

Xinjia Peng, Jun Lang and Zhuo Jing-Schmidt*

Core-grammatical innovation in real time: radical and rapid constructionalization in Chinese

<https://doi.org/10.1515/cog-2025-0167>

Received October 20, 2025; accepted September 23, 2026; published online October 8, 2026

Abstract: Recent research in Diachronic Construction Grammar and Creative Construction Grammar has begun to uncover nongradual cases of linguistic creativity. Two gaps nevertheless remain: the scarcity of documented cases of radical and rapid innovation in core grammar, and the need to specify the converging mechanisms by which such innovation arises and spreads in real time. This study addresses both through the Anticanonical Passive Construction (APC) in Chinese, in the form [*bèi* X], where X violates conventional passivization constraints. As a form-meaning innovation in core grammar, the APC radically retools the canonical passive into a resource for the satirical construal of subjects as undergoing coercion, loss of agency, and distortion of truth. Combining corpus and survey data, we show that the APC spread rapidly, admitted a wide range of canonically non-passivizable predicates, and is represented in native-speaker knowledge as both interpretable and productively extendable. We explain this development through three converging mechanisms: cognitive biases, including anomaly-triggered grammatical surprise and heightened attention to negative events; joint affect, which facilitates interpretive uptake; and platform amplification, whereby digital affordances and incentives accelerate diffusion. The study expands current accounts of constructionalization by documenting and explaining a core-grammatical innovation in real time.

Keywords: constructionalization; core-grammatical innovation; anticanonical passive; Chinese

*Corresponding author: **Zhuo Jing-Schmidt**, East Asian Languages and Literatures, University of Oregon, Eugene, OR, USA, E-mail: zjingsch@uoregon.edu. <https://orcid.org/0000-0002-9164-7824>

Xinjia Peng, Key Laboratory of Linguistics, Corpus and Computational Linguistics Center, Institute of Linguistics, Chinese Academy of Social Sciences, Beijing, China. <https://orcid.org/0000-0001-6803-4678>

Jun Lang, Asian Languages and Literatures, Pomona College, Claremont, CA, USA. <https://orcid.org/0000-0003-3572-4677>

1 Introduction

1.1 Theoretical background

Linguistic research on adult-led grammatical innovation has long been guided by a gradualist model supported by robust diachronic evidence across languages. Its clearest expression is morpheme-based grammaticalization, in which lexical items develop into grammatical morphemes along “evolutionary pathways” (Heine and Reh 1984: 15; Lehmann 1985). Diachronic Construction Grammar, while treating constructions as the basic units of change, likewise emphasizes gradual changes leading to constructionalization (Fried 2013; Traugott and Trousdale 2010, 2013). Accordingly, Traugott and Trousdale (2010: 7) define grammaticalization as “a constructional (form-meaning) change that occurs in micro-steps,” and Traugott and Trousdale (2013) describe constructionalization as the creation of a new form-meaning pairing through a succession of microsteps. Well-established, the gradual model of grammatical innovation remains a highly fruitful framework in historical linguistics.

At the same time, Diachronic Construction Grammar has begun to attend more explicitly to nongradual aspects of constructional innovation. Traugott and Trousdale (2013: §4.8), for example, observe “instant node creation” in several lexical patterns, noting that while the schema may develop gradually, the individual lexical micro-constructions modeled on it are “clearly not gradual developments.” Although the general model remains incremental overall, this recognition of instant node creation points toward growing interest in creative and discourse-driven innovation.

Methodologically, however, Diachronic Construction Grammar has, as Hilpert (2021: 61) notes, still relied largely on historical documents. More recently, scholars have increasingly addressed this limitation through the study of emerging constructions in contemporary English. Following Traugott and Trousdale’s (2013: 184) account of “snowclones” as instances of “lexical constructionalization of a schema,” studies have examined playful and stance-marked semi-schematic patterns: Michaelis and Feng (2015) on the sarcastic split interrogative construction, Bohmann (2016) on [*because* X], Hilpert and Bourgeois (2020) on *sarcastic much?*, and Hartmann and Ungerer (2024) on two cases of creative snowclones. Hoffmann and Turner (2025) bring this line of research into theoretical focus in Creative Construction Grammar, arguing that constructional blending provides a domain-general mechanism of constructional creativity. Their proposal finds an early precursor in Hermann Paul’s concept of *Konstruktionsmischung* as a key mechanism of change (Paul 1909: §118), a line of thought largely overlooked outside German language scholarship.

While the innovations discussed above have greatly advanced our understanding of linguistic creativity in use, they mostly involve rhetorical or stylistic patterns rather than innovation in the core grammar of event representation. Similarly, while blending offers a valuable account of certain types of constructional creativity, it is less well suited to radical rule-shifting constructionalization in online discourse, which requires a broader explanatory model that identifies and integrates not only cognitive but also affective, pragmatic, social and technological factors. These limitations suggest two gaps in current accounts of constructional change: the scarcity of documented cases of nongradual innovation in core grammar, and the need to specify the converging mechanisms by which such innovation arises and spreads in real time. The present study addresses both through the case of the Anticanonical Passive Construction (APC) in contemporary Chinese online discourse, which combines radical retooling of a core grammatical construction with rapid digital expansion shaped by cognitive, affective, and sociotechnological forces.

1.2 The canonical passive construction (CPC) in Mandarin Chinese

The Mandarin Chinese CPC, commonly referred to as the *bèi* construction, takes the form $[NP_1 \text{ } bèi \text{ } (NP_2) \text{ } VP]$, where *bèi* marks the clause as passive and introduces an event in which the subject (NP_1) is affected by the action of an external agent (NP_2), which may be expressed or omitted (Sun 1996). The verb phrase (VP) is obligatorily transitive, capable of assigning an agent role to NP_2 and a patient role to NP_1 . Consider the constructed examples in (1) and (2):¹

- (1) 他被姐姐骂了
Tā bèi jiějie mà le
 3SG PASS older.sister scold PFV
 ‘He was scolded by his older sister.’
- (2) 我被抢了
wǒ bèi qiǎng le
 1sg pass rob pfv
 ‘I was robbed.’

In both cases, the verb (*scold*, *rob*) is transitive and projects a patient role for the subject. The syntactic requirement for a transitive predicate aligns with

¹ Abbreviations used in the glosses: 1sg = first person singular; 3sg = third person singular; PASS = passive; PFV = perfective.

crosslinguistic norms of prototypical passive constructions (Comrie 1977; Givón 1979; Keenan 1975; Shibatani 1985).

Semantically, the CPC entails a negative outcome for the affected subject and is often described as the “adversative passive” (Chao 1968: 703; Li and Thompson 1989; Sun 1996; Wang 1957). This feature is typologically salient among passive constructions in Asian languages (Givón 1979; Shibatani 1985). It also restricts the CPC to a relatively narrow set of transitive predicates associated with adverse consequences.² Crucially, intransitive, nominal, and other nontransitive predicates cannot be passivized with *bèi*. For this reason, novel expressions such as *bèi zìshā* ‘be suicide-ed’ and *bèi zìyuàn* ‘be voluntary-ed,’ are syntactically anomalous. However, it is precisely such rule-violating usages that instantiate the grammatical innovation at the center of this study – the APC – to which we now turn.

1.3 The anticanonical passive construction

The APC, schematized as [*bèi* X], emerged in early 21st-century Chinese digital discourse. It inherits the passive marker *bèi* from the CPC and retains its canonical semantic prosody but departs radically from the canonical passive syntax: X is explicitly nontransitive. The construction represents a core grammatical innovation that radically reconfigures the CPC. The news headline in (3) offers an illustration:

- (3) 学生家长“被自愿”购买“天价跳绳”
Xuésheng jiāzhǎng “bèi zìyuàn” gòumǎi “tiānjià tiàoshéng”
 student parent PASS voluntary purchase sky-high.price jump.rope
 ‘Students’ parents were “voluntary-ed” into purchasing “sky-high-priced jump ropes.”
 (Hebei Consumer Rights Network, 5 April 2013)

In this example, 被自愿 *bèi zìyuàn* ‘be voluntary-ed’ is a novel expression that violates canonical passive syntax. The stative, nontransitive *zìyuàn* ‘be voluntary’ occupies the X slot even though it cannot be ordinarily passivized. Although the institutional source of coercion is inferable from the discourse, no agent is expressed in the APC itself. The construction thus reconfigures the passive form-meaning pairing and its event structure: instead of profiling a canonical transitive event in which an agent acts on a patient, it construes the subject as affected by an externally imposed narrative or status. Here, parents compelled to make a purchase are

² In contemporary Chinese, neutral and even positive usages have developed under the influence of European languages (Xiao et al. 2006).

ironically represented as having done so voluntarily, exemplifying a defining function of the emergent APC.

The earliest documented instance of APC appeared in spring 2008 after the suspicious prison death of a whistleblower drew national attention. Although the forensic report ruled the death a suicide, the political sensitivity of the case fueled public doubt about the official narrative. In this context, a new expression – *bèi zìshā* ‘be suicide-ed’ – appeared and quickly went viral. By forcing the non-passivizable ‘suicide’ into a passive frame, the expression was as syntactically jarring as the event was shocking. In doing so, it instantiated a core-grammatical innovation: the passive schema was extended beyond a canonical transitive event to encode an externally imposed account of what happened to the subject.

This initial exemplar soon sparked a wave of related innovations, many likewise prompted by salient events widely reported online. One prominent case was *bèi jiùyè* ‘be job-ed,’ which emerged in summer 2009 after a college graduate posted on social media about being unemployed despite the university’s boast of high placement rates. The sarcastic expression captured the frustration of graduates whose names had been entered into fake employment agreements without their consent. The use of the intransitive *jiùyè* ‘to enter the workforce’ in the passive frame signals both distortion of truth and erasure of agency.³ Whereas *bèi zìshā* responded to an isolated case of suspicious death, *bèi jiùyè* gave voice to a widely shared grievance among college graduates, making the usage socially resonant. It quickly spawned similar expressions like *bèi zēngzhǎng* ‘be growth-ed’ and *bèi píngjūn* ‘be average-ed,’ extending the construction to the critique of institutional metrics more broadly.

New coinages soon became increasingly outlandish, recruiting nominal, adjectival, and nontransitive compounds in the X slot, as in *bèi gāotiě* ‘be high-speed-rail-ed,’ *bèi fùyù* ‘be wealthy-ed,’ *bèi dì-yī* ‘be number one-ed,’ *bèi dāngdiē* ‘be become-father-ed.’ As the range of lexical categories admitted into the X slot expanded, the construction’s departure from canonical passive syntax became progressively more radical. These expressions construe the subject as subjected to an externally imposed process, state, role, or outcome. Voted by netizens as an “expression of the year” in 2009, the new usage attracted attention in state media and was eventually recognized in the *Modern Chinese Dictionary* in 2012.

3 When glossing individual X-slot fillers, we follow their conventional lexical meaning (e.g., *jiùyè* ‘enter the workforce’); when glossing whole APC expressions, we use deliberately non-idiomatic renderings such as ‘be job-ed’ to reflect the construction’s syntactic anomaly.

1.4 Previous studies on APC

The APC has attracted considerable attention among Chinese linguists. The more substantive studies cited here (Chen and Hu 2020; Huang and Liu 2014; Jing-Schmidt 2024; Peng 2017; Shen 2010; Shi 2013; Yao et al. 2013; Yuan and Zhang 2026) broadly agree that grammatical deviance and stance-taking are central to the construction. This characterization has also entered normative reference works: the *Modern Chinese Dictionary* (7th edition, 2016: 59) glosses the APC as indicating “untrue” or “coerced” circumstances “with a sarcastic undertone.” While this semantic-pragmatic summary reflects broad consensus, some studies distinguish two constructional subtypes corresponding to these senses (e.g., Huang and Liu 2014; Peng 2017; Shen 2010; J. Wang 2011; Y. Wang 2011), whereas others emphasize contextual inferences from a general meaning of falsehood (e.g., Chen and Hu 2020; Shi 2013; Zhang 2020).

Although previous studies have examined the semantic, syntactic, and pragmatic dimensions of the APC, few have situated it within broader discussions of grammatical change. Huang and Liu (2014) interpret the APC as a typological shift from analytic back to synthetic structure in modern Chinese, but do not address the temporal profile of its emergence. Xiao (2016), invoking grammaticalization and reanalysis, similarly adopts historical concepts without supporting them with historical evidence. Yuan and Zhang (2026) offer one of the most systematic recent descriptions of the APC’s usage trajectory, based on data from a small self-compiled news corpus. Their snowclone analysis, however, is ontologically unconvincing: although the construction is culturally salient, productive, and extravagant, its extensions are grounded less in a fixed lexical source than in the radical reworking of the canonical passive construction as a core-grammatical construction. Thus, while their study connects the APC to current debates in Creative and Diachronic Construction Grammar, it leaves unresolved the fundamental question of how its constructionalization relates to innovation in core grammar. The present study takes up this issue by examining the APC explicitly as a case of radical and rapid core-grammatical innovation.

1.5 This study

This study investigates the APC’s frequency profile, categorial expansion, rate of constructionalization, semantic coverage, current representation in native-speaker knowledge, and explains its mechanisms of innovation.

Our analysis is situated within Diachronic Construction Grammar, which takes constructions to be the locus of grammatical change (Barðdal and Gildea 2015; Noël

2007; Traugott and Trousdale 2013) and within Creative Construction Grammar, which focuses on linguistic creativity from a cognitive-constructionist perspective (Hoffmann and Turner 2025). We also adopt the functionalist view that change is shaped both by language internal features and by external forces that condition communication (Labov 2001; Paul 1909). More specifically, we address two gaps in this literature: the scarcity of documented cases of nongradual innovation in core grammar, and the need to identify and specify the converging mechanisms that enable such innovation.

To this end, we triangulate corpus and survey data to examine the community-wide usage and speakers' mental representation of the construction, providing complementary evidence for radical and rapid core-grammatical innovation in contemporary Chinese. We argue that its development was enabled by three mutually reinforcing factors: cognitive biases, joint affect, and platform amplification. By bringing digitally mediated, high-affect innovation in core grammar into view, the study makes both theoretical and empirical contributions to Diachronic Construction Grammar and Creative Construction Grammar.

2 Corpus data and methods

2.1 Corpus source and retrieval

To analyze the APC, we retrieved data from the mainland Chinese domain (ca. 6.8 billion words) of the Simplified Chinese Web Corpus zhTenTen2017 in Sketch Engine.⁴ We used the following CQL string in advanced concordance search: [word = “”][word = “被”][word = “”]. The use of quotation marks served as a heuristic for capturing noncanonical and perceptibly marked uses of the [bèi X] pattern in online discourse. This choice was motivated both empirically and technically. A preliminary inspection of online examples suggested that innovative [bèi X] forms were typically enclosed in quotation marks as signals of markedness. In addition, because [bèi X] shares its surface format with canonical passive expressions, a broad search without quotation marks returned more than 7,600,000 concordances, far beyond a manageable range for reliable retrieval and annotation.

⁴ The corpus is a POS tagged, overwhelmingly topic- and genre-unspecified “all-purpose Chinese corpus covering the largest possible variety of genres, topics, text types and web sources” in simplified Chinese characters, comprising ca. 13 billion words. Details are available on the Sketch Engine website. Although a newer zhTenTen21 edition had become available by the time this study was completed, we did not adopt it because the current study already combines zhTenTen17 for corpus analysis with survey evidence on the construction's current status in native-speaker knowledge.

Quotation-marked search therefore provided the most effective available means of isolating innovative uses of the pattern. As the APC became conventionalized, such overt marking likely decreased; the present dataset should therefore be understood as preferentially capturing the construction's innovation phase rather than the full range of later usage. This limitation is addressed in part by the survey component, which probes current speaker knowledge of APC (see Section 3).

The query returned 17,276 concordances comprising 3,655 distinct types, all downloaded as raw data. Two researchers independently annotated the 3,655 types, coding each instance as either 1 = valid APC or 0 = false positive. The primary annotation criterion was the transitivity, and thus passivizability, of X. Inter-coder agreement was substantial (Cohen's $\kappa = 0.71$, $p < 0.001$). Discrepant cases were then reviewed in concordance context, with attention to pragmatic indicators such as sarcasm and with reference to the *Modern Chinese Dictionary* (2016) and resolved through discussion. This process eliminated 2,114 false positives, leaving 1,541 valid APC types totaling 9,399 tokens.

2.2 Metadata on attestation dates

To estimate the pace of APC constructionalization, we selected the 100 most frequent types as a representative sample and established their earliest attestation dates. Given the construction's digital origin, we adopted a Web-as-corpus approach. Using the simplified Chinese platform of Google News, which pools content from Chinese-language news sources, we searched each expression and recorded the month and year of its earliest publicly documented attestation (see Supplementary Material I). Where no relevant news result could be located, we consulted the Sketch Engine source links for the concordances and used the earliest available timestamps as a proxy.

2.3 Analytical methods

2.3.1 Frequency, entrenchment, and productivity

To characterize the frequency and usage profile of the APC, we examined its type frequency, token frequency, and hapax frequency. Type frequency indexes the breadth of lexical expansion, token frequency reflects recurrence and entrenchment, and hapax frequency provides an indicator of productive extension (Baayen 1993; Barðdal 2006; Bybee 2010; De Smet 2020; Säily et al. 2018).

2.3.2 Syntactic radicalization

To assess the APC's syntactic radicalization and constructional recategorization, we analyzed the lexical-class distribution of X-slot fillers. Two researchers independently coded their parts of speech according to their conventional meaning and function outside the APC, combining dictionary meaning with dominant semantic function (Croft 2001). Referential forms were coded as nouns, predicating forms as verbs – transitive if they license an external object, otherwise intransitive – and predicative forms denoting stable properties as adjectives.⁵ Intercoder reliability was high (Cohen's $\kappa = 0.94$) and disagreements were resolved through discussion.

2.3.3 Cumulative innovation trajectory of core types

Using first-attestation metadata for the 100 most frequent APC types, we plotted cumulative type counts at half-year intervals. Because exhaustive first-attestation dating is not feasible for the full APC inventory, this analysis should be understood as a conservative reconstruction of the emergence of the construction's high-frequency core.

2.3.4 Semantic vector analysis

To explore the semantic dimensions of the APC, we used semantic vector analysis grounded in distributional semantics, which measures semantic similarity between words based on their collocational patterns. To examine similarities and differences among the fillers of the X slot, we adopted the Word2Vec approach, a widely used word embedding technique in natural language processing (NLP), in which semantically related words occupy similar positions in high-dimensional vector space. Vector representations for the most frequent 100 APC types were obtained from the Tencent AI Lab Embedding Corpus (Song et al. 2018), a large pretrained Chinese word embedding model. Trained with an enhanced Directional Skip-Gram approach, this model represents semantic information in a 100-dimensional space based on over eight million Chinese words and phrases (Mikolov et al. 2013). After retrieving the 100-word vectors, we normalized each vector to unit length so that Euclidean distance would behave equivalently to cosine similarity, the standard measure of semantic proximity in NLP.

5 Some Chinese lexemes may occur in more than one syntactically defined class, with one use being more conventional than others (Guo 2004). We code *bèi xìngfú* according to the adjectival/stative sense with which this APC use was initially associated, as in *Nǐ xìngfú ma?* 'Are you happy?'

We then applied k-means clustering to group words with similar vector representations, and thus similar meanings (Moisl 2015). To determine the optimal number of clusters, we used the elbow method, testing values of k from 1 to 10 and identifying the point at which additional clusters yielded diminishing reductions in within-cluster variance (Syakur et al. 2018). This procedure identified $k = 6$ as the most appropriate solution.

To visualize the clustering results, we performed Principal Component Analysis to reduce the 100-dimensional vectors to two dimensions while retaining as much variance as possible (Adnan et al. 2021). The resulting coordinates were plotted with each word shown as a labeled point. Words belonging to the same cluster share a color, and each cluster is enclosed by an ellipse with a descriptive label based on its dominant lexical profile.

All corpus analyses and visualizations were performed in R using the following packages: *tidyverse*, *ggplot2*, *ggforce*, and *showtext*.

3 Survey data and methods

To address the limits of the corpus data, especially the lack of post-2017 data and the resulting uncertainty about the current status of APC, we conducted an anonymous online survey to assess the current representation of APC in native-speaker knowledge. Supplementary Material III provides the survey instrument in the original Chinese, together with a brief English description.

3.1 Survey data

In addition to demographic information including age, gender, education level, platforms used, and daily online exposure, the survey comprised four tasks: (1) metalinguistic judgment, (2) exemplar interpretation, (3) perception of productivity, and (4) a verbal-protocol task eliciting additional examples. In the metalinguistic judgment task, participants evaluated four [*bèi* X] expressions *tā bèi jiùyè le* ‘He is job-ed’, *tā bèi zìyuàn le* ‘He is voluntary-ed’, *tā bèi xìngfú le* ‘He is happy-ed’, and *tā bèi jiānqiáng le* ‘He is strong-ed’, selected to differ sharply in token frequency (444, 215, 95, and 3, respectively). One canonical passive item, *tā bèi dǎshāng le* ‘He is beaten up’, was included as a quality-control item. Item order was randomized. For each item, participants rated three dimensions: (i) familiarity, (ii) confidence in meaning, and (iii) likelihood of self-use in everyday contexts. In the interpretation task, participants selected appropriate meanings for *wǒmén bèi jiùyè le* ‘We were job-ed’ and *wǒ bèi xiūjià le* ‘I was vocation-ed,’ two items selected to differ in token frequency

(444 vs. 22), from a list containing construction-consistent interpretations and a distractor option. The perceived-productivity task asked participants to judge, on a four-point scale from Absolutely Impossible to Absolutely Possible, whether new [*bèi* X] expressions could be created as needed. Finally, participants were invited to provide additional [*bèi* X] usages they knew.

The survey was administered through Qualtrics and targeted adult native speakers of Chinese. Recruitment was conducted via faculty contacts at five universities in northern, southern, northeastern, southeastern, southwestern China, as well as through the researcher's personal networks. The semi-controlled sample was not population-representative and is used here to assess current native-speaker awareness, interpretation, and perceived productivity of the construction, not to support population-level claims. The survey research was part of a larger Chinese linguistics project, which was reviewed by the University of Oregon Research Compliance Services (RCS) under the 2018 Common Rule and determined to qualify for exemption under Title 45 CFR 46.104(d)(2)(i) Tests, surveys, interviews, or observation (non-identifiable)).

A total of 449 responses were received, of which 83 unfinished responses were excluded; after quality screening, the final sample comprised 311 valid responses. Participants were distributed across four age groups: 18–25 ($n = 161$), 26–35 ($n = 58$), 36–45 ($n = 58$), and 46+ ($n = 34$). The gender distribution was female ($n = 208$), male ($n = 98$), and prefer not to say ($n = 5$). Daily internet exposure was under 1 h ($n = 10$), 1–3 h ($n = 84$), 3–6 h ($n = 107$), and 6+ h ($n = 110$). Education level was heavily skewed toward college and postgraduate participants (college/university $n = 129$; graduate and above $n = 159$; high school $n = 5$; middle school or below $n = 5$) and was therefore not included in the present analysis.

3.2 Survey analytical methods

For statistical analysis, participants' metalinguistic judgments – familiarity, confidence in meaning, and likelihood of use – were treated as ordinal dependent variables. Ordered logistic regression models were fitted to examine the effects of age group, gender, and daily online exposure on these three judgment dimensions. For interpretation, participants' responses were coded binarily as Correct if at least one target-consistent meaning was selected and Incorrect if none was selected. Fisher's exact test was then used to examine age-group differences in interpretation accuracy. Finally, perceived productivity was operationalized as participants' judgment of whether new [*bèi* X] expressions could be created as needed. Age-group differences were evaluated using the Kruskal–Wallis test. Survey data analyses and

visualizations were performed in R using the following packages: *readxl*, *tidyverse*, *ggplot2*, *ggforce*, *showtext*, *patchwork*, *cowplot*, *emmeans*.

4 Results

4.1 Frequency and productivity

Table 1 presents the frequency profile of the APC as captured in the corpus data. The data show 1,541 distinct types, 714 hapax legomena, and 9,399 tokens (1.4 per million words in the mainland Chinese domain of zhTenTen17). The 10 most frequent types each occur between 147 and 444 times, making up 26 % of all tokens (see Supplementary Material I). Together, these measures reveal a distribution characterized by a small set of highly frequent exemplars with a long tail of low-frequency types.

4.2 Syntactic radicalization

The X slot extends well beyond the canonically required transitive verb, accommodating fillers from several lexical categories. Table 2 shows that intransitives and nouns predominate, together accounting for nearly 90 % of the inventory, followed by adjectives. Transitive verbs used sarcastically in the APC are less than 3 %.

Table 1: Frequency profile of the APC.

Construction	Type frequency	Token frequency	Hapax legomena
[<i>bèi</i> X]	1,541	9,399	714

Table 2: Distribution of lexical classes among the fillers of X slot.

Lexical category	Type count	Percentage
Intransitive verbs	722	46.85 %
Nouns	649	42.12 %
Adjectives	128	8.31 %
Transitive verbs	38	2.47 %
Numerals	4	0.26 %
Total	1,541	100.00 %

4.3 Timeline of type expansion

The 100 most frequent APC types were attested from 2008 to 2016. Their earliest located attestation dates were aggregated by half-year intervals and plotted cumulatively in Figure 1.

Figure 1 provides a conservative view of the APC’s early innovation trajectory by tracing the accumulation of its 100 most frequent types. It reveals a concentrated burst of rapid type expansion between 2009 and 2011. Only five of the 100 most frequent APC types were attested by 2009 H1, but the cumulative total rose sharply to 33 by 2009 H2, 60 by 2010 H1, 75 by 2010 H2, and 84 by 2011 H1. Growth then slowed markedly, with the inventory increasing only gradually from 84 types in 2011 H1 to the full set of 100 by 2016 H2. This pattern points to an explosive early phase of constructional expansion in the APC’s high-frequency core, followed by slower growth.

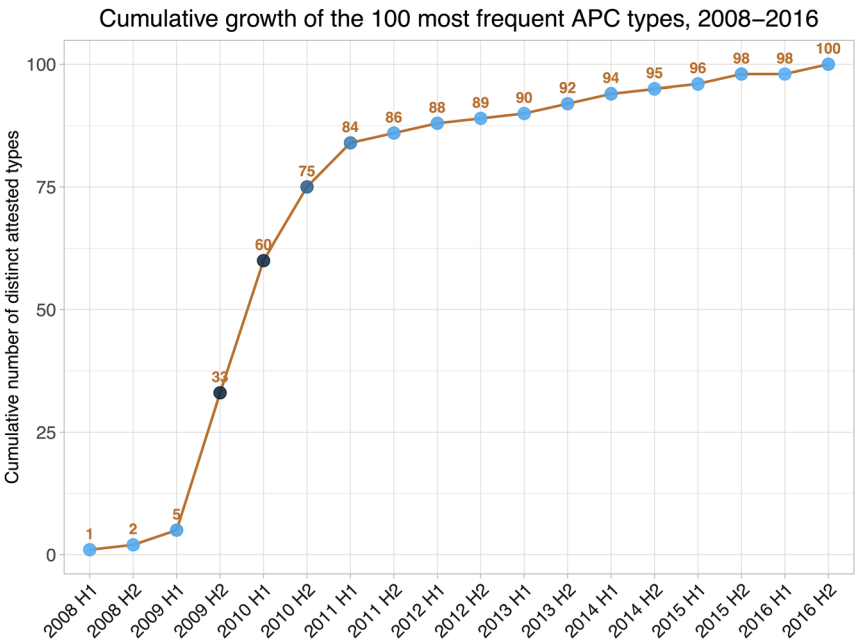


Figure 1: Cumulative growth of the 100 most frequent APC types from 2008 to 2016, based on earliest located attestations and plotted at half-year intervals. The curve shows a sharp early burst of type expansion between 2009 H2 and 2011 H1, followed by slower incremental growth. As the figure tracks only the most frequent types, it represents a conservative visualization of the construction’s overall expansion.

4.4 Semantic coverage of APC

Figure 2 visualizes the vector-space distribution of the items occupying the X slot in the 100 top [bèi X] types projected onto the first two principal components (PC1 and PC2). Under the $k = 6$ k-means solution, the types fall into six semantic groupings, labeled as Education/advancement, Social positioning, Biographical crisis, Socio-economic administration, Benchmarking, and Policy regulations. These labels are heuristic abstractions based on the lexical makeup of each cluster and are intended to aid interpretation of the vector-space distribution and do not imply categorically discrete semantic classes. The six cluster labels, their descriptive definitions, and illustrative examples are provided in Table 3; Supplementary Material II lists all items in the clusters with pinyin and gloss.

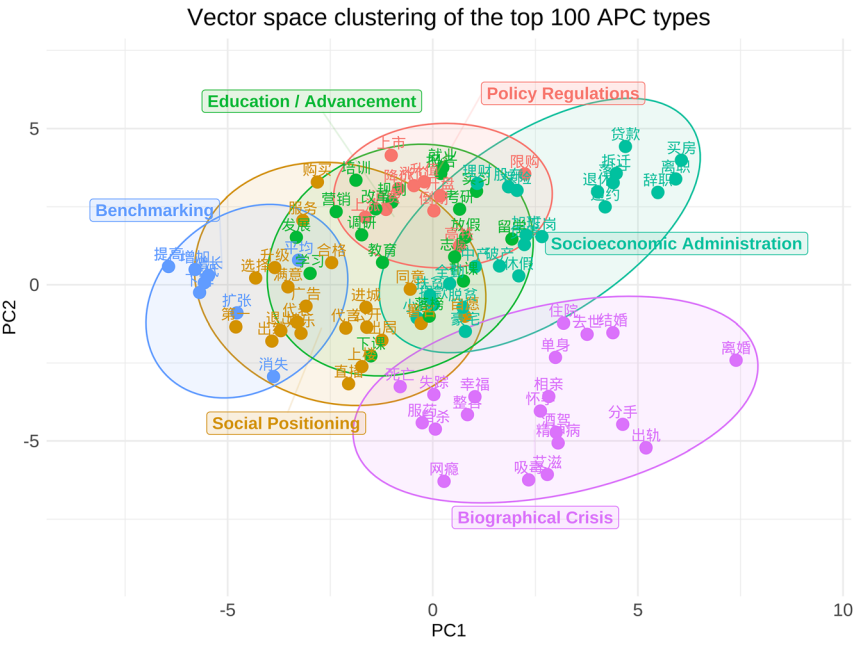


Figure 2: Semantic clustering of the top 100 APC types in vector space. Chinese labels identify individual types; colored ellipses mark broad semantic clusters. PC1 and PC2 represent the first two principal components of the semantic vector space. English cluster labels are provided to guide interpretation for readers not familiar with Chinese.

Table 3: Descriptive summary of the six semantic clusters with examples.

Cluster label	Brief description	Examples of X
Education/advancement	Education, training, qualification, and institutional mobility	就业 <i>jiùyè</i> ‘enter the workforce’
Social positioning	Representation, evaluation, publicity, and social position	满意 <i>mǎnyì</i> ‘be satisfied’
Biographical crisis	Life events, bodily states, stigma, and personal affectedness	离婚 <i>lǐhūn</i> – ‘divorce’
Socioeconomic administration	Administration of finance, housing, labor, welfare, and livelihood	贷款 <i>dàikuǎn</i> ‘take a loan’
Benchmarking	Quantification, aggregate comparison, and statistic	平均 <i>píngjūn</i> ‘average’
Policy regulations	Regulation, restriction, and policy-governed administration	高铁 <i>gāotiě</i> ‘high-speed rail’

5 Survey results

The overall survey results show a clear asymmetry between age-related variation in participants’ metalinguistic judgments of the four APC expressions and their uniformly demonstrated ability to interpret the two test items accurately, recognize the productivity of the schematic [*bèi* X] pattern, and produce both corpus-attested and novel instances.

5.1 Metalinguistic judgement

Across the three metalinguistic judgment measures – familiarity, confidence in meaning, and likelihood of self-use – the most consistent predictor was age cohort. Gender and daily internet use did not significantly predict responses across the three tasks (all *p* > 0.05). The age effect, however, was not uniform across measures: it was clearest for likelihood of use, robust for familiarity, and weaker and less consistent for confidence in meaning.

For familiarity, the proportional odds assumption was satisfied (Brant omnibus $\chi^2(21) = 22.96$, *p* = 0.35). The 18–25 group showed the lowest predicted familiarity (EMM = –2.61, 95 % CI [–3.04, –2.17]), followed by the 26–35 group (EMM = –1.62, 95 % CI [–2.18, –1.06]), the 46+ group (EMM = –1.37, 95 % CI [–2.05, –0.70]), and the 36–45 group (EMM = –1.26, 95 % CI [–1.78, –0.73]). Bonferroni-corrected pairwise comparisons showed that the 18–25 group was significantly less familiar than all three

older groups: 26–35 ($b = -0.99$, $z = -3.26$, $p = 0.0066$), 36–45 ($b = -1.35$, $z = -4.55$, $p < 0.0001$), and 46+ ($b = -1.23$, $z = -3.37$, $p = 0.0044$). No significant differences were found among the three older groups. Descriptively, high-frequency types were rated as more familiar than low-frequency types. Figure 3 shows the age-group distributions of self-reported familiarity with the four APC expressions with the left panel showing the pooled results across the four items.

For confidence in meaning, the age effect was weaker and less regular than for familiarity. The estimated marginal means show a non-monotonic pattern, peaking in the 26–35 group and declining on either side. At the same time, the proportional-odds assumption was violated for this model (Brant omnibus $\chi^2(21) = 93.4$, $p < 0.001$), so the ordinal-logit results should be interpreted with caution. Estimated marginal means again placed the 18–25 group lowest (EMM = -1.45 , 95 % CI [-2.05 , -0.86]), while the 26–35 group showed the highest predicted confidence (EMM = -0.43 , 95 % CI [-1.10 , 0.24]), followed by 36–45 (-0.72 , 95 % CI [-1.41 , -0.03]) and 46+ (-0.85 , 95 % CI [-1.56 , -0.14]). After Bonferroni correction, only one age-group contrast reached significance: 18–25 versus 26–35 ($b = -1.02$, $z = -3.56$, $p = 0.002$). No other pairwise differences were significant. Descriptively, high-frequency types also received higher confidence-in-meaning ratings than low-frequency types. Figure 4 shows the age-group distributions of self-reported confidence in meaning for the four APC expressions with the left panel showing the pooled results across the four items.

For self-reported likelihood of use, age differences were strongest. Estimated marginal means showed the strongest age differentiation. The 18–25 group again had the lowest predicted value (EMM = -2.51 , 95 % CI [-3.07 , -1.95]), followed by the

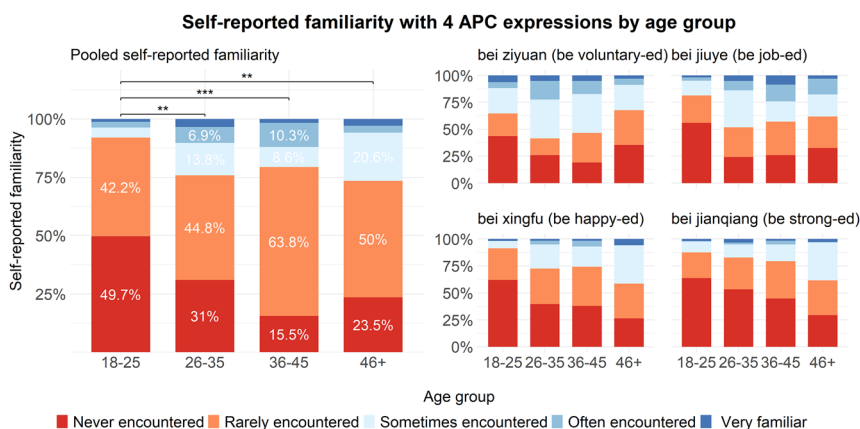


Figure 3: Age group distributions of self-reported familiarity with the four target [bèi X] expressions from Never Encountered to Very Familiar. The left panel shows the pooled response distribution across items; the four panels on the right show item-specific distributions.

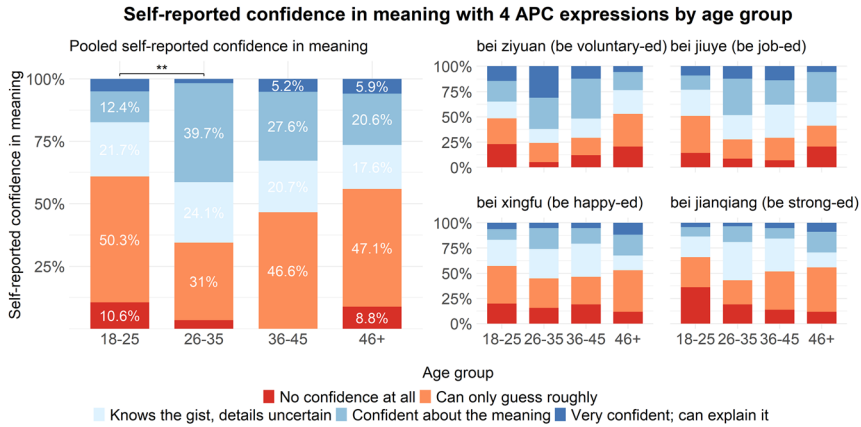


Figure 4: Age-group distributions of self-reported confidence in meaning comprehension for the four target [*bèi* X] expressions from No confidence at all to Very confident, can explain it. The left panel shows the pooled response distribution across items; the four panels on the right show item-specific distributions.

26–35 group (EMM = -1.20 , 95 % CI [-1.84 , -0.56]), the 36–45 group (EMM = -0.86 , 95 % CI [-1.47 , -0.25]), and the 46+ group (EMM = -0.10 , 95 % CI [-0.73 , 0.54]). Bonferroni-corrected pairwise comparisons showed that the 18–25 group was significantly less likely than all three older groups to report that they would use the target expressions: 26–35 ($b = -1.31$, $z = -4.14$, $p = 0.0002$), 36–45 ($b = -1.65$, $z = -5.13$, $p < 0.0001$), and 46+ ($b = -2.41$, $z = -5.83$, $p < 0.0001$). No significant differences were found among the three older groups. Descriptively, high-frequency types were rated as more usable than low-frequency types. Figure 5 shows the distributions of self-reported use tendency by age group for the four APC expressions with the left panel showing the pooled distributions across the four items.

Taken together, the three judgment measures show partially convergent age effects. Familiarity and likelihood of use both display a robust cohort contrast in which the 18–25 group patterns significantly below all older groups. By contrast, confidence in meaning shows a weaker and less uniform age effect, with only the 18–25 versus 26–35 comparison reaching significance.

5.2 Accuracy of interpretation

Across all age groups, over 90 % of participants correctly interpreted both test items, as can be seen in Figure 6. Age group differences were not significant for *bèi jiùyè* (Fisher's exact test, $p = 0.178$). Although an overall Fisher's exact test reached

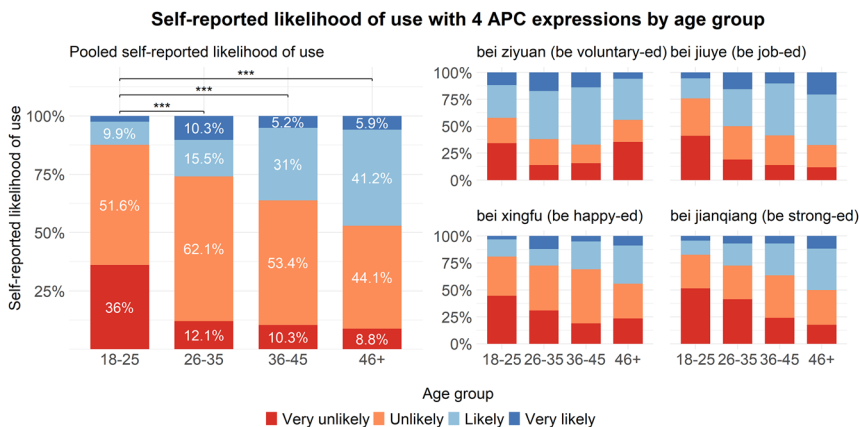


Figure 5: Age-group distributions of self-reported likelihood of using the four target [*bei* X] expressions from Very unlikely to Very likely. The left panel shows the pooled response distribution across items; the four panels on the right show item-specific distributions.

significance for *bèi xiūjià* ($p = 0.015$), post-hoc pairwise comparisons did not identify any significant group difference.

5.3 Perceived productivity

Responses to the question whether new [*bèi* X] expressions can be created as needed were predominantly positive across all age groups. As Figure 7 shows, the combined proportion selecting Mostly possible and Absolutely possible was 73.9 % among the 18–25 group, 77.6 % among the 26–35 group, 82.8 % among the 36–45 group, and 70.6 % among the 46+ group. A Kruskal–Wallis test indicated that the overall age-group difference did not reach significance, $H(3) = 7.11$, $p = 0.069$, $\varepsilon^2 = 0.023$.

5.4 Spontaneous elicitation

Of the 311 survey participants, 198 participants (64 %) supplied at least one valid APC example, whereas 113 (36 %) either provided no example or provided non-APC examples. A chi square test found no statistically significant association between age group and valid APC elicitation $\chi^2(3, N = 311) = 3.99$, $p = 0.263$. The task yielded 240 APC tokens representing 102 distinct types. The four most frequently elicited examples – *bèi dàibiǎo* ‘be involuntarily represented’ (17), *bèi bìyè* ‘be registered as having graduated’ (16), *bèi jiéhūn* ‘be allegedly/involuntarily married’ (16), and *bèi zìyuàn* ‘be

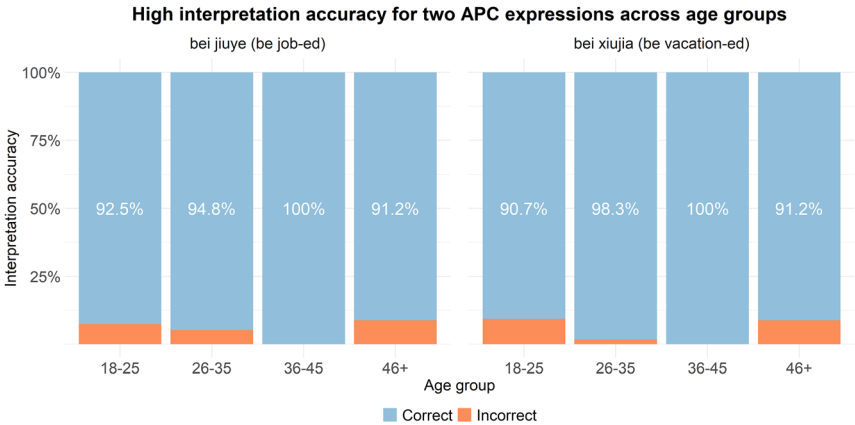


Figure 6: Stacked bars show distributions of correct and incorrect interpretations for *bèi jiùyè* ‘be job-ed’ and *bèi xiūjià* ‘be vacation-ed’ across four age groups. Interpretation accuracy exceeds 90 % for both APC expressions in every age group.

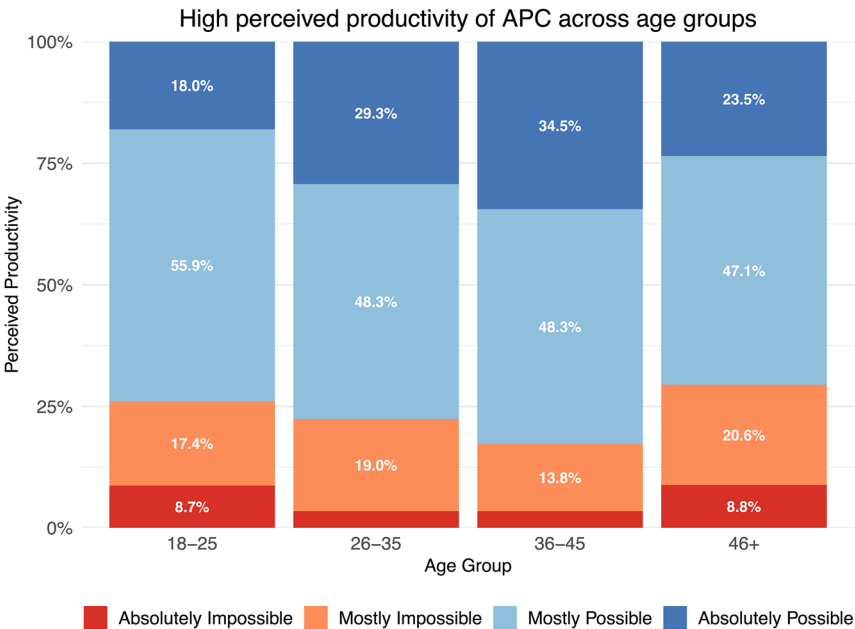


Figure 7: Stacked bars show the distribution of responses to whether new *[bèi X]* expressions can be created as needed, from Absolutely Impossible to Absolutely Possible. Perceived productivity of the *[bèi X]* pattern is consistently high across all age groups.

voluntary-ed' (14) – accounted for 26 % of all tokens. The ten most frequent types accounted for 42 %. Overall, 23 types were cited by at least three participants, 38 by at least two, 64 only once.

Of the 102 elicited APC types, 77 are attested in the zhTenTen2017 corpus, whereas 25 are unattested there. Illustrative examples of these include *bèi gāo shōurù* 'be high-income-ed' and *bèi xiǎosān* 'be mistress-ed', showing that speakers can readily extend the pattern beyond corpus-attested exemplars. Comparison with zhTenTen2017 corpus frequencies shows that elicitation frequency was positively, though moderately, associated with corpus token frequency (Spearman's $\rho = 0.36$, $p = 0.001$). Exceptions appear to reflect sample relevance: college-related *bèi biyè* 'be registered as having graduated' was frequently elicited despite low corpus frequency, whereas *bèi tuōpín* 'be classified as having escaped poverty' showed the reverse pattern.

6 Discussion

6.1 Interpreting the corpus findings

Within the attested ten-year window from the earliest known APC exemplar in 2008 to the endpoint of the zhTenTen17 data in 2017, the corpus findings provide strong evidence that the APC is productive. Its frequency profile combines high token frequency and high type frequency with a long tail of hapaxes, indicating recurrence, entrenchment, and schematic productivity (Baayen 1993; Barðdal 2006; De Smet 2020; Säily et al. 2018).

The pattern also exhibits substantial syntactic radicalization, freely admitting major non-passivizable categories – not only intransitive verbs but also nouns and adjectives – to construe the subject as noncanonically subjected to an externally imposed narrative, event, process, state, role, or outcome.

The cumulative trajectory of its 100 most frequent types shows that the construction's high-frequency core emerged in a sharply concentrated early burst from 2009 to 2011. This pattern is broadly consistent with Yuan and Zhang's (2026) type-frequency trajectory, which shows an early rise from 2008, a peak around 2010, and later stabilization. Although their dataset is not directly comparable to ours, this convergence supports our conservative finding that the APC underwent a sharp early surge in types before entering a more stable phase.⁶

⁶ Yuan and Zhang (2026) also report type-token ratio and hapax proportion, but we consider type frequency the more reliable indicator in their small self-compiled corpus, since ratio-based measures are more unstable and therefore less reliable under low token counts (cf. Säily et al. 2018).

The semantic vector analysis further shows that the APC's high-frequency types span a broad conceptual space. Five of the six groupings overlap densely, whereas Biographical Crisis occupies a relatively distinct region. The overlapping clusters are better understood as related subregions within a broader semantic field, many of them involving macro-level institutional imposition, administrative categorization, and top-down narratives to varying degrees. By contrast, Biographical Crisis centers on personal life events, bodily states, and stigmatized conditions. Its relative semantic distinctiveness suggests that the speakers also recruit the APC to construe private experiences as forms of externally imposed identity, diminished agency, and loss of control – unwanted situations thrust upon them. Overall, the APC's semantic coverage exemplifies what Goldberg (2019: 16) calls a “hyper-dimensional conceptual space”: a construction that accommodates semantically diverse but conceptually interrelated experiences and integrates them through a shared construal. As a new node in the passive constructional network, the APC has become an expressive resource for both social critique and the articulation of personal disempowerment.

Read together, these patterns point to the APC's radical and rapid constructionalization, whose scale, speed, and categorial expansion are not readily explained as incremental change.

6.2 Interpreting the survey findings

The survey findings provide independent evidence of the APC's continuing vitality, complementing the corpus findings with data on speaker knowledge, perception, and production. Across tasks, the construction emerges as accessible, interpretable, and productively available, although access to particular exemplars varies across age cohorts.

The clearest cohort effects appear in the metalinguistic judgments. Participants aged 18–25 reported lower familiarity with the target expressions, lower confidence in explaining their meanings, and lower likelihood of use than older groups. A plausible explanation is differential historical exposure. When the earliest APC expression emerged in 2008 and the construction rose to prominence in the late 2000s and early 2010s, members of the youngest cohort had either not yet been born or were still too young to encounter these expressions at the height of their public salience. Their lower ratings thus likely reflect weaker familiarity with historically salient exemplars rather than weaker access to the constructional pattern itself. This account may also explain why the 26–35 group showed the highest confidence in meaning: they were both old enough to have encountered the expressions in their period of greatest visibility and young enough to have grown up with their exemplar-based diffusion.

In contrast, interpretation accuracy was uniformly high across age groups, and participants overwhelmingly judged the [bèi X] pattern to be productive. These findings suggest a distinction between metalinguistic familiarity with particular lexicalized exemplars and schematic constructional knowledge (Goldberg 2019; Traugott and Trousdale 2013).

The elicitation results reinforce this interpretation. They reveal both a shared core of highly accessible exemplars and a long tail of less frequently retrieved types. The positive association between corpus and elicitation frequency suggests that community-wide conventionalization is generally reflected in the entrenchment of APC exemplars in individual speakers (Schmid 2015). The elicitation of corpus-unattested types further shows that the [bèi X] pattern is readily extendible beyond corpus exemplars, consistent with participants' productivity judgments.

Overall, the survey results reveal both age-related variation and broad shared knowledge of the APC. While metalinguistic knowledge of particular exemplars varies by cohort, interpretation, productivity judgments, and elicitation all indicate that the APC remains accessible, interpretable, and productive in contemporary Chinese.

Taken together, the corpus and survey evidence points to the APC's rapid constructionalization and community-wide conventionalization, as well as its entrenchment and accessibility in the mind of speakers – interrelated processes central to Schmid's (2015) Entrenchment-and-Conventionalization Model. Yet, the EC-Model typically foregrounds gradual individual and communal adaptations underlying constructionalization (Schmid 2020). The APC's exceptional rate of innovation thus raises a key theoretical question: What drove the radical and rapid constructionalization of APC? As Traugott (2018: 21) observes, "change involves production as well as perception during acquisition throughout life," highlighting the dynamic interplay between innovative speaker strategies and receptive listener uptake. In what follows, we specify three mutually reinforcing mechanisms underlying the APC's radical and rapid constructionalization: cognitive biases, joint affect, and platform amplification.

6.3 Mechanisms of radical and rapid constructionalization of APC

6.3.1 Grammatical surprise

That there are strong cognitive, communicative, and social incentives for speakers to adhere to the conventions of their native language is well established (Bloomfield 1933; Croft 2000; Goldberg 2019; Labov 2012). If so, what leads them to produce

expressions that so blatantly violate grammatical rules? One factor, we argue, is the attentional effect of GRAMMATICAL SURPRISE, a concept proposed by Jing-Schmidt (2024) to capture the surprise induced when jarring grammatical violations generate prediction error.

Croft (2000:115) argues that communication serves social goals of the interlocutors whose language use responds to the novelty of communicative situations, and that “sometimes achieving those goals, or even achieving just the goal of communication, is better done by stretching or breaking linguistic conventions.” We argue that the initial impetus for APC arises directly from speakers’ need to articulate extraordinary social events using extraordinary linguistic means. In other words, the illicit and surprising linguistic form mirrors the perceived absurdity of reality.

This communicative strategy may recall what Haspelmath (1999: 1057) calls “extravagance,” the speaker’s “imaginative and vivid” performance of verbal novelty to attract attention. In the APC, that extravagance resides precisely in the striking effect of syntactic anomaly, which generates what Schmid and Günther (2016) call “salience by surprisal.” Grammatical surprise elaborates the notion of extravagance by explicitly linking grammatical deviance to the broader psychology of surprise: deviant forms violate speakers’ cognitive models of grammar, generating prediction error and, consequently, surprise. For speakers, it not only heightens expressivity and stance, turning deviant form into discursive force, but also enables a new way of saying *new* things, rather than merely “a new way of saying old things” in Haspelmath’s account of extravagance. For listeners, the unexpected form triggers prediction error, and cognitive research on surprise associates such prediction error with heightened attention, enhanced memory, and better learning (Kagan 2002; Reisenzein et al. 2019). As Goldberg (2019: 56) notes, “surprising information creates a stronger error signal, which results in greater weight changes and more learning.” In digital discourse, these effects make the APC expressions more likely to be noticed, shared, and remembered. By placing non-passivizable elements in the X slot, the APC creates surprise and sensation through direct violation of passive syntax, making core grammar itself the locus of innovation.

6.3.2 Negativity bias

Another factor is negativity bias. The APC is not only syntactically jarring but also semantically negative, centering on death, coercion, or falsehood. These themes activate what cognitive psychology identifies as negativity bias – the tendency for negative stimuli to command greater attention, elicit stronger emotional response,

and leave deeper memory traces (Baumeister et al. 2001; Rozin and Royzman 2001).⁷ We argue that this cognitive-affective intensification works in tandem with grammatical surprise, helping to explain why the early coinages gained immediate traction despite their syntactic deviance. The threat they encoded amplifies the salience of the grammatical anomaly, making the expressions memorable, shareable, and viral.

Previous research (Peng 2017; Shen 2010; Shi 2013) has argued that the APC's negativity is inherited from the adversative semantics of the CPC, as in *tā bèi shā le* 'he was killed.' We argue that the negativity driving the radical and rapid constructionalization of APC differs from its canonical counterpart not just in degree, but in kind. While canonical *bèi shā* 'be killed' denotes a severely negative event, *bèi zishā* 'be suicide-ed' adds further layers: narrative dispossession, systemic injustice, obscured truth, and feigned agency, whereby an ostensibly voluntary act is exposed as coerced or staged.

The APC thus does more than depict adversity; it satirically exposes injustice and distortion, inviting readers to question dominant narratives and authority structures. Its layered negativity lies not only in what happened, but in how it is represented. In this way, affect, irony, and critique become central to APC's memorability and transmissibility in digital discourse.

6.3.3 Joint affect

Grammatical surprise and affective intensity may capture attention and strengthen memory, but a construction cannot spread unless it is also intelligible. How, then, do listeners make sense of an unprecedented ungrammatical expression?

Following Croft's (2000: §4.3) model of non-conventional coordination devices, we suggest that listeners rely on joint salience – the shared accessibility of contextual cues and encyclopedic knowledge – to construct meaning. In *bèi zishā* 'be suicide-ed', for example, the viral visibility of a suspicious death clashes with the shared understanding that suicide is, by definition, self-initiated and voluntary. This incongruity makes the syntactic anomaly jointly salient, enabling audiences to infer coercion, cover-up, or institutional manipulation.

This inferential process is affectively charged. The grammatical anomaly attracts attention while fostering emotional alignment, especially in contexts of moral outrage, fear, or shared cynicism. We therefore extend Croft's notion of joint salience into "joint affect," understood as emotional resonance that enables interpretive uptake by inviting the listener into a shared stance of indignation.

⁷ For research on the linguistic pragmatics of negativity bias, see Jing-Schmidt (2007) and Foolen (2015).

Grice's (1989) principle of cooperative conversation undergirds this inferential process. When speakers produce a syntactically jarring expression that appears to violate the Maxim of Manner, listeners nevertheless assume that the utterance is relevant and seek an interpretation consistent with the Maxim of Relation. In digital discourse, the resulting implicature is affective as well as informational: the APC conveys both what happened and how it is felt.

6.3.4 Platform amplification

Cognitive biases and joint affect explain why the APC captures attention and invites alignment, but not the speed and scale of its diffusion. Its spread was amplified by the affordances and incentives of digital platforms through which these expressions were created, circulated, and transformed. Explaining the APC's productivity thus requires attention to the sociotechnical infrastructure that gave this bottom-up innovation visibility and traction in real time.

The emergence of APC coincided with the rise of Web 2.0 platforms in China, including Sina Weibo, Tencent QQ, and later WeChat. These platforms created a new sociotechnological ecology for language use, in which users could coin novel expressions and disseminate them instantaneously to large and diverse audiences. Their affordances – speed, amplification, and social indexing – served as accelerants of constructional spread.

Digital interaction further amplified the virality of linguistic innovations marked by shock value, emotional charge, and critical stance. Users drawn to the APC's striking form and satirical force shared, reshared, and commented on these expressions, thereby catalyzing further innovation. This pattern of networked participation supports the emergence of a sociolinguistic “small world”: although the digital public is vast, users are densely connected through reposts, likes, and shared discourse (Fagyal et al. 2010). As Ke et al. (2008) show, linguistic innovations diffuse especially efficiently in regular and small-world networks, where highly connected agents act as linguistic “hubs” that greatly increase the visibility and uptake of novel forms. In this sense, APC diffusion was not purely usage-based in the traditional sense, but also network-enabled.

At the same time, in the digital attention economy, visibility is currency. For influencers, media outlets, and everyday content creators alike, standing out in an oversaturated information landscape affects engagement, sponsorship, and monetized exposure. In this environment, linguistic novelty becomes economically valuable, as research on clickbait has shown (e.g., Lang and Jing-Schmidt 2024; Scott 2021). With its syntactic incongruity and emotional intensity, the APC works as a form of grammatical clickbait: its shock value invites click, comments, and reposts, which

in turn fuel algorithmic promotion and expand visibility. The affordances and incentives of platform capitalism thus acted as amplifiers for the APC's rise and spread.

Schmid (2016: 547) urges cognitive linguists to “spell out, rather than just acknowledge, the social and pragmatic dimension of language.” The analysis presented here has done just that: it shows how cognitive, affective, pragmatic, and sociotechnological forces go hand in hand to propel the APC's dramatic rise and spread.

7 Conclusions

This study has provided converging corpus and survey evidence that the APC emerged through radical syntactic innovation and spread with striking rapidity, and that native speakers not only access and interpret its exemplars but also perceive the pattern itself as productive. The development of the APC exemplifies a pathway of constructionalization driven not by gradual change over micro steps, but by the interplay of cognitive biases, joint affect, and platform amplification, resulting in the rapid emergence of a radical grammatical node. The study contributes to current work in Diachronic Construction Grammar and Creative Construction Grammar by documenting radical innovation in core grammar and spelling out the converging mechanisms that enable it to emerge in real time.

The APC is not a historical anomaly. It aligns with known patterns of innovation in response to a changed sociocultural environment that opens up new expressive territory. Comparable dynamics can be seen in earlier core-grammatical shifts, such as the rise of the Mandarin *bǎ*-construction in its modern form during the Yuan dynasty (1,271–1,368), alongside the emergence of vernacular storytelling and theater, genres that favored vivid representation of dramatic events (Jing-Schmidt 2005; Jing-Schmidt and Tao 2009). In this sense, the case fits a general view of language as a complex dynamic and adaptive system, in which structures emerge from interaction and innovations spread through social selection and community uptake (Beckner et al. 2009; Croft 2000; Elman 1995).

Methodologically, the individual components of the study have limitations. The corpus analysis is constrained by quotation-marked retrieval and the partial reconstruction of construction's type growth. The survey reflects a predominantly college- and postgraduate-based participant profile. Future research could trace the spread of unmarked APC uses in newer corpora, reconstruct cumulative innovation on the basis of a fuller APC inventory, examine the relation between individual adoption and broader conventionalization, and test speaker knowledge in a more population-representative sample. These limitations notwithstanding, a major empirical strength of this study lies in the triangulation of corpus and survey evidence, which captures

both community-level diffusion and individual-level knowledge. Their convergence demonstrates the radical and rapid constructionalization of the APC in core grammar, expanding current models of constructional innovation beyond pathways of gradual change.

Acknowledgments: We thank the three anonymous reviewers and the editors for their insightful and constructive comments, which helped us strengthen this article in important ways. We are also grateful to Sketch Engine for granting access to the full set of retrieved corpus data beyond the standard 1,000-type download limit, and to all participants in the survey research for their time and contributions. An earlier version of this study was presented at the 45th Annual Conference of the German Linguistic Society (Deutsche Gesellschaft für Sprachwissenschaft, DGfS). We thank the audience for their valuable feedback. All remaining errors, of course, are our own.

Data availability: The corpus and survey data supporting the findings of this study are openly available on Open Science Framework (OSF) at <https://osf.io/qswg5/>.

References

- Adnan, Morshed J., Matthias L. Hemmje & Michael A. Kaufmann. 2021. Social media mining to study social user group by visualizing tweet clusters using Word2Vec, PCA and K-Means. In *Joint proceedings of the second workshop on bridging the gap between information science, information retrieval and data science, and third workshop on evaluation of personalisation in information retrieval co-located with 6th ACM SIGIR conference on human information interaction and retrieval (CHIIR 2021)*, 40–51. CEUR-WS. <https://ceur-ws.org/Vol-2863/> (accessed 10 July 2026).
- Baayen, R. Harald. 1993. On frequency, transparency and productivity. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 1992*, 181–208. Dordrecht: Kluwer.
- Barðdal, Jóhanna. 2006. Predicting the productivity of argument structure constructions *Proceedings of the annual meeting of the Berkeley Linguistics Society*, vol. 32, 467–478. Berkeley, CA: Berkeley Linguistics Society.
- Barðdal, Jóhanna & Spike Gildea. 2015. *Diachronic construction grammar*. Amsterdam: John Benjamins.
- Baumeister, Roy F., Ellen Bratslavsky, Catrin Finkenauer & Kathleen D. Vohs. 2001. Bad is stronger than good. *Review of General Psychology* 5(4). 323–370.
- Beckner, Clay, Richard Blythe, Joan Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman & Tom Schoenemann. 2009. Language is a complex adaptive system: Position paper. *Language Learning* 59. 1–26.
- Bloomfield, Leonard. 1933. *Language*. London: Henry Holt.
- Bohmann, Axel. 2016. Language change because Twitter? Factors motivating innovative uses of *because* across the English-speaking Twittersphere. In Lauren Squires (ed.), *English in computer-mediated communication: Variation, representation, and change*, 149–178. Berlin/Boston: De Gruyter.
- Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Chao, Yuen Ren. 1968. *Language and symbolic systems*. Cambridge: Cambridge University Press.

- Chen, Rong & Xiaoxia Hu. 2020. Be suicided: A construction grammar analysis of the innovative *bèi* construction in Chinese. *Cognitive Semantics* 6(1). 83–106.
- Comrie, Bernard. 1977. In defense of spontaneous demotion: The impersonal passive. In Peter Cole & Jerrold M. Sadock (eds.), *Grammatical relations (Syntax and semantics 8)*, 47–58. New York: Academic Press.
- Croft, William. 2000. *Explaining language change: An evolutionary approach*. Harlow, England: Longman.
- Croft, William. 2001. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- De Smet, Hendrik. 2020. What predicts productivity? Theory meets individuals. *Cognitive Linguistics* 31(2). 251–278.
- Elman, Jeffrey L. 1995. Language as a dynamical system. In Robert F. Port & Timothy van Gelder (eds.), *Mind as motion: Explorations in the dynamics of cognition*, 195–225. Cambridge, MA: The MIT Press.
- Fagyal, Zsuzsanna, Samarth Swarup, Anna María Escobar, Les Gasser & Kiran Lakkaraju. 2010. Centers and peripheries: Network roles in language change. *Lingua* 120(8). 2061–2079.
- Foolen, Ad. 2015. Word valence and its effects. In Ulrike M. Lüdtkke (ed.), *Emotion in language: Theory – Research – Application*, 241–256. Amsterdam, Netherlands: John Benjamins.
- Fried, Mirjam. 2013. Principles of constructional change. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford handbook of construction grammar*, 419–437. Oxford: Oxford University Press.
- Givón, Talmy. 1979. From discourse to syntax: Grammar as a processing strategy. In Talmy Givón (ed.), *Discourse and syntax*, 81–112. Leiden: Brill.
- Goldberg, Adele E. 2019. *Explain me this: Creativity, competition, and the partial productivity of constructions*. Princeton: Princeton University Press.
- Grice, Paul. 1989. *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- Guo, Rui (郭锐). 2004. *Xiandai hanyu cilei yanjiu* 现代汉语词类研究 [On word classes in modern Chinese]. Beijing: The Commercial Press.
- Hartmann, Stefan & Tobias Ungerer. 2024. Attack of the snowclones: A corpus-based analysis of extravagant formulaic patterns. *Journal of Linguistics* 60(3). 599–634.
- Haspelmath, Martin. 1999. Why is grammaticalization irreversible? *Linguistics* 37(6). 1043–1068.
- Heine, Bernd & Mechthild Reh. 1984. *Grammaticalization and reanalysis in African languages*. Hamburg: Helmut Buske.
- Hilpert, Martin. 2021. *Ten lectures on diachronic construction grammar*. Leiden: Brill.
- Hilpert, Martin & Samuel Bourgeois. 2020. From confrontation to solidarity in the sarcastic *much*? Construction. *Constructions and Frames* 12(1). 96–120.
- Hoffmann, Thomas & Mark Turner. 2025. *Creative construction grammar*. Cambridge: Cambridge University Press.
- Huang, C.-T. James & Na Liu. 2014. Xinxing feidianxing beidongshi “bei XX” de jufa yu yuyi jiegou 新兴非典型被动式“被XX”的句法与语义结构 [The syntax and semantics of the new non-canonical Bei XX construction]. *Yuyan kexue* 语言科学 [Language Science] 70(5). 225–241.
- Jing-Schmidt, Zhuo. 2005. *Dramatized discourse: The Mandarin Chinese ba-construction*. Amsterdam/Philadelphia: John Benjamins.
- Jing-Schmidt, Zhuo. 2007. Negativity bias in language: A cognitive-affective model of emotive intensifiers. *Cognitive Linguistics* 18(3). 417–443.
- Jing-Schmidt, Zhuo. 2024. Yufa jingyi: lun yuyan xijuhua de renzhi qinggan jizhi 语法惊异: 论语言戏剧化的认知情感机制 [Grammatical surprise: On the cognitive affective mechanism of linguistic dramaticity]. *Dangdai yuyanxue* 当代语言学 [Contemporary Linguistics] 26(3). 317–335.

- Jing-Schmidt, Zhuo & Hongyin Tao. 2009. The Mandarin disposal constructions: Usage and development. *Language and Linguistics* 10(1). 29–58.
- Kagan, Jerome. 2002. *Surprise, uncertainty, and mental structures*. Cambridge, MA: Harvard University Press.
- Ke, Jinyun, Tao Gong & William S.-Y. Wang. 2008. Language change and social networks. *Communications in Computational Physics* 3(4). 935–949.
- Keenan, Edward. 1975. Some universals of passives in universal grammar. In *Papers from the eleventh regional meeting of the Chicago Linguistic Society*, 340–352. Chicago: Chicago Linguistics Society.
- Labov, William. 2001. *Principles of linguistic change, Vol. 2: Social factors*. Oxford: Blackwell.
- Labov, William. 2012. What is to be learned: The community as the focus of social cognition. *Review of Cognitive Linguistics* 10(2). 265–293.
- Lang, Jun & Zhuo Jing-Schmidt. 2024. The blurry lines between popular media and party propaganda: China's convergence culture through a linguistic lens. *PLoS One* 19(1). e0297499.
- Lehmann, Christian. 1985. Grammaticalization: Synchronic variation and diachronic change. *Lingua e stile* 20. 303–318.
- Li, Charles N. & Sandra A. Thompson. 1989. *Mandarin Chinese: A functional reference grammar*. Oakland: University of California Press.
- Michaelis, Laura A. & Hanbing Feng. 2015. What is this, sarcastic syntax? *Constructions and Frames* 7(2). 148–180.
- Mikolov, Tomas, Ilya Sutskever, Kai Chen, Gre Corrado & Dean Jeffrey. 2013. Distributed representations of words and phrases and their compositionality. In Christopher J. Burges, Léon Bottou, Zoubin Ghahramani & Kilian Q. Weinberger (eds.), *Advances in neural information processing systems* 26, 3111–3119. Red Hook, NY: Curran Associates, Inc.
- Moisl, Hermann. 2015. *Cluster analysis for corpus linguistics*. Berlin/Boston: Mouton de Gruyter.
- Noël, Dirk. 2007. Diachronic construction grammar and grammaticalization theory. *Functions of Language* 14(2). 177–202.
- Office of Lexicography, Institute of Linguistics, Chinese Academy of Social Sciences. 2012. *Xiandai hanyu cidian (di 6 ban) 现代汉语词典(第六版)[Modern Chinese dictionary (6th edition)]*. Beijing: The Commercial Press.
- Office of Lexicography, Institute of Linguistics, Chinese Academy of Social Sciences. 2016. *Xiandai hanyu cidian (di 7 ban) 现代汉语词典(第七版)[Modern Chinese dictionary (7th edition)]*. Beijing: The Commercial Press.
- Paul, Hermann. 1909. *Prinzipien der Sprachgeschichte*. Vierte Auflage. Halle an der Saale: Verlag von Max Niemeyer.
- Peng, Xinjia. 2017. *Linguistic innovations in Chinese: Internal and external factors*. Eugene: University of Oregon PhD dissertation.
- Reisenzein, Rainer, Gernot Horstmann & Achim Schützwohl. 2019. The cognitive-evolutionary model of surprise: A review of the evidence. *Topics in Cognitive Science* 11(1). 50–74.
- Rozin, Paul & Edward B. Royzman. 2001. Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review* 5(4). 296–320.
- Säily, Tanja, Victorina González-Díaz & Jukka Suomela. 2018. Variation in the productivity of adjective comparison in present-day English. In Voclav Brezina, Robbie Love & Karin Aijmer (eds.), *Corpus approaches to contemporary British speech: Sociolinguistic studies of the spoken BNC2014*. London: Routledge.
- Schmid, Hans-Jörg. 2015. A blueprint of the entrenchment-and-conventionalization model. *Yearbook of the German Cognitive Linguistics Association* 3(1). 3–26.

- Schmid, Hans-Jörg. 2016. Why cognitive linguistics must embrace the social and pragmatic dimensions of language and how it could do so more seriously. *Cognitive Linguistics* 27(4). 543–557.
- Schmid, Hans-Jörg. 2020. How the entrenchment-and-conventionalization model might enrich diachronic construction grammar: The case of *(the) thing is (that)*. *Belgian Journal of Linguistics* 34. 306–319.
- Schmid, Hans-Jörg & Franziska Günther. 2016. Toward a unified socio-cognitive framework for salience in language. *Frontiers in Psychology* 7. Article 1110. <https://doi.org/10.3389/fpsyg.2016.01110>.
- Scott, Kate. 2021. You won't believe what's in this paper! Clickbait, relevance and the curiosity gap. *Journal of Pragmatics* 175. 53–66.
- Shen, Jiaxuan (沈家煊). 2010. Shishuoxinyu sanze pingshuo-bei-zisha, xixiao gongzuo, you haojiu 世说新语三则评说--被自杀·细小工作·有好酒 [Commentary on three neologisms: Bei suicide, small work, have good wine]. *Dangdai xiucixue* 当代修辞学 [Contemporary Rhetoric] (4). 93–95.
- Shi, Chunhong (施春宏). 2013. Xin bei-zishi de shengcheng jizhi, yuyi lijie ji yuyong xiaoying 新“被”字式的生成机制, 语义理解及语用效应 [The mechanism of emergence, comprehension and pragmatic effect of the new BEI pattern]. *Dangdai xiucixue* 当代修辞学 [Contemporary Rhetoric] (1). 12–28.
- Shibatani, Masayoshi. 1985. Passives and related constructions: A prototype analysis. *Language* 61(4). 821–848.
- Song, Yan, Shuming Shi, Jing Li & Haisong Zhang. 2018. Directional skip-gram: Explicitly distinguishing left and right context for word embeddings. In *Proceedings of the 2018 conference of the North American chapter of the Association for Computational Linguistics: Human language technologies*, vol. 2, 175–180. New Orleans: Association for Computational Linguistics.
- Sun, Chaofen. 1996. *Word order change and grammaticalization in the history of Chinese*. Stanford: Stanford University Press.
- Syakur, Muhammad A., Bain K. Khotimah, Eka Mala Sari Rochman & Budi Dwi Satoto. 2018. Integration k-means clustering method and elbow method for identification of the best customer profile cluster. *IOP Conference Series: Materials Science and Engineering* 336(1). 012017.
- Traugott, Elizabeth Closs. 2018. Rethinking the role of invited inferencing in change from the perspective of interactional texts. *Open Linguistics* 4(1). 19–34.
- Traugott, Elizabeth Closs & Graeme Trousdale. 2010. Gradience, gradualness and grammaticalization: How do they intersect? In *Gradience, gradualness and grammaticalization*, 19–44. Amsterdam: John Benjamins.
- Traugott, Elizabeth Closs & Graeme Trousdale. 2013. *Constructionalization and constructional changes*. Oxford: Oxford University Press.
- Wang, Li (王力). 1957. *Hanyu Shigao* 汉语史稿 [History of Chinese language]. Beijing: Zhonghua Book Company.
- Wang, Junxi (王均熙). 2011. *Hanyu xinci cidian* 汉语新词词典 (2005–2010) [Dictionary of Chinese neologisms (2005–2010)]. Shanghai: Academia Press.
- Wang, Yin (王寅). 2011. Xin bei-zi goushi de cihui yazhi jixi – dui bei-ziyuan yilei xinbiaoda de renzhi goushi yufa yanjiu “新被字构式”的词汇压制解析--对“被自愿”一类新表达的认知构式语法研究 [A lexical coercion analysis on the new BEI construction: A cognitive constructionist analysis on expressions like “bei ziyuan”]. *Waiguoyu* 外国语 [Journal of Foreign Languages] 34(3). 13–20.
- Xiao, Yun. 2016. The rise of catchword “bei-X”: Grammaticalization and reanalysis. *Chinese as a Second Language* 51(1). 79–97.
- Xiao, Richard, Tony McEnery & Yufang Qian. 2006. Passive constructions in English and Chinese: A corpus-based contrastive study. *Languages in Contrast* 6(1). 109–149.
- Yao, Jun, Jie Song & Michael Singh. 2013. The ironical Chinese *bei*-construction and its accessibility to English speakers. *Journal of Pragmatics* 55. 195–209.

- Yuan, Yongxue & Yi Zhang. 2026. The Chinese non-canonical “Bei X” construction: A snowclone construction analysis. *Lingua* 340. 104197.
- Zhang, Yi. 2020. Non-canonical passives in Chinese. A mental space account. *Chinese Language and Discourse* 11(1). 84–106.

Supplementary Material: This article contains supplementary material (<https://doi.org/10.1515/cog-2025-0167>).