| RESEARCH ARTICLE

Semantic Frames as Interlingual Representations for Crosslinguistic Comparison: a Quantitative Analysis of Spontaneous Texts

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

| ABSTRACT

Recent advancements in frame semantics have shifted toward multilingual applications, proposing semantic frames as interlingual representations or platforms for comparison in contrastive linguistics. Building on these proposals, this study demonstrates the effectiveness and descriptive power of frames and frame systems in crosslinguistic comparison, through a quantitative analysis of customer reviews for a dish soap product in English, Chinese, and Japanese. Departing from traditional reliance on translation data, we utilize these spontaneous texts to uncover different inherent linguistic preferences of frames and lexical choices within a shared conceptual domain.

---

1. Introduction

Pluwak (2021) notes the shift in scientific interest from developing the lexicography project of (English) FrameNet (Fillmore et al., 2003) towards multilingual studies, within the field of frame semantics. In this context, Boas (2005, 2013, 2020) has been advocating for the use of frames as interlingual representations or universal platforms for describing and contrasting multiple languages, beyond just English, since frames are "universal" and not tied to any particular language. Specifically, these studies propose using frames for building multilingual lexical databases, exploring contrastive lexicon fragments, and analyzing translation equivalents, by which (polysemous) meanings and syntactic structures of linguistic expressions can be effectively described. Additionally, Čulo (2013), Czulo (2017) and Czulo et al. (2023) have been emphasizing the descriptive and explanatory power of frame semantics, particularly oriented towards translation studies. Recently, Pluwak (2021) argues that not only individual frames, but also systems of frames and their relations can be used as interlingual representations for comparative studies.

　　Building on these previous studies, this research aims to provide a quantitative case study, supporting the practicality of using frames and frame systems as interlingual representations for crosslinguistic comparison. Translation data has been most commonly used in previous studies. However, instead of focusing on (mis-)matches in translation, this study uses parallel spontaneous texts, which provide more insight into the inherent preferences of different languages within a specific frame system. Multilingual reviews of a dish soap product in English, Japanese, and Chinese serve as the data for this study. This study focuses on how the frame system of Skin_feel helps uncover and present similarities and differences in linguistic usages across these languages.

2. Literature Review
2.1 Frame Semantics and Semantic Frames

The concept of semantic frames, or frames, and the theory of frame semantics (Fillmore, 1982) were initially introduced by Charles Fillmore in the 1970s and 1980s. Currently, frame semantics is a well-established linguistic theory. Frames are schematic representations of situations involving various participants, props, and other conceptual roles (Johnson & Fillmore, 2000). For example, according to the analysis of frame semantics (Fillmore, 1982), the term "weekend" does not merely denote a period of

rest within the week; rather, it reflects a specific frame. Its origin is tied both to the seven-day cycle of the calendar and to a particular socio-economic practice—one in which individuals work for a relatively extended, consecutive period and then reserve the remaining two consecutive days for personal life. If there were only a single day of rest, the term weekend would be unnecessary; people could simply refer to that day by name. Similarly, if the work-rest pattern consisted of three workdays followed by four days off, the period designated for personal life would unlikely be given that name.

Frames have been widely used for analyzing and describing linguistic usages. For example, the Apply_heat frame describes a common situation involving a Cook, some Food, and a Heating_Instrument, and is evoked by words such as *bake*, *boil*, *broil*, and *steam* (Ruppenhofer et al., 2016). In the example of (1), *fry* is the target lexical unit evoking the Apply_heat frame, and roles such as Cook are called frame elements.

(1)   [_{<COOK>}Matilde] *fried*^{Tgt} [_{<FOOD>}the catfish] [_{<HEATING_INSTRUMENT>}in a heavy iron skillet].

A frame system is a collection of interrelated frames, in which the less dependent, or more abstract frame can be called the Super frame and other dependent, or less abstract frames can be called the Sub frames. Get_a_job and Hiring are Sub frames of Begin_employment. The relationships between Super frames and Sub frames can also be further classified into various types, according to Ruppenhofer et al. (2016).

2.2 Frames in Crosslinguistic Comparison

Frame semantics can serve as a universal "interlingua" for intralingual and interlingual descriptions and comparisons (Boas, 2005, 2020; Hasegawa et al., 2011). One reason is that the perspective of frames is from semantics (Baker & Ruppenhofer, 2002), abstracting away the morpho-syntactic differences (Hasegawa et al., 2014). Words or phrases grouped together in the same or related frames might be semantically similar but have different syntactic behaviors. This is particularly useful in crosslinguistic studies since different languages have typological differences in various aspects, but are comparable at the semantic level. Furthermore, Pluwak (2021) argues that frame systems help explain why seemingly unrelated expressions in different languages can become semantic-pragmatic equivalents from a broader perspective.

The most common application of frames in crosslinguistic comparison concerns translation studies, where frames are used for objective evaluation or identification of translation accuracy or divergence (Hasegawa et al., 2014; Ohara, 2020; Ellsworth et al., 2021; Czulo et al., 2023). Such divergence can be termed "frame shift" (Yong et al., 2022). For example, the verb-centered English phrase *Parkinson's disease can attack those younger than 40* might be translated into a noun-centered Japanese phrase such as *those who are under 40 can have symptoms of Parkinson's disease*, where the Attack frame is translated into the Getting_disease frame (Hasegawa et al., 2014). The process of translation can be seen as aiming to achieve maximum frame-to-frame comparability between the original and the translation (Czulo, 2017).

Crosslinguistic applications of frames have also uncovered language-specific frames and lexical items (Schmidt, 2009; Bertoldi & Chishman, 2012; VanNoy, 2017). For example, Bertoldi & Chishman (2012) revealed that, due to cultural differences, the Arraignment frame in American English legal frames does not have direct equivalents in Brazilian Portuguese legal systems. In addition, the comparison of specific frames and frame systems, and their conceptualizations in different languages has also been studied, such as the Risk frame (Ohara, 2009) and the frame systems in lease contracts (Pluwak, 2021).

Pluwak (2021) argues for the use of parallel spontaneous texts to analyze frame systems, instead of using translation data commonly employed in previous studies, because the former is not influenced by translation strategies like direct translation, where translators aim to maintain the syntactic and morpho-syntactic structure of the original text. However, Pluwak (2021) is a qualitative analysis of frames and does not focus on the corresponding lexical choices. This study adopts spontaneous texts for analysis but analyzes frame and word usage quantitatively.

3. Materials and Methods

This study used spontaneous texts of online customer reviews of the same product. Customer reviews of a dish soap product were collected from three online marketplaces: Amazon.co.uk (United Kingdom), JD.com (China), and Amazon.co.jp (Japan). The languages under comparison are English, Chinese, and Japanese.

Following Pluwak (2021), domain-specific frames focusing on how customers evaluate the skin feel of this dish soap product were designed, rather than using FrameNet frames as given. Under the frame system of Skin_feel, 5 Sub frames were summarized from the data: Negative_product_impact, Positive_product_impact, Negative_skin_change, Positive_skin_change, and Customer_reaction_to_product. Frame elements in this frame system include Product, Skin, and Customer. The English examples of each Sub frame are given in (2) – (6).

Note that in this study, the analysis of frames differs from FrameNet (Ruppenhofer et al., 2016: 15), in its treatment of transitive and intransitive usages (namely the causative-inchoative alteration: *the hand dries* vs. *dry the hand*), which are treated as the same frame in this study. The reason is that, semantics rather than syntax should be valued in frame semantics, as discussed above. Causative-inchoative alterations can be seen as different "profilings" of the same event in cognitive grammar (Langacker, 1991). Additionally, change and impact are differentiated in these 5 Sub frames, which have intrinsic differences. For instance, change-of-state verbs emphasize the result, tending to have causative-inchoative alternations (Haspelmath, 1993; Alexiadou, 2015). This

also supports the treatment of transitive and intransitive usages in the present analysis.
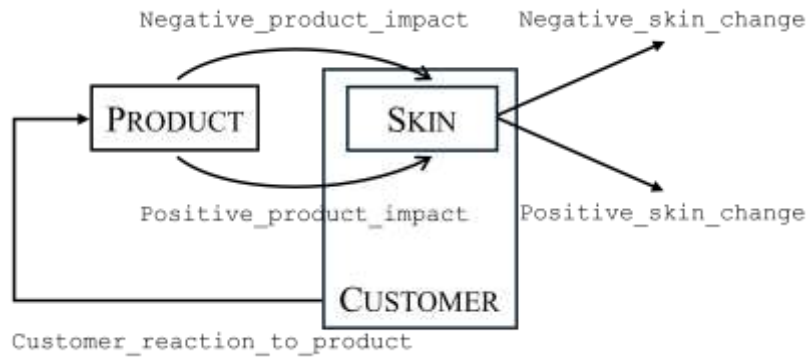

Figure 1: The schema for the Skin_feel frame system.

(2) Negative_product_impact: It is the only [<PRODUCT>washing up liquid] that does not *irritate*[Tgt] [<SKIN>my skin].

(3) Positive_product_impact: [<CUSTOMER>We] have [<SKIN>sensitive skin] and [<PRODUCT>this] is so *good*[Tgt].

(4) Negative_skin_change: [<SKIN>My hands] are not *dry*[Tgt] anymore.

(5) Positive_skin_change: [<SKIN>Hands] are *okay*[Tgt] with [<PRODUCT>this product].

(6) Customer_reaction_to_product: [<CUSTOMER>I] personally *recommend*[Tgt] [<PRODUCT>this] if you have [<SKIN>sensitive hands].

4 Result and Discussion

From the collected data, 100 random instances of target lexical units (LUs) evoking Skin_feel and corresponding Sub frames were identified for each language, through careful manual inspection. The results are summarized in Figure 2.
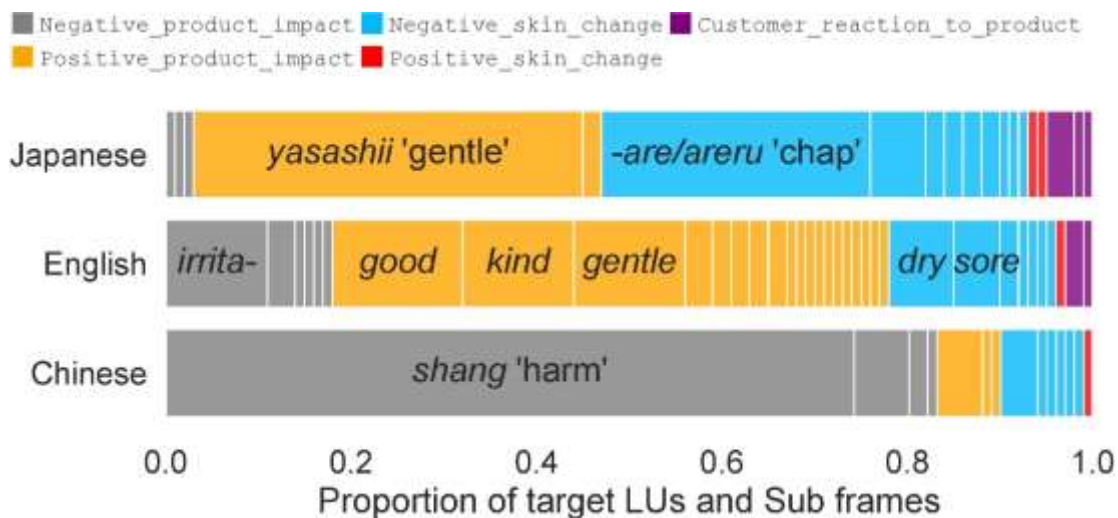

Figure 2: Distribution of target LUs and corresponding Sub frames of Skin_feel in the dish soap customer reviews. The proportion of each LU is displayed in the figure, with only the representative LUs labeled.

This crosslinguistic analysis employing frame semantics reveals both similarities and, more saliently, differences in the preferences of these three languages. The most prominent difference lies in their distinct preferences for Sub frames when evaluating the skin feel of the dish soap. Chinese strongly favors Negative_product_impact, which is less prevalent in English and particularly rare in Japanese. In contrast, English mostly favors Positive_product_impact, while Japanese exhibits a preference for both Positive_product_impact and Negative_skin_change. Note that these preference differences should be attributed to internal factors within the language systems, rather than extralinguistic real-world factors. The results do not follow the assumption that the frequent appearance of a particular frame in linguistic usage might be due to its salience in perception or its frequent occurrence in the real world (see the case of *protect* in Bosque [2011] and Kjellmer [1991: 114] for examples of real-world factors). This is because, despite similar socio-cultural contexts in these parallel texts, these languages exhibit significant differences in their linguistic preferences.

Moreover, there are notable differences in the specific lexical units preferred by each language (within the same Sub frame). Chinese favors *shang* 'harm'; English favors *good*, *kind* and *gentle*; Japanese favors *yasashii* 'gentle' and *-are/areru* 'chap'. Illustrative

examples include *bu shang shou* 'does not harm hands', *kind to skin* and *te ni yasashii* 'gentle to hands'. Liu (2010) highlights the challenges in cross-linguistic comparison of linguistic usages, in establishing equivalent word pairs across languages (e.g., it is hard to determine a single translation equivalent for Japanese *yasashii* among words like *gentle*, *mild*, *kind,* and *good*). However, this challenge can be addressed by employing frames as a universal semantic platform, where words in different frames inherently differ from those within the same frame. For instance, to determine whether *-are/areru* 'chap' is more favored in Japanese compared to its semantic counterparts in Chinese and English, one can compare it with expressions in Negative_skin_change like *dry* and *sore*. As a result, although *-are/areru* 'chap' may correspond to *dry*, it occurs significantly more frequently than the latter.

The frame usages in these different languages also exhibit certain similarities. Positive_skin_change and Customer_reaction_to_product are often neglected in the customer reviews when compared to other Sub frames. Moreover, in each Sub frame, the frequencies of target lexical units vary, with certain dominant words exhibiting significantly higher frequency, following a Zipfian distribution.

Methodologically, frames offer a practical platform to uncover, organize, and present the similarities and differences in linguistic usages across different languages. The linguistic items examined in this study focus on how customers describe the harmlessness of this dish soap product on their hand skins. These expressions are seemingly similar pragmatically, but their differences become evident when analyzed within the context of the distinct Sub frames they evoke. Comparing expressions of different languages is not a straightforward task; however, synonymous expressions within the same domain can be systematically collected and analyzed using frame semantics.

5. Conclusion

This study highlights the effectiveness and descriptive power of frames and frame systems as a powerful platform for systematic crosslinguistic analysis and interpretation. By examining customer reviews of a dish soap product across three languages—English, Chinese, and Japanese—this study has demonstrated that the frame system of Skin_feel provides valuable insights into their distinct preference for frames and lexical choices within a shared conceptual domain. Methodologically, the use of spontaneous texts, as opposed to translation data, elucidates the inherent linguistic preferences of these languages.

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] Alexiadou, A. (2015). The causative alternation revisited: Constraints and variation. *English Linguistics*, *32*(1), 1–21. https://doi.org/10.9793/elsj.32.1_1

[2] Baker, C. F., & Ruppenhofer, J. (2002). FrameNet's frames vs. Levin's verb classes. In J. Larson & M. Paster (Eds.), *Proceedings of the 28th annual meeting of the Berkeley linguistics society* (pp. 27–38). Berkeley Linguistics Society.

[3] Bertoldi, A., & Chishman, R. (2012). Frame Semantics and legal corpora annotation: Theoretical and applied challenges. *Linguistic Issues in Language Technology*, *7*. https://doi.org/10.33011/lilt.v7i.1277

[4] Boas, H. C. (2005). Semantic frames as interlingual representations for multilingual lexical databases. *International Journal of Lexicography*, *18*(4), 445–478. https://doi.org/10.1093/ijl/eci043

[5] Boas, H. C. (2013). Frame Semantics and translation. In A. Rojo & I. Ibarretxe-Antuñano (Eds.), *Cognitive Linguistics and Translation: Advances in Some Theoretical Models and Applications* (pp. 125–158). De Gruyter Mouton.

[6] Boas, H. C. (2020). A roadmap towards determining the universal status of semantic frames. In R. Enghels, B. Defrancq, & M. Jansegers (Eds.), *New approaches to contrastive linguistics: Empirical and methodological challenges* (pp. 21–52). De Gruyter Mouton.

[7] 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.

[8] Čulo, O. (2013). Constructions-and-frames analysis of translations: The interplay of syntax and semantics in translations between English and German. *Constructions and Frames*, *5*(2), 143–167. https://doi.org/10.1075/cf.5.2.02cul

[9] Czulo, O. (2017). Aspects of a primacy of frame model of translation. In S. Hansen-Schirra, O. Czulo, & S. Hofmann (Eds.), *Empirical Modelling of Translation and Interpreting* (pp. 465–490). Language Science Press.

[10] Czulo, O., Willich, A., Ziem, A., & Torrent, T. T. (2023). A multilingual approach to the interaction between frames and constructions: Towards a joint framework and methodology. *Constructions and Frames*, *15*(1), 59–90. https://doi.org/10.1075/cf.00067.czu

[11] Ellsworth, M., Baker, C., & Petruck, M. R. L. (2021). FrameNet and typology. In E. Vylomova, E. Salesky, S. Mielke, G. Lapesa, R. Kumar, H. Hammarström, I. Vulić, A. Korhonen, R. Reichart, E. M. Ponti, & R. Cotterell (Eds.), *Proceedings of the third workshop on computational typology and multilingual NLP* (pp. 61–66). Association for Computational Linguistics.

[12] Fillmore, C. J. (1982). Frame semantics. In Linguistic Society of Korea (Ed.), *Linguistics in the morning calm* (pp. 111–137). Hanshin.

[13] Fillmore, C. J., Johnson, C. R., & Petruck, M. R. L. (2003). Background to framenet. *International Journal of Lexicography*, *16*(3), 235–250. https://doi.org/10.1093/ijl/16.3.235

[14] Hasegawa, Y., Lee-Goldman, R., & Fillmore, C. J. (2014). On the universality of frames: Evidence from English-to-Japanese translation. *Constructions and Frames*, *6*(2), 170–201. https://doi.org/10.1075/cf.6.2.03has

[15] Hasegawa, Y., Lee-Goldman, R., Kong, A., & Kimi, A. (2011). FrameNet as a resource for paraphrase research. *Constructions and Frames*, *3*(1),

104–127. https://doi.org/10.1075/cf.3.1.04has

[16] Haspelmath, M. (1993). More on the typology of inchoative/causative verb alternations. In B. Comrie & M. Polinsky (Eds.), *Causatives and transitivity* (pp. 87–120). John Benjamins.

[17] Johnson, C., & Fillmore, C. J. (2000). The FrameNet tagset for frame-semantic and syntactic coding of predicate-argument structure. *Proceedings of the 1st North American Chapter of the Association for Computational Linguistics Conference*, 56–62.

[18] Kjellmer, G. (1991). A mint of phrases. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics: Studies in honour of Jan Svartvik* (pp. 111–127). Longman.

[19] Langacker, R. W. (1991). *Foundations of cognitive grammar: Vol. II: Descriptive application*. Stanford University Press.

[20] 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

[21] Ohara, K. H. (2009). Frame-based contrastive lexical semantics in Japanese FrameNet: The case of risk and kakeru. In H. C. Boas (Ed.), *Methods and Applications* (pp. 163–182). De Gruyter Mouton.

[22] Ohara, K. H. (2020). Finding corresponding constructions in English and Japanese in a TED talk parallel corpus using frames-and-constructions analysis. In T. T. Torrent, C. F. Baker, O. Czulo, K. Ohara, & M. R. L. Petruck (Eds.), *Proceedings of the International FrameNet Workshop 2020: Towards a Global, Multilingual FrameNet* (pp. 8–12). European Language Resources Association.

[23] Pluwak, A. (2021). The frame system as an interlingual representation for parallel texts. *Intercultural Pragmatics*, *18*(5), 657–685. https://doi.org/10.1515/ip-2021-5004

[24] Ruppenhofer, J., Ellsworth, M., Schwarzer-Petruck, M., Johnson, C. R., & Scheffczyk, J. (2016). *FrameNet II: Extended theory and practice*. International Computer Science Institute.

[25] Schmidt, T. (2009). The Kicktionary – a multilingual lexical resource of football language. In H. C. Boas (Ed.), *Multilingual FrameNets in computational lexicography: Methods and applications* (pp. 101–134). De Gruyter Mouton.

[26] VanNoy, A. (2017). *Culture specific aspects of semantic frames in multilingual frame descriptions* [Dissertation]. University of Texas at Austin.

[27] Yong, Z. X., Watson, P. D., Timponi Torrent, T., Czulo, O., & Baker, C. (2022). Frame shift prediction. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the thirteenth language resources and evaluation conference* (pp. 976–986). European Language Resources Association.