784. <https://doi.org/10.1002/gps.5302>
- Song, J., Hu, M., Li, S., & Ye, X. (2023). The Impact Mechanism of Household Financial Debt on Physical Health in China. *International Journal of Environmental Research and Public Health*, 20(5), 4643. <https://doi.org/10.3390/ijerph20054643>
- Sparkes, M., Gummy, J., & Burchell, B. (2018). Debt: Beyond homo economicus. In *The Cambridge handbook of psychology and economic behaviour, 2nd ed* (pp. 198–233). Cambridge University Press. <https://doi.org/10.1017/9781316676349.007>
- Stănescu, C. G. (Ed.). (2022). *Regulation of Debt Collection in Europe: Understanding Informal Debt Collection Practices*. Routledge. <https://doi.org/10.4324/9781003343158>
- Stanley, L., Deville, J., & Montgomerie, J. (2015). Digital Debt Management: The Everyday Life of Austerity. *New Formations*, 2015(87). <https://doi.org/10.3898/NEWF.87.4.2016>
- Stout, N. (2016). #Indebted: Disciplining the Moral Valence of Mortgage Debt Online. *Cultural Anthropology*, 31(1), Article 1. <https://doi.org/10.14506/ca31.1.05>

- Sweet, E., Nandi, A., Adam, E., & McDade, T. (2013). The High Price of Debt: Household financial debt and its impact on mental and physical health. *Social Science & Medicine (1982)*, *91*, 94–100. <https://doi.org/10.1016/j.socscimed.2013.05.009>
- Turunen, E., & Hiilamo, H. (2014). Health effects of indebtedness: A systematic review. *BMC Public Health*, *14*(1), 489. <https://doi.org/10.1186/1471-2458-14-489>
- World Bank. (2024). *Overview—China* [Text/HTML]. World Bank. <https://www.worldbank.org/en/country/china/overview>
- Wu, P. (2021). *Online Feminist Groups Seek New Home as Douban Chills*. <https://www.sixthtone.com/news/1009258>
- Xiao, J. J., Yan, C., Bialowolski, P., & Porto, N. (2021). Consumer debt holding, income and happiness: Evidence from China. *International Journal of Bank Marketing*, *39*(5), 789–809. <https://doi.org/10.1108/IJBM-08-2020-0422>
- Xu, H., Lu, X., & Brezina, V. (2019). Acquisition of the Chinese Particle *le* by L2 Learners: A Corpus-Based Approach. In X. Lu & B. Chen (Eds.), *Computational and Corpus Approaches to Chinese Language Learning* (pp. 197–216). Springer Nature. https://doi.org/10.1007/978-981-13-3570-9_10
- Yao, J., Zhang, Z.-X., Brett, J., & Murnighan, J. K. (2017). Understanding the trust deficit in China: Mapping positive experience and trust in strangers. *Organizational Behavior and Human Decision Processes*, *143*, 85–97. <https://doi.org/10.1016/j.obhdp.2016.12.003>
- Zappavigna, M. (2011). Ambient affiliation: A linguistic perspective on Twitter. *New Media & Society*, *13*(5), 788–806. <https://doi.org/10.1177/1461444810385097>

Declaration of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The author is an Editorial Board Member/Editor-in-Chief/Associate Editor/Guest Editor for this journal and was not involved in the editorial review or the decision to publish this article.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Jiaqi Feng Guo

John Marshall