International Journal of Linguistics, Literature and Translation
ISSN: 2617-0299 (Online); ISSN: 2708-0099 (Print)
DOI: 10.32996/ijllt
Journal Homepage: www.al-kindipublisher.com/index.php/ijllt

IJLLT
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Why Do Chinese Speakers Prefer the Word 'Shang' (Harm) ? Exploring the Semantic Motivations Underlying Nativelike Selections

Huasheng Zhang
College of Liberal Arts, Sichuan Normal University, China
Corresponding Author: Huasheng Zhang, **E-mail**: zhang.huasheng@foxmail.com

| ABSTRACT
This study challenges the assumption that nativelike selections—linguistic expressions preferred by native speakers over grammatical alternatives—are inherently arbitrary. Focusing on Chinese speakers' preference for *shang* 'harm' in *bu shang shou* 'not harm hands' to praise dish soaps' skin-friendliness, we reveal the semantic motivation that drives this lexical choice. This semantic motivation is rooted in *shang*'s pragmatic nuance, specifically its prototype of "self-care suggestions for everyday life". The findings support the hypothesis that a concept is more readily expressed by a certain linguistic expression whose prototype aligns with that concept.

| KEYWORDS
Semantic motivation, nativelike selection, prototype, arbitrary collocation

| ARTICLE INFORMATION
ACCEPTED: 02 February 2025 **PUBLISHED:** 25 February 2025 **DOI:** 10.32996/ijllt.2025.8.2.16

1. Introduction

This is a follow-up study to Zhang (2025a, 2025b), which analyzed multilingual customer reviews of a dish soap product. These studies discovered that Chinese speakers prefer to use *shang* 'harm' as in *bu shang shou* 'not harm hands' to praise the skin feel of this product, whereas English and Japanese speakers favor other lexical choices such as *kind to skin* and *te ni yasashii* 'gentle to hands'. This is a long-discussed phenomenon that falls under the concept of "nativelike selection"—linguistic expressions preferred by native speakers over other grammatical alternatives to convey a specific concept.

The case of Chinese speakers' preference for *shang* 'harm' in these texts, poses an explanatory challenge to the widely accepted view of nativelike selections. Over the years, arbitrariness—or conventionality—has been assumed for the property of nativelike selections. A radical perspective within this framework asserts that "[a]ll lexical items are arbitrary—they are simply the consensus of what has been institutionalised" (Lewis, 1997: 17). However, it is difficult to view *bu shang shou* 'not harm hands' as an arbitrary collocation, favored by native speakers conventionally. There is nothing special about the verb *shang* 'harm' in this fully-transparent phrase; *shang* can freely combine with numerous other words to convey the same meaning of "harm".

This study aims to explore the semantic motivation in the underlying mechanism of nativelike selections, moving beyond the simplistic attribution to arbitrariness. Specifically, this study focuses on a case study of Chinese nativelike selection *bu shang shou* 'not harm hands', and employs the theoretical framework in Grondelaers & Geeraerts (2003) to investigate the semantic motivations underlying nativelike selections. Grondelaers & Geeraerts (2003) propose a hypothesis that if a concept aligns with a certain word's prototype, then it is preferably expressed with that word by native speakers.

This kind of research into how and why a concept is expressed by different linguistic items is relatively uncommon (Geeraerts, 2015: 242). Recent studies by Liu (2013) and Mehl (2018) have further supported Grondelaers & Geeraerts's (2003) hypothesis on the semantic motivation underlying nativelike selections. This study contributes new evidence, enriching our understanding of this understudied but crucial aspect of the dynamics of language.

The following sections first introduce studies that challenge the arbitrariness of nativelike selections, highlighting the need for a semantic explanation of this phenomenon. Such semantic motivation can be captured by Grondelaers & Geeraerts's (2003)

hypothesis, which proposes that a concept is more readily expressed by a certain linguistic expression whose prototype aligns with that concept. Grondelaers & Geeraerts's (2003) relevant studies, including the prototype theory, are then discussed. Finally, the corpus analysis reveals the language-specific pragmatic nuances of *shang*'s prototype, elucidating the semantic motivation (why it is favored by native speakers) behind the nativelike selection *bu shang shou* 'not harm hands' in customer reviews of the dish soap.

2. Literature Review

2.1 Studies Challenging the Arbitrariness of Nativelike Selections

The idea of almost equating nativelike selections with arbitrary word combinations can been seen in the extensive literature of formulaic language studies (Pawley & Syder, 1983; De Cock, 1998; Erman & Warren, 2000; Erman, 2009; Smiskova et al., 2012; Wray, 2012; Ortaçtepe, 2013; Erman et al., 2016). Due to this belief, it is advised by researchers and teachers in the second language classroom that these linguistic expressions should be taught and learned as an integral whole rather than being taken apart and analyzed in isolation for their underlying rules (Liu, 2010; Woolard, 2000).

However, there are opposing evidence and viewpoints challenging this idea. Walker (2008), Liu (2010), and Bosque (2011) questioned the assumed arbitrariness of some collocations, highlighting their potential analyzability and deducibility (e.g., **powerful coffee* does not necessarily prove the arbitrariness of *strong coffee* when considering the nuanced semantic distinctions between *powerful* and *strong*); Alonso Ramos (2017) also summarized views in this perspective. This deducibility is usually analyzed from a "predicate-driven" perspective rather than the traditional "from argument/base to predicate/collocate" direction (Almela-Sánchez, 2019). Nesselhauf (2005: 32, 35) also noted the challenge in categorizing collocations, pointing out that some combinations can be interpreted as either arbitrary or not (e.g., *?commit a lie* may not affirm the arbitrariness of *commit a suicide* if the latter is viewed as a stand-alone sense of *commit*).

In addition, the phenomenon of nativelike selection extends beyond words that conventionally co-occur. Even though Foster (2009) and Zaytseva (2016) consider nativelike selections as (arbitrary) formulaic language, they also observed "nativelike selection of individual words". For instance, learners tend to use broad-meaning basic vocabularies (e.g., *understand*) in contrast to native speakers, who often prefer narrowly defined terms (e.g., *notice*, *realize*, *pay attention,* and *care*). This indicates that nativelike selections include the nuanced selection of individual words, which need not be directly linked to arbitrariness.

2.2 Semantic Motivation Derived from Lexical Prototypes

The above findings suggest that not all nativelike selections are necessarily arbitrary, allowing room for semantic explanation and analyzability. This issue can be addressed through the hypothesis proposed by Geeraerts et al. (1994), and Grondelaers & Geeraerts (2003). They highlight prototypicality as the underlying semantic motivation for a lexical item to express a concept, rather than attributing it to arbitrariness. The hypothesis under investigation in this study is as follows:

"A referent is more readily named by a lexical item if it is a salient member of the category denoted by that item" (Grondelaers & Geeraerts 2003: 74).

This delineates the empirical findings in Geeraerts et al. (1994), which compiled a database of clothing photos paired with their corresponding referring words from fashion magazines, and discovered "In plain language: when you have to name something, you preferentially choose those items of which the thing to be named is a typical representative". In recent years, this hypothesis has been supported by Liu (2013) and Mehl (2018). Liu (2013) investigated native speakers' use of synonyms in different contexts, finding that they prefer synonyms whose prototypes align closely with the intended contextual meaning. Liu (2013) noted that this phenomenon aligns with the findings of Geeraerts et al. (1994). Similarly, Mehl (2018) validated this hypothesis through a study on English verbs such as *make*, *take*, and *give*. The prototypes of these verbs are concrete rather than abstract, and native speakers are more likely to use them when referring to concrete uses as opposed to their abstract alternatives, such as *produce*, *collect*, and *provide*.

2.3 Prototype Theory

To explore the prototypes of words, this section introduces prototype theory. Prototype theory, first proposed by Rosch (1973), originated in the field of psychology. This theory posits that members of a category exhibit varying degrees of typicality. For any given category, there are "good examples" and "bad examples", as well as "typical members" and "marginal members". The central members of a category are its prototypes. For instance, in the category of 'bird', a prototypical bird is characterized by features such as feathers, a beak, and the ability to fly. Ostriches and penguins are marginal members of the "bird" category, whereas sparrows and swallows are prototypical. The category of "bird" is defined by its prototype (typical birds), while ostriches and penguins form a peripheral subset. Following its introduction in psychology, linguists have extensively applied prototype theory to the semantic analysis of words. One early study is Fillmore (1977), which employed prototype theory to analyze the semantics of the word "bachelor".

Gilquin (2008) points out that there are different yet interconnected definitions for the concept of prototype. Of them, the most relevant ones to this study, are the most cognitively salient exemplar among members of a category, and the most frequent

item in the corpus. To ensure objectivity, practicality, and reproducibility in this study, this study adopts the latter perspective, utilizing corpus-based methods to investigate the prototypes of lexical items.

2.4 Usage Clusters as Categorizations of Lexical Prototypes

This study adopts the view that a word can have multiple prototypes, different from the conventional view which only posits a single prototype (Gilquin, 2008). This approach enables a more fine-grained analysis of word usage and its prototypes. The notion that an individual sense of a word can have a prototype, can be found in Geeraerts (1993, 1997: 61), Lewandowska-Tomaszczyk (2007), and Fu (2015). This can be exemplified by Ambridge's (2020) illustration: spoons are generally small and metal or large and wooden; but there can not be a single prototype that is intermediate in these attributes.

To analyze the categorizations of words' prototypes, this study employs a cluster-based methodology to address the challenges in identifying different senses that correspond to different prototypes. Determining word senses is "notoriously problematic" (Gries, 2006; Artstein & Poesio, 2008), particularly in addressing the issues of "splitting or lumping" or "vagueness vs. polysemy" problem (Geeraerts, 1993; Tuggy, 1993). Different scholars (Taylor, 2012: 223) and dictionaries (Jackson, 2002: 88) favor different approaches of these two. Also, every polysemy test is not without problems, as extensively discussed by Riemer (2005). Moreover, word meanings are inherently flexible and lack clearly defined boundaries (Geeraerts, 2006: 74).

The underlying principle behind the cluster-based analysis of prototype categorizations is "senses and corresponding prototypes as clusters of usages", based on distributionalism which holds that different meanings are associated with different contexts (Lenci, 2008). This idea has been widely put into practice applied in natural language processing under the term "word sense induction/discrimination" (Navigli, 2009; Nasiruddin, 2013), where "the number of clusters indicates the number of the target word's senses" (Ghayoomi, 2021). This principle is similarly reflected in the work of linguists such as Glynn (2010a, 2016), Polguère (2018), and Gries (2019).

3. Materials and Methods

To investigate the semantic motivation behind the nativelike selection *bu shang shou* 'not harm hands'—specifically, why it is favored by native speakers—this study follows previous studies that emphasize the role of lexical prototypes. The prototype of *shang* 'harm' is analyzed in comparison with its near-synonyms or its near equivalents in Chinese (*shanghai*), English (*harm*), and Japanese (*kizutsukeru*). For a fine-grained analysis, only the present negative forms of these 4 words were collected and analyzed because syntactic variations influence word usage (Stubbs, 2001; Glynn, 2010a, 2010b). The results aim to uncover why the present negative form of *shang* is preferred to praise the skin feel of dish soap.

This study employs the CC-100 corpus (Conneau et al., 2020), a multilingual dataset derived from web-crawled texts from 2018. From CC-100, 5160 instances of the present negative forms of *shang*, *shanghai*, *harm,* and *kizutsukeru* were randomly sampled respectively. The following processes were performed on the collected data: a cluster-based analysis of *shang*'s prototype categorization, and a behavioral profile analysis (Divjak & Gries, 2006) of *shang*'s pragmatic nuance in its prototype.

Following Velasco et al. (2022), and Saidi & Jarray (2023), the steps in cluster-based analysis include generating embeddings for sentences containing the target word, and clustering of these embeddings; 5160 sentences were analyzed for each word. The steps in behavioral profile analysis (Divjak & Gries, 2006) include annotation of semantic features, and correspondence analysis of the annotation results; 100 sentences were annotated for each word, resulting in a total of 11,600 tags annotated.

4. Results and Discussion

4.1 Results of *Shang*'s Prototype Categorization

By conducting categorization of *shang*'s prototypes, rather than analyzing them as a whole, this fosters a more fine-grained analysis of *shang*'s pragmatic nuance in the nativelike selection of *bu shang shou* 'not harm hands', which helps uncover the semantic motivation underlying this nativelike selection.

The results of *shang*'s prototype categorization are summarized in Figure 1 and are compared with *shanghai*, *harm,* and *kizutsukeru*. The results show that the prototype categorizations of these 4 words are similar, which is consistent with their status as near-synonyms, and indicates the reliability of this analysis. *Shang*'s usages can be divided into 2 clusters, indicating two distinct prototypes with different meanings, as suggested in Section 2.4.

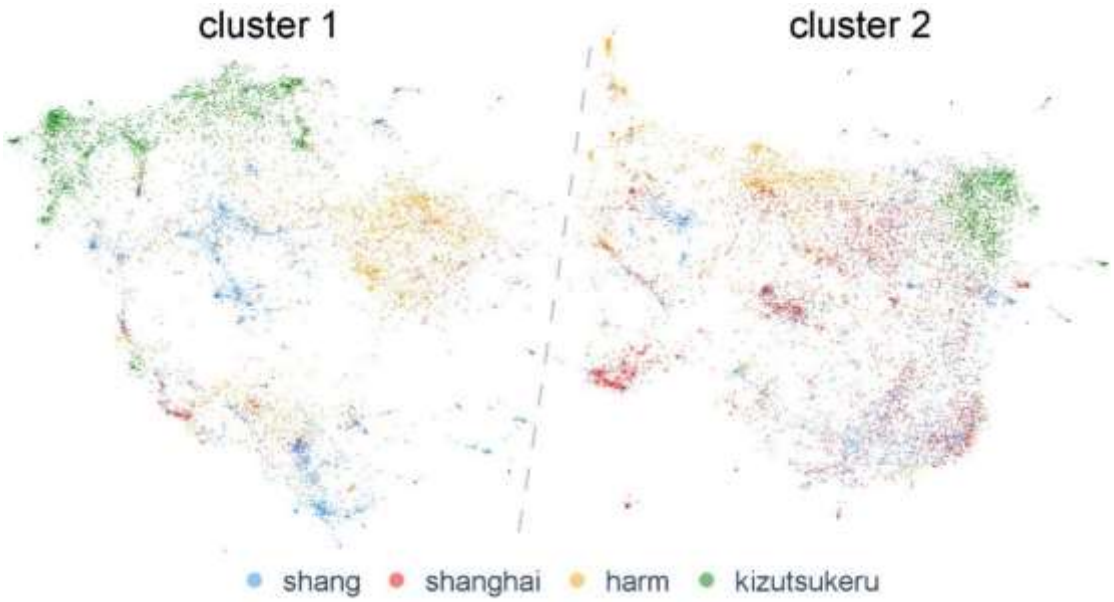

Figure 1: Visualization of the prototype categorization for *shang*, *shanghai*, *harm*, and *kizutsukeru*.

By manually inspecting the sentences in these two cluster, it is identified that cluster 1, rather than cluster 2, corresponds to *shang*'s usage in *bu shang shou* 'not harm hands'. The analysis indicates that cluster 1 is about health-related harm as shown in (1), while cluster 2 is about emotional or destructive harm as demonstrated in (2) and (3).

(1) *Dan zhiyao jizhu yixia zhe qi dian, paobu budan bu shang xi, hai neng zhuang xi.*
'But as long as you remember these seven points, running not only doesn't harm the knees but can even strengthen them.'

(2) *"Xiang'ai bu yiding yao zai yiqi", ye chengwei le jujue duifang de, zui bu shang ren de liyou.*
''Loving each other doesn't mean we have to be together', has become the rejection that does not harm anyone.'

(3) *Yong fangshexing wuzhi sha si tamen, weifan le guoji jian bu shang pingmin de zhunze.*
'Using radioactive substances to kill them violates the international principle of not harming civilians.'

Additionally, the findings from manual inspection are supported by analyzing the top five most frequent objects of *shang* 'harm' across the clusters. The results further demonstrate that cluster 1 is characterized by harm to body parts, contrasting with cluster 2, where the usage of emotional harm is dominant.

Table 1: Comparison of *shang's* usage in cluster 1 and 2.

Usage type	Most frequent objects of *shang*
Cluster 1	*shen* 'body', *shou* 'hand', *shenti* 'body', *yan* 'eye', *pifu* 'skin'
Cluster 2	*ren* 'person', *ganqing* 'emotion', *xin* 'heart', *shen* 'body', *zizun* 'self-respect'

4.2 Results of *Shang*'s Pragmatic Nuance in Its Prototype

Findings in Section 4.1 indicate that the semantic motivation behind the nativelike selection *bu shang shou* 'not harm hands', can be revealed by focusing on the prototype of cluster 1 in *shang*'s usage. A behavioral profile analysis was conducted on the prototype of *shang*'s cluster 1, as described in Section 3. The results are summarized and visualized in Figure 2. For details of how to interpret this plot, see Bock (2011).

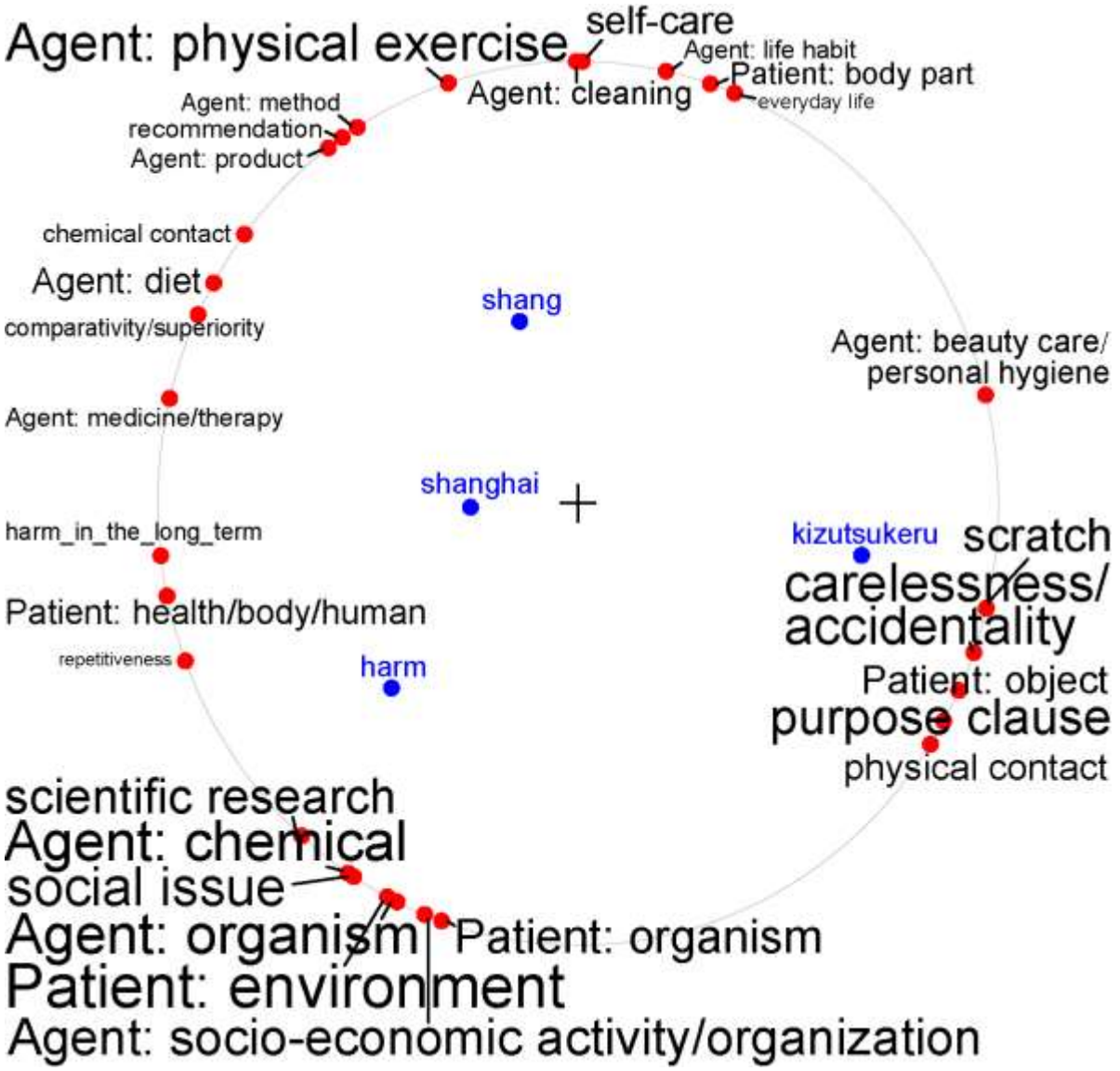

Figure 2: *Shang*'s pragmatic nuance in comparison with *shanghai*, *harm,* and *kizutsukeru.* The relative associations between words and semantic features are represented by their spatial distance and font sizes (e.g., *shang* is associated with features in the upper circle).

Results in Figure 2 reveal *shang*'s language-specific pragmatic nuance in its prototype, distinguishing it from the other words. The four words occupy distinct positions in the plot, reflecting their different nuances in meaning. In summary, *shang* is distinctively associated with self-care suggestions for everyday life; *shanghai* and *harm* are more general in meaning without strong pragmatic nuances like *shang*; *kizutsukeru* is distinctively associated with careless or accidental scratches.

The prototypical meaning of Chinese *shang* 'harm' is closely linked to self-care suggestions for everyday life. It is associated with "everyday life" and related agents like "life habit", "diet", "cleaning", "physical exercise", and "product"; additionally, *shang* is high in raw percentages for "repetitiveness" and "harm_in_the_long_term" which are related to everyday life. Moreover, it is associated with "self-care", which is about ways to take care of personal health. This is related to features like "Agent: method", "comparativity/superiority" and "recommendation", which are about the ways. Health-related features include "Agent: medicine/therapy" and "Patient: body part". In fact, the features about everyday life and self-care are inherently interconnected, because self-care is about what we choose to do in everyday life routines.

As Segall & Goldstein (1989) summarized, self-care is associated with elements such as laypersons rather than professionals, consultation with non-medical health care practitioners, lifestyle and routine health maintenance activities (e.g., eating, sleep, exercise, and personal hygiene), and chronic morbidity. These are closely related to the semantic features associated with *shang* in Figure 2.

The prototype of *shang* can be exemplified in (4) and (5). They recommend self-care tips in everyday life that teach the correct life habits for eating and exercising. When eating and exercising, we have to choose the foods and methods. However, we are

concerned that if we do not follow the right practices, these routines could harm our health in the long term. Normal ways of breastfeeding and jumping rope may be harmful to our health and knees to a degree, especially in the long term. Therefore, by using *shang*, better ways which avoid or minimize the harm are recommended instead.

(4) *Women de shenghuo zhong you bushao yongcan wuqu, name zenme chi cai neng zai yuezi li rang xin mamamen ji you zugou yingyang gongji baobao naishui er you neng bu shang ziji shenti ne?*
'There are many misconceptions about eating habits in our lives. So, how can new moms eat during the postpartum confinement to ensure sufficient nutrition for breastfeeding while not harming their own bodies?'

(5) *Tiaosheng bu shang xigai jiu yao jiangjiu fangshi fangfa, zhengque de fangshi, bujin zengjia xiaoguo, hai neng baohu hao xigai.*
'If you want that jumping rope doesn't harm your knees, you need to pay attention to the method. A correct method not only enhances the benefits but also protects your knees.'

In summary, when we say *A bu shang B* 'A does not harm B', we typically mean the following pragmatic nuance in *shang*'s prototype that is about self-care suggestions for everyday life:

a. We choose how to do things and what to use in everyday life habits (e.g., what to eat and what products to use).
b. B is usually a body part.
c. Poor choices result in worsening the health of B in the long term.
d. We want to maintain good body health.
e. A is recommended because it does not harm B's health and is considered an appropriate and better choice.

4.3 Semantic Motivation Underlying the Nativelike Selection

Bu shang shou 'not harm hands' is a nativelike selection favored by Chinese speakers to praise the skin feel of the dish soap product. An analysis of lexical collocations with *shou* 'hand' in the customer reviews reveals that *shang* 'harm' is the most frequent word. If the analysis were limited to this observation, one might conclude that the frequent collocation of them is arbitrary in the texts, as there appears to be no obvious reasons for preferring *shang* 'harm' over alternatives such as *hao* 'good' or *gan* 'dry'. However, the underlying semantic motivation can be revealed if a broader analysis is involved by further studying *shang*'s entire usages beyond its combination with *shou* 'hand'. This uncovers its special pragmatic nuance in its prototype, which explains Chinese speakers' preference of it in the customer reviews.

Specifically, the analysis in this study is grounded in the hypothesis of Grondelaers & Geeraerts (2003). According to this hypothesis, if a concept aligns with a certain word's prototype, then it is preferably expressed with that word by native speakers. The concept of "praising the skin feel of the dish soap product" is about praising its favorable effect on people's health, and recommending this household necessity that people use in daily life. This concept aligns with *shang*'s prototype which is associated with self-care suggestions for everyday life. Consequently, *bu shang shou* 'not harm hands' becomes the nativelike selection in this customer review. Additionally, an interesting comparison can be made with English and Japanese linguistic usages in the same texts. Neither the English term *harm* nor the Japanese term *kizutsukeru* 'harm' carries the same pragmatic nuance, which explains why they are not chosen as nativelike selections by their native speakers.

5. Conclusion

This study challenges the conventional assumption that nativelike selections are fundamentally arbitrary, by demonstrating the role of semantic motivation in shaping linguistic preferences. Through a semantic analysis of the Chinese nativelike selection *bu shang shou* 'not harm hands' in the customer review texts, this research reveals that the verb *shang* 'harm' is favored not merely by convention but because of its pragmatic nuance in its prototype. The Corpus-based clustering analysis and behavioral profile analysis identified two distinct prototypes of *shang*: one centered on health-related harm and is associated with "self-care suggestions for everyday life", and the other on emotional or destructive harm. The former aligns with the pragmatic context of product reviews, where *bu shang shou* 'not harm hands' functions as a recommendation for maintaining skin health through appropriate choice of household necessity (i.e., the dish soap) that people use in everyday life. These findings validate Grondelaers & Geeraerts' (2003) hypothesis that a concept is more readily be expressed by a word whose prototype aligns with that concept. There are also limitations in this research. How many nativelike selections are arbitrary or analyzable by investigating the semantic motivations is beyond the scope of this study, and this can be further studied in future research.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Almela-Sánchez, M. (2019). Collocation and selectional preferences: A frame-based approach. *Journal of English Studies*, *17*, 3–41. https://doi.org/10.18172/jes.3905

[2] Alonso Ramos, M. (2017). Can collocations be deduced?: A lexically driven analysis from the perspective of language production. In Sergi Torner & Elisenda Bernal (Eds.), *Collocations and other lexical combinations in Spanish: Theoretical, lexicographical and applied perspectives* (pp. 21–40). Routledge.

[3] Ambridge, B. (2020). Against stored abstractions: A radical exemplar model of language acquisition. *First Language*, *40*(5–6), 509–559. https://doi.org/10.1177/0142723719869731

[4] Artstein, R., & Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics*, *34*(4), 555–596. https://doi.org/10.1162/coli.07-034-R2

[5] Bock, T. (2011). Improving the display of correspondence analysis using moon plots. *International Journal of Market Research*, *53*(3), 307–326. https://doi.org/10.2501/IJMR-53-3-307-326

[6] Bosque, I. (2011). Deducing collocations. In I. Boguslavsky & L. Wanner (Eds.), *Proceedings of the 5th international conference on meaning-text theory* (pp. vi–xxiii). Universitat Pompeu Fabra.

[7] Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2020). Unsupervised cross-lingual representation learning at scale. In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 8440–8451). Association for Computational Linguistics.

[8] De Cock, S. (1998). A recurrent word combination approach to the study of formulae in the speech of native and non-native speakers of English. *International Journal of Corpus Linguistics*, *3*(1), 59–80. https://doi.org/10.1075/ijcl.3.1.04dec

[9] Divjak, D., & Gries, S. T. (2006). Ways of trying in Russian: Clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory*, *2*(1), 23–60. https://doi.org/10.1515/CLLT.2006.002

[10] Erman, B. (2009). Formulaic language from a learner perspective: What the learner needs to know. In Roberta Corriga, Edith A. Moravcsik, Hamid Ouali, & Kathleen Wheatley (Eds.), *Formulaic languagee: Vol. 2. Acquisition, loss, psychological reality, and functional explanations* (pp. 323–346). John Benjamins.

[11] Erman, B., Forsberg Lundell, F., & Lewis, M. (2016). Formulaic language in advanced second language acquisition and use. In K. Hyltenstam (Ed.), *Studies on Language Acquisition* (pp. 111–148). De Gruyter Mouton.

[12] Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text & Talk*, *20*(1), 29–62. https://doi.org/10.1515/text.1.2000.20.1.29

[13] Fillmore, C. J. (1977). Scenes-and-frames semantics. In A. Zampolli (Ed.), *Linguistic Structures Processing*. North-Holland Publishing Company.

[14] Foster, P. (2009). Lexical diversity and native-like selection: The bonus of studying abroad. In B. Richards, M. H. Daller, D. D. Malvern, P. Meara, J. Milton, & J. Treffers-Daller (Eds.), *Vocabulary studies in first and second language acquisition: The interface between theory and application* (pp. 91–106). Palgrave Macmillan.

[15] Fu, X. (2015). Meaning Prototype: A Study of '110' in Chinese. *Language and Linguistics*, *17*(3), 321–349. https://doi.org/10.1177/1606822X15586689

[16] Geeraerts, D. (1993). Vagueness's puzzles, polysemy's vagaries. *Cognitive Linguistics*, *4*(3), 223–272. https://doi.org/10.1515/cogl.1993.4.3.223

[17] Geeraerts, D. (1997). *Diachronic prototype semantics: A contribution to historical lexicology*. Clarendon Press.

[18] Geeraerts, D. (2006). *Words and other wonders: Papers on lexical and semantic topics*. De Gruyter Mouton.

[19] Geeraerts, D. (2015). Sense individuation. In N. Riemer (Ed.), *The Routledge handbook of semantics*. Routledge.

[20] Ghayoomi, M. (2021). Word sense induction in Persian and English: A comparative study. *Journal of Information Systems and Telecommunication (JIST)*, *9*(36), 263–274. https://doi.org/10.52547/jist.9.36.263

[21] Gilquin, G. (2008). What you think ain't what you get: Highly polysemous verbs in mind and language. In J.-R. Lapaire, G. Desagulier, & J.-B. Guignard (Eds.), *Du fait grammatical au fait cognitif. From gram to mind: Grammar as cognition* (Vol. 2, pp. 235–255). Presses Universitaires de Bordeaux.

[22] Glynn, D. (2016). Quantifying polysemy: Corpus methodology for prototype theory. *Folia Linguistica*, *50*(2), 413–447. https://doi.org/10.1515/flin-2016-0016

[23] Glynn, D. (2010a). Synonyme, lexical fields, and grammatical constructions. A study in usage-based cognitive semantics. In H.-J. Schmid & S. Handl (Eds.), *Cognitive Foundations of Linguistic Usage Patterns* (pp. 89–118). De Gruyter Mouton.

[24] Glynn, D. (2010b). Testing the hypothesis. Objectivity and verification in usage-based cognitive semantics. In D. Glynn & K. Fischer (Eds.), *Quantitative methods in cognitive semantics: Corpus-driven approaches* (pp. 239–270). De Gruyter Mouton.

[25] Gries, S. T. (2006). Corpus-based methods and cognitive semantics: The many senses of to run. In S. T. Gries & A. Stefanowitsch (Eds.), *Corpora in cognitive linguistics: Corpus-based approaches to syntax and lexis* (pp. 57–99). De Gruyter Mouton.

[26] Gries, S. T. (2019). Polysemy. In E. Dąbrowska & D. Divjak (Eds.), *Cognitive linguistics—Key topics* (pp. 23–43). De Gruyter Mouton.

[27] Grondelaers, S., & Geeraerts, D. (2003). Towards a pragmatic model of cognitive onomasiology. In H. Cuyckens, R. Dirven, & J. R. Taylor (Eds.), *Cognitive approaches to lexical semantics* (pp. 67–92). De Gruyter Mouton.

[28] Jackson, H. (2002). *Lexicography: An Introduction*. Routledge.

[29] Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *Italian Journal of Linguistics*, *20*(1), 1–31.

[30] Lewandowska-Tomaszczyk, B. (2007). Polysemy, prototypes, and radial categories. In D. Geeraerts & H. Cuyckens (Eds.), *The Oxford handbook of cognitive linguistics*. Oxford University Press.

[31] Lewis, M. (1997). *Implementing the lexical approach: Putting theory into practice*. Language Teaching Publications.

[32] Liu, D. (2010). Going beyond patterns: Involving cognitive analysis in the learning of collocations. *TESOL Quarterly*, *44*(1), 4–30. https://doi.org/10.5054/tq.2010.214046

[33] Liu, D. (2013). Salience and construal in the use of synonymy: A study of two sets of near-synonymous nouns. *Cognitive Linguistics*, *24*(1), 67–113. https://doi.org/10.1515/cog-2013-0003

[34] Mehl, S. (2018). What we talk about when we talk about corpus frequency: The example of polysemous verbs with light and concrete senses.

Corpus Linguistics and Linguistic Theory, *17*(1), 223–247. https://doi.org/10.1515/cllt-2017-0039

[35] Nasiruddin, M. (2013). A state of the art of word sense induction: A way towards word sense disambiguation for under-resourced languages. In F. Boudin & L. Barrault (Eds.), *Proceedings of RECITAL 2013* (pp. 192–205). ATALA.

[36] Navigli, R. (2009). Word sense disambiguation: A survey. *Acm Computing Surveys*, *41*(2), 1–69. https://doi.org/10.1145/1459352.1459355

[37] Nesselhauf, N. (2005). *Collocations in a learner corpus*. John Benjamins.

[38] Ortaçtepe, D. (2013). Formulaic language and conceptual socialization: The route to becoming nativelike in L2. *System*, *41*(3), 852–865. https://doi.org/10.1016/j.system.2013.08.006

[39] Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication*. Longman.

[40] Polguère, A. (2018). A lexicographic approach to the study of copolysemy relations. *Russian Journal of Linguistics*, *22*(4), 788–820. https://doi.org/10.22363/2312-9182-2018-22-4-788-820

[41] Riemer, N. (2005). The semantics of polysemy: Reading meaning in English and Warlpiri. In *The semantics of polysemy*. De Gruyter Mouton.

[42] Rosch, E. H. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive Development and Acquisition of Language* (pp. 111–144). Academic Press. https://doi.org/10.1016/B978-0-12-505850-6.50010-4

[43] Saidi, R., & Jarray, F. (2023). Sentence transformers and DistilBERT for Arabic word sense induction. *Proceedings of the 15th International Conference on Agents and Artificial Intelligence*, *3*, 1020–1027.

[44] Segall, A., & Goldstein, J. (1989). Exploring the correlates of self-provided health care behaviour. *Social Science & Medicine*, *29*(2), 153–161. https://doi.org/10.1016/0277-9536(89)90163-9

[45] Smiskova, H., Verspoor, M., & Lowie, W. (2012). Conventionalized ways of saying things (CWOSTs) and L2 development. *Dutch Journal of Applied Linguistics*, *1*(1), 125–142. https://doi.org/10.1075/dujal.1.1.09smi

[46] Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Blackwell.

[47] Taylor, J. R. (2012). *The mental corpus: How language is represented in the mind*. Oxford University Press.

[48] Tuggy, D. (1993). Ambiguity, polysemy, and vagueness. *Cognitive Linguistics*, *4*(3), 273–290. https://doi.org/10.1515/cogl.1993.4.3.273

[49] Velasco, D. J., Alba, A., Pelagio, T. G., Ramirez, B. A., Chua, U., Samson, B. P., Cruz, J. C. B., & Cheng, C. (2022). *Towards automatic construction of Filipino WordNet: Word sense induction and synset induction using sentence embeddings*. arXiv. https://doi.org/10.48550/arXiv.2204.03251

[50] Walker, C. (2008). Factors which influence the process of collocation. In F. Boers & S. Lindstromberg (Eds.), *Cognitive linguistic approaches to teaching vocabulary and phraseology* (pp. 291–308). De Gruyter Mouton.

[51] Woolard, G. (2000). Collocation: Encouraging learner independence. In M. Lewis (Ed.), *Teaching collocation: Further developments in the lexical approach* (pp. 28–46). Language Teaching Publications.

[52] Wray, A. (2012). What do we (think we) know about formulaic language? An evaluation of the current state of play. *Annual Review of Applied Linguistics*, *32*, 231–254. Cambridge Core. https://doi.org/10.1017/S026719051200013X

[53] Zaytseva, V. (2016). *Vocabulary acquisition in study abroad and formal instruction: An investigation on oral and written lexical development* [Dissertation]. Universitat Pompeu Fabra.

[54] Zhang, Huasheng. (2025a). Automatic discovery of nativelike selections through topic modeling: An experiment based on online reviews . *International Journal of Linguistics, Literature and Translation*, *8*(2), 101-104. https://doi.org/10.32996/ijllt.2025.8.2.11

[55] Zhang, Huasheng. (2025b). Semantic frames as interlingual representations for crosslinguistic comparison: A quantitative analysis of spontaneous texts . International Journal of Linguistics, Literature and Translation, 8(2), 116-120. https://doi.org/10.32996/ijllt.2025.8.2.14