Language and Cognition (2024), 1–23
doi:10.1017/langcog.2024.6

ARTICLE

Contrasting the semantic space of 'shame' and 'guilt' in English and Japanese

Eugenia Diegoli[1] and Emily Öhman[2]

[1]Department of Interpreting and Translation, University of Bologna, Bologna, Italy
[2]Faculty of International Research and Education, School of International Liberal Studies, Waseda University, Tokyo, Japan
Corresponding author: Eugenia Diegoli; Email: eugenia.diegoli2@unibo.it

(Received 31 May 2023; Revised 24 January 2024; Accepted 28 January 2024)

Abstract
This article sheds light on the significant yet nuanced roles of shame and guilt in influencing moral behaviour, a phenomenon that became particularly prominent during the COVID-19 pandemic with the community's heightened desire to be seen as moral. These emotions are central to human interactions, and the question of how they are conveyed linguistically is a vast and important one. Our study contributes to this area by analysing the discourses around shame and guilt in English and Japanese online forums, focusing on the terms *shame*, *guilt*, *haji* ('shame') and *zaiakukan* ('guilt'). We utilise a mix of corpus-based methods and natural language processing tools, including word embeddings, to examine the contexts of these emotion terms and identify semantically similar expressions. Our findings indicate both overlaps and distinct differences in the semantic landscapes of shame and guilt within and across the two languages, highlighting nuanced ways in which these emotions are expressed and distinguished. This investigation provides insights into the complex dynamics between emotion words and the internal states they denote, suggesting avenues for further research in this linguistically rich area.

Keywords: corpus-based contrastive analysis; corpus linguistics; emotion detection; guilt; language of emotions; moral behaviour; natural language processing; semantic similarity; shame; word embeddings

1. Introduction

The prominent role of shame and guilt in contemporary societies became apparent during the COVID-19 pandemic. The desire to be seen as moral personae, and the shame and guilt resulting from the failure to be perceived as such, presumably played a key role in fostering moral behaviours. This, however, is by no means a novel phenomenon. Long before COVID-19, shame and guilt have been understood as modifiers of behaviour (Bedford & Hwang, 2003), markers of personal identity (Hultberg, 1988), and mechanisms for social control across cultures (Creighton,

1990). Yet, their effects on the emoter[1] (i.e., whoever experiences the emotion; Bednarek, 2008, p. 14; Glynn, 2014, p. 72) are controversial. In the field of psychology, it is now widely acknowledged that shame and guilt are central in fostering socially responsible behaviours or avoidance of behaviours that may lead to disapproval (Sabiston & Castonguay, 2014). However, some studies positively correlate shame and guilt with a variety of negative behavioural, psychological, and physical outcomes such as depression and anxiety (Cavalera, 2020; Sabiston & Castonguay, 2014, p. 626). The degree to which these aspects of shame and guilt are shared across cultures is also controversial, as they have been categorised differently by different scholars – even within the same field, including linguistics and psychology. Unlike anger, surprise, disgust, enjoyment, fear and sadness, which have traditionally been considered to be primary emotions shared by all humans (e.g., Ekman, 1992),[2] secondary emotions are learnt through socialisation, and are, therefore, culture-specific (Wierzbicka, 1999). In this study, we subscribe to the view that shame and guilt are secondary emotions. The inherently culture-dependent nature of shame and guilt, and the actual behaviours associated with them, have been examined across cultures from a psychological perspective (Arimitsu, 2001; Sheikh & Janoff-Bulman, 2010; Smith et al., 2002; Suzuki, 2007; Tangney & Dearing, 2002). Linguists have also tackled these phenomena, focusing on the linguistic structures that give us access to shame and guilt as emotion concepts and their variability across languages (e.g., Fabiszak & Hebda, 2007; Krawczak, 2017, 2018; Kumamoto, 2019; Tissari, 2006). The underlying theoretical framework in these studies and ours is that language is usage-based. That is, the idea that linguistic knowledge is shaped by the context and frequency of language use, thus the language of shame and guilt is also 'usage-based' (e.g., Fabiszak et al., 2016; Geeraerts, 2010; Glynn, 2007, 2010a, 2010b; Langacker, 1987; Vigliocco et al., 2009, p. 222), and can be better understood by exploring the different contexts in which expressions of shame and guilt are present.

This study adds to the existing literature on the topic by proposing a linguistic approach to shame and guilt in two typologically and pragmatically different languages, namely Japanese and English. It does so by asking how these two emotions are metadiscursively framed in two online forums that are relatively similar in terms of audience, aim, and structure – a factor that increases comparability. Hence, the main focus is on only one aspect of the highly complex phenomena of shame and guilt, namely how people *talk about* them. By using computational tools from the field of natural language processing (NLP), we extract data in novel ways and further add to existing methodologies in corpus linguistics. Our aim is to further contribute to existing theories of emotions and the language through which such emotions are verbalised by extracting new insights with novel approaches and comparing the

[1]It is important to note our use of the term 'emoter', used in the field of cognitive linguistics by Glynn (2014). While this term effectively captures the agent experiencing or expressing an emotion in our analysis, we acknowledge that it is not widely utilised in the broader linguistic community. Our choice to use 'emoter' aligns with Glynn's conceptual framework but may differ from more conventional terminology in emotional studies.

[2]The existence of universal emotions has been questioned in more recent research (Barrett et al., 2011; Jack et al., 2012) as well as by Ekman's contemporaries (e.g., Mead, 1972). However, proponents still exist (Keltner et al., 2019). In subsequent work, Ekman (1992) himself raises the possibility that *shame* and *guilt* could also have universal characteristics. This model is further developed, most notably by Plutchik (1980), who adds to the list of basic emotions *trust* and *anticipation* – but these additions are more controversial.

https://doi.org/10.1017/langcog.2024.6 Published online by Cambridge University Press

results to previous findings. We aim to enhance academic perspectives on emotions by focusing on how people discuss them. This approach is based on the premise that overlooking people's conceptualisations could lead to the creation of analytical artifacts. These artifacts might lack real-world relevance and fail to accurately represent the experiences of those involved in the interaction.

2. 'Shame' and 'guilt': what they are and what they do

This section presents a working definition of 'shame' and 'guilt', in single quotation marks when we want to indicate that they are not the English words *shame* and *guilt* actually used in interaction, but more general emotional labels that are shared, to some degree, across languages and cultures. It then presents the link between (moral) emotions and evaluation. Finally, it reviews some important studies that tackle the language of emotions across linguacultures.

2.1. A working definition

Drawing from Tangney and Dearing (2002, p. 25), we tentatively define 'shame' and 'guilt' as negatively valenced moral emotions typically experienced in interpersonal contexts. They are *negatively valenced* because they imply a negative evaluation by others of one's conduct/identity, and *moral* because they guide and are guided by our sense of good and bad. Since 'shame' and 'guilt' are not predetermined genetically, but learnt, negotiated, and eventually challenged through socialisation, it is reasonable to assume that their conceptualisations may differ across linguacultures (Bedford & Hwang, 2003).

2.2. Emotions and evaluation

The presumed causal link between 'shame' and 'guilt' and a higher moral order (i.e., 'a culture-specific ideology about what counts as right or wrong' (Culpeper & Tantucci, 2021, p. 148) is also evident in their characterisation as *moral* emotions that 'provide immediate punishment (or reinforcement) of behavior' (Tangney & Dearing, 2002, p. 133). This definition makes apparent the evaluative nature of moral emotions. On this basis, and drawing from the evaluative tradition on emotions (Scarantino, 2016), 'shame' and 'guilt' – and other valenced emotions – are here conceptualised as forms of *evaluation*, which in turn is intended in a sense described by Hunston and Thompson (2000, p. 5) as the indication that something is good or bad. In the case of 'shame' and 'guilt,' the latter applies: 'shame' and 'guilt' are emotions that originate from the producer's awareness of having failed to be or behave in accordance with the standards recognised as proper by the group. They entail a negative judgement of actions or behaviours deemed shameful or guilt-worthy and, by extension, of the individuals responsible for these actions. Looking at (the linguistic manifestation of) emotions as a form of evaluation accommodates the inherent normative aspect of 'shame' and 'guilt'. It is important to note that the methods employed in this study do not give us access to the emotions themselves. Emotions are internal phenomena and, as such, not amenable to direct empirical observation. However, we can use the linguistic manifestations of emotions as a proxy for the actual emotions (Laaksonen et al., 2023).

https://doi.org/10.1017/langcog.2024.6 Published online by Cambridge University Press

2.3. Emotion talk across languages

Bednarek (2008) makes an important distinction between emotion talk and emotional talk.[3] The former indicates the language *about* emotions and 'is constituted by all those expressions in the dictionary that denote affect/emotion, for example, *love*, *hate*, *joy*, *envy*, *sad*, *mad*, *enjoy*, *dislike* and so on'. The latter relates to 'all those constituents (linguistic and non-linguistic) that conventionally express or signal affect/emotion (whether genuinely experienced or not, whether intentional or not)' (Bednarek, 2008, p. 11). This study focuses on the former, i.e., emotion talk, as we look at 'shame' and 'guilt' only insofar as the linguistic expressions that denote the internal states conventionally associated with them are present in the text. Such explicit linguistic resource pertains in appraisal theory (Martin & White, 2005) to 'Judgement' – a sub-system of 'Affect' concerned with 'resources for assessing behaviour according to various normative principles' (Martin & White, 2005, p. 35).

In the definition of Judgement given above, we have again the complex interrelationship of moral emotions, evaluations and norms: emotions are a form of evaluation in that they always encode interactants' point of view towards something. When they are *moral* in nature, evaluations are based on normative standards that a given group or community sees as proper, that is, ideologies (Garfinkel, 1967; Heinrich, 2012; Verschueren, 2011). Just as evaluations in terms of (in)appropriateness can vary from one community to the other, and even among individuals of the same group, also the words conventionally associated with the emotions that are manifestations of such evaluations may acquire different meanings in different contexts. It follows that, even when we have a direct translation of English metalexemes, such as *shame* and *guilt* in other languages, what they conventionally index may differ across linguacultures (Kádár & Haugh, 2013; Kádár & Ran, 2019; Kumamoto, 2019; Soares da Silva, 2020). This, however, is a point often overlooked in the literature, where scientific metalabels (which are almost invariably in English) are rarely problematised. Some notable exceptions are Wierzbicka and Harkins (2001) and Pavlenko (2008), who point out that emotion concepts may not overlap completely in different languages or cultures. In the Japanese context specifically, Imada (1989) demonstrates that the English and the Japanese notions of 'anxiety', 'fear' and 'depression' differ, with *fuan* 'anxiety' being closer to *yūutsu* 'depression' than to *kyōfu* 'fear', while *anxiety* and *fear* are more similar than *anxiety* and *depression* (Imada, 1989, p. 12).

2.4. Computational approaches to shame and guilt detection

In the field of NLP specifically, despite the increasing prevalence of sentiment analysis and emotion detection, 'shame' and 'guilt' are somewhat underexplored. The few studies we did find (e.g., Adoma et al., 2020; Meque et al., 2023) tend to focus on ISEAR (International Survey on Emotion Antecedents and Reactions) as their main source of 'shame' and 'guilt' labels and mainly perform classification tasks on the data, that is, the focus is on the classification task where automatic detection of 'shame' and 'guilt' is attempted, not on how 'shame' and 'guilt' are expressed. In the

[3]This overlaps with Pavlenko's (2008, p. 148) distinction between emotion words and emotion-laden words. We favour Bednarek's (2008) terminology because she adopts a *linguistic* approach to emotions that nicely fits ours. Hence, we take the term *talk* to include written interactive forms of communication like the one explored here.

https://doi.org/10.1017/langcog.2024.6 Published online by Cambridge University Press

adjacent field of corpus linguistics, there is some prior research that employs multivariate corpus methods to investigate cross-linguistic use of 'shame' and 'guilt' (see Krawczak, 2014a, 2014b, 2018; Krawczak & Badio, 2015). The present work builds on these studies but presents a relatively novel methodological perspective (a combination of corpus and NLP methods) and focuses on a different type of data (online written forums).

Other researchers have investigated adjacent moral emotions linked to social control, such as condolence and empathy in online communities (Zhou & Jurgens, 2020) and hope and regret detection (Sidorov et al., 2023). The former concluded that online and in-person engagement with condolences and empathy, in general, were based on quite different social clues, suggesting that posting about struggles online is about seeking positive reinforcement rather than 'comments that require emotional effort to engage with complex emotions' (Zhou & Jurgens, 2020, p. 617). Many studies have also tried to model or detect suicidal tendencies online, which often partly include 'guilt' or 'shame' as components or parameters in the detection model (see, e.g., Guidère, 2020).

Emotion detection, in general, is a highly active research field. From a purely NLP perspective, these approaches mostly aim to improve models in terms of accuracy metrics. Such models are, therefore, almost invariably based on supervised machine learning and tend to work only within a specific domain. However, it has been argued that more real-world congruent results with reusable methods can be obtained by using lexicons either independently (Teodorescu & Mohammad, 2022) or together with data-driven methods (Öhman, 2021). Öhman and Rossi (2023) use emotion lexicons to create affective word embeddings that allow them to create domain-specific models that take semantic shifts into account when attempting to use affect as a proxy for mood in literary texts.

3. Methods and aims

Building on these earlier studies, the present work combines NLP and corpus-based collocational methods to investigate how people talk about, negotiate, and eventually challenge *shame* and *guilt* and *haji* and *zaiakukan* in two online web forums. The working hypothesis is that they may denote slightly different experiences and concepts.

3.1. Why shame and guilt and haji and zaiakukan

The selection of the English search items was quite straightforward because *shame* and *guilt* are discussed at length in the psychology literature on emotions (e.g., Ekman, 1992; Sabiston & Castonguay, 2014; Tangney & Dearing, 2002). The next step was selecting their Japanese translations. The Genius English–Japanese dictionary (6th edition) proposes the following possible translations[4]:

[4]A number of translations that do not directly describe emotional states have been omitted, for example, *tsumi* and *hanzai* for *guilt*, which are closer to (criminal/legal) offence. Also, note the reoccurring character 恥 *haji* 'shame' in the 'shame' column for all words. A case could be made for the cultural centrality of these concepts simply based on the existence or non-existence of a specific character for each.

https://doi.org/10.1017/langcog.2024.6 Published online by Cambridge University Press

Shame	*Guilt*
恥ずかしさ *hazukashisa*	罪悪感 *zaiakukan*
恥ずかしい思い *hazukashii omoi*	自責 *jiseki*
羞恥心 *shūchishin*	後ろめたさ *ushirometasa*
恥 *haji*	

Among these, we selected items that have received some attention in the Japanese literature on emotions (e.g., Higuchi, 2002; Inaba, 2009; Suzuki, 2007), that is, *zaiakukan* for *guilt*, and *haji* and *shūchishin* for *shame.* A search on the web corpus JaTenTen11 revealed that *haji* (96,201) is overwhelmingly more frequent than *shūchishin* (16,794), hence the former was preferred.

Importantly, we understand that experiences of 'shame' and 'guilt' can be verbalised in a multitude of ways that go well beyond the explicit use of emotion talk, let alone two sets of specific nouns. In Section 4.1, we attempt to incorporate as many such expressions as possible into our work by utilising word embeddings. This method allows us to find words and phrases that are closely semantically related.

3.2. Research questions

Our primary research question is (1) What are the main similarities and differences between the experiences verbally labelled as *shame* and *guilt* in English and *haji* and *zaiakukan* in Japanese? We also touch upon the more specific questions: (2) Who feels 'ashamed' and 'guilty' and for what? (3) Do people differentiate between these two experiences?

3.3. The data sources

The English data, amounting to 115,582,531 tokens, come from Reddit (https://www.reddit.com/), a predominantly North American pseudo-anonymous online discussion forum with 52 million regular users. Our data were extracted from the relationship_advice subreddit that centre on the topic of relationships.

The Japanese data are more restricted in size, amounting to 1,137,135 tokens after segmentation. They come from Hatsugen Komachi (https://komachi.yomiuri.co.jp/), which has been operated by the Yomiuri Newspaper (one of the largest newspapers in Japan) since 1999. Hatsugen Komachi, which literally means 'little town of speech', is a forum for *dare ni mo kikenakatta onna no nayami* (lit. 'I couldn't ask anyone'; worries of women') originally addressed exclusively to women. According to Yahoo! Japan, it averages 2,000 posts a day and more than 100 millions monthly page views, and a much broader user base than when it was first launched. The fact that, despite its popularity, it has not received much attention from scholars to date possibly attests to resistance in Japan to approaching online forms of data in a scientifically adequate way (Miyake, 2022). The present study investigates online forums as scientifically adequate sources of linguistic data.

3.4. Tools for the data collection and analysis

Although we focus on *shame* and *guilt* and their Japanese counterparts, these emotions are not always expressed using these specific words. Previous research

https://doi.org/10.1017/langcog.2024.6 Published online by Cambridge University Press

suggests that the related terms *embarrassed* and *ashamed* are also used (Krawczak, 2014b). In Japanese, negative emotions are rarely expressed directly and, similarly to English, *hazukashii* 'embarrassing' is also a frequently used emotion term (Farese, 2016). Consequently, we utilise word embeddings to computationally extract words and expressions that are used in semantically similar contexts to *shame*, *guilt*, *haji*, and *zaiakukan*.

Word embeddings, or semantic vector space models, are shallow neural networks that reconstruct the linguistic context of words as vectors by iterating over a corpus to learn associations between words and mapping semantically similar words to geometrically close embedding vectors (Mikolov et al., 2013). Cosine similarity provides the angle between two vectors and is the most used similarity measure for word similarity calculations (Sidorov et al., 2014). Such approaches have been previously used within a usage-based linguistic framework (specifically Pankratz & Van Tiel, 2021). We use the collected data to build language- and context-specific vector space representations and examine which words and expressions are semantically closest to the keywords by using cosine similarity measures and word ngrams. This allows us to find many more examples of how 'shame' and 'guilt' have been expressed beyond the words themselves, including periphrastic expressions. We also employ manual evaluations of the results to ensure they are robust and not random (Antoniak & Mimno, 2018; Pierrejean & Tanguy, 2018).

The orthography of the Japanese language presents additional challenges as NLP applications rely on tokenisation and lemmatisation of words (i.e., splitting up the text into word units in their base form), a process which is fairly simplistic and achieves near-perfect accuracies in English with the tools that exist today. However, as there are no spaces separating word-like units, Japanese texts first need to be segmented into 'words' – a concept that is hard to define in any language but can be largely ignored in languages that have rules for where to place break spaces (Grefenstette & Tapanainen, 1994; Papandropoulou & Sinclair, 1974). Many different segmentation tools exist, but they all output slightly different segments based on different logic, making comparisons between lexical items a challenge. Issues with segmentation also affect the data at hand in that when *zaiakukan* 罪悪感 'guilt' is present in a text, it is split into *zaiaku* 罪悪 'crime' and *kan* 感 'feeling' (cf. Figure 3 and Table 2). However, in our sample, *zaiaku* (n = 24) appears almost exclusively as part of *zaiakukan* (n = 23). This justifies the choice to focus on *zaiaku* for the word embeddings (considering the segmentation issues mentioned earlier), whilst the concordance analysis illustrated in Section 4.2.6 examines *zaiakukan*.

4. Results

4.1. Vector space representations

Figures 1 and 2 show a map of the semantic vector space of the seed words (*shame*, *guilt*, *embarrassment*, *regret*, *remorse*) and their most semantically similar words as measured by cosine similarity. *Embarrassment*, *regret* and *remorse* were added as separate keywords after looking at the most similar words of *shame* and *guilt*. This approach suggests that the semantically most similar expressions to *shame* are related to *disgrace*, *dishonor* and *embarrassment*, whereas *guilt*-related expressions are more closely related to *fault*, *conscience*, *culpability*, (*sincere*) *remorse*, but also *grief* and *sorrow*.

https://doi.org/10.1017/langcog.2024.6 Published online by Cambridge University Press

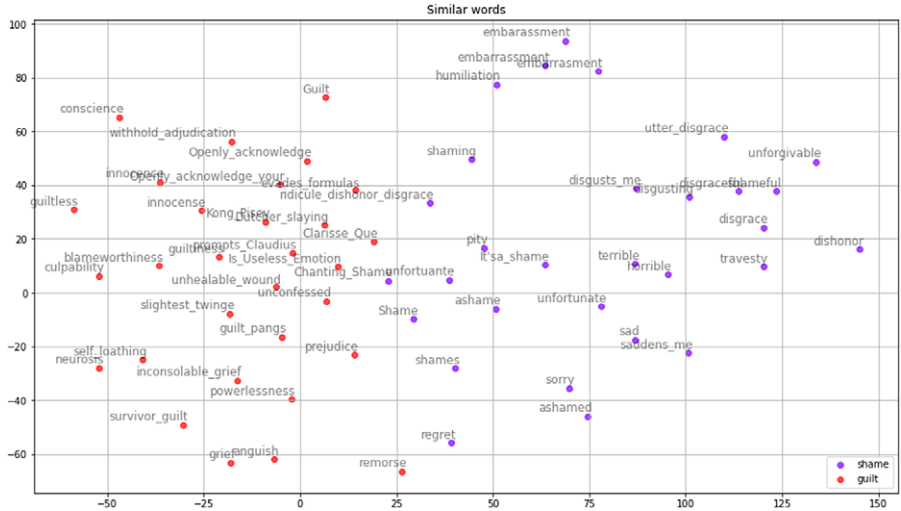

Figure 1. Words semantically similar to *shame* and *guilt* in English.

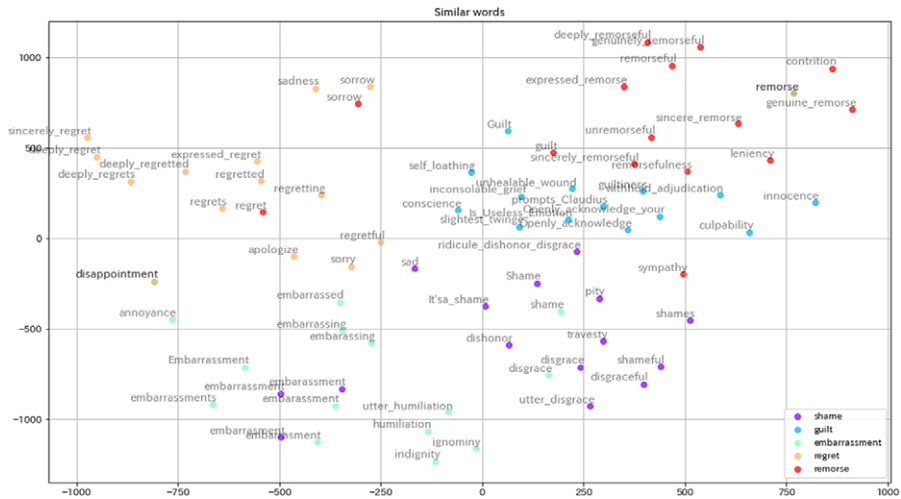

Figure 2. Words semantically similar to *shame*, *guilt*, *embarrassment*, *regret* and *remorse* in English.

Figure 3 shows a map of the semantic vector space of the target words in Japanese. Loss of face *(oime)*, distrust *(fushin)*, and regret (*ushirometa**) are linked to *zaiaku* 罪悪, while embarrassment (*hazukashii*) and dishonour (*akahaji*) are more related to *haji* 恥.

Tables 1 and 2 refer to Figures 2 and 3, respectively. Note that they were pruned to improve legibility by excluding spelling variations (e.g., embarasing/embarrasing/embarrassing/embarassing) and different verbs (*kaka* かか, *kakasu* かかす, *kaku-kara* かくから, *kakisute* かきすて, etc.) and adjectives (*hazu* 恥ず, *hazukashii* 恥ずかしい) forms. Some English compound terms (e.g., *deeply_regrets*, *deeply_regretted*, *deeply_regret*, *expressed_regret*), Japanese morphemes with no clear meaning on their

https://doi.org/10.1017/langcog.2024.6 Published online by Cambridge University Press

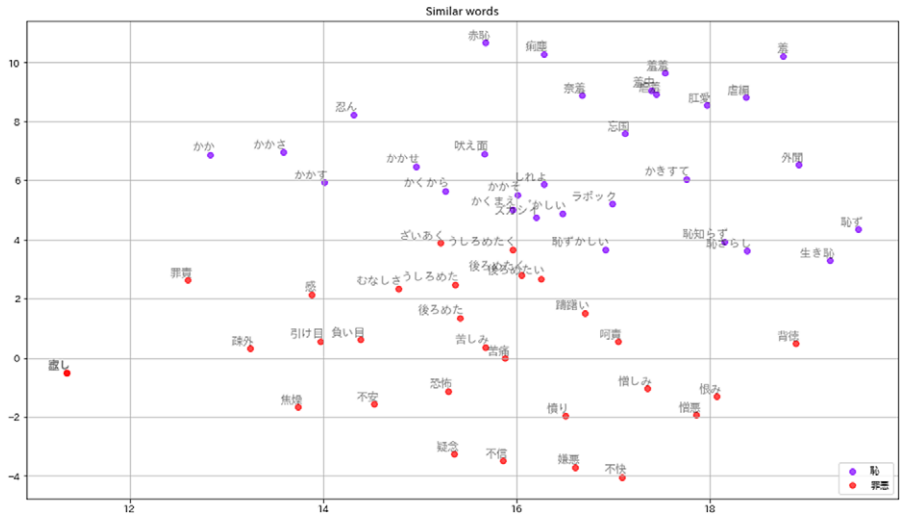

Figure 3. Words semantically similar to *haji* and *zaiaku* in Japanese.

Table 1. The semantically closest matches to keywords in English

shame	disgrace disgraceful	pity dishonor	embarrassment shameful	travesty unforgivable	sad
guilt	unhealable_wound inconsolable_grief	culpability innocence	conscience self_loathing	Is_Useless_Emotion Openly_acknowledge	
embarrassment	humiliation disgrace	ignominy annoyance	shame indignity		
regret	sorrow sorry	sadness remorse	disappointment apologize		
remorse	contrition unremorseful	leniency sorrow	guilt sympathy	regret moral_culpability	sympathy remorseful

own (*sukashi* スカシ, kashii かしい *rabokku* ラボック, *i* イ) and idiosyncratic uses employed, for example, as part of *manga* or TV shows titles (*kōai* 肛愛, *shūchū* 羞中, etc.) were also excluded from the tables but can be seen in the vector space visualisations.

Word embeddings are complemented by the corpus-based analysis of texts, which was carried out with the tools offered by Lancsbox (Brezina et al., 2020), a recently developed software package for the analysis of language data and corpora. We used collocation analysis to access meanings that recur across different pieces of text. In the attempt to access 'non-obvious general semantic preferences' (Partington, 2004, p. 164) and picture what the semantic space of 'shame' and 'guilt' may look like in the two linguacultures, the analysis is not limited to first-order collocates (i.e., the collocates of our search items), but we have further travelled 'the collocational network' (Marchi, 2023). When relevant, we accessed the extended concordance lines to zoom in on specific and often highly context-dependent linguistic

https://doi.org/10.1017/langcog.2024.6 Published online by Cambridge University Press

https://doi.org/10.1017/langcog.2024.6 Published online by Cambridge University Press

Table 2. The semantically closest matches to keywords in Japanese

恥 *haji* shame	赤恥 *akahaji* dishonour 恥ずかしい *hazukashii* embarrassing	羞中 *shūchū* shame 生き恥 *ikihaji* live in shame	痴塵 *rijin* disgrace しれよ *shire yo* shame on you	かか* *kaka** humiliate 忍ん *nin* patience	外聞 *gaibun* reputation 恥知らず *haji shirazu* shameless	羞羞 *shūshū* shame 恥さらし *haji sarashi* disgrace oneself	虐羞 *gyakushū* shame	吠え面 *hoetsura* regret	
罪悪 *zaiaku* guilt	後ろめた* *ushirometa** regret 焦燥 *shōsō* irritation 疎外 *sogai* alienation	負い目 *oime* loss of face 躊躇い *tamerai* hesitation 不安 *fuan* worry	嫌悪 *keno* hatred むなしさ *munashisa* vanity 不快 *fukai* unpleasant	不信 *fushin* distrust 罪責 *zaiseki* responsibility 憎しみ *nikushimi* hatred	憤り *ikidoori* resentment 引け目 *hikeme* inferiority 憎み *nikumi* hate	虚し *munashi* useless 苦しみ *kurushimi* pain 憎悪 *zouo* hatred	恐怖 *kyōfu* fear 苦痛 *kutsū* pain 寂し *sabishi* sad	感 *kan* feeling 呵責 *kashaku* conscience	疑念 *ginen* doubt 背徳 *haitoku* immorality

constructions. The findings shed light on some frequent and meaningful patterns that collocate with *shame*/*haji*(-related) and *guilt*/*zaiakukan*(-related) expressions.

4.2. Collocational and concordance analysis

A preliminary examination of the data using corpus tools indicated a notable difference in the frequency of the search words. For instance, in the English corpus, occurrences of *guilt* (9,894) were more numerous than those of *shame* (6,363). Contrastingly, in the Japanese corpus, *haji* (representing 'shame') appeared 560 times, surpassing the occurrences of *zaiakukan* (representing 'guilt'), which appeared only 23 times. However, these figures should be approached with caution. Raw frequency counts do not necessarily equate to linguistic or cultural prominence and can be influenced by various factors unrelated to the emotional salience of these terms. Additionally, our findings seem to diverge from external sources such as the COCA database, where *shame* appears more frequently than *guilt.* This discrepancy highlights the importance of considering multiple data sources and methodologies when examining linguistic phenomena. Furthermore, the relevance of these occurrences in relation to the emotional states they represent must be carefully considered. Not all instances of *shame* and *guilt* in the English corpus may pertain directly to the emotional states. Therefore, a more in-depth analysis that discerns the context of each occurrence is necessary to draw meaningful conclusions about the emotional landscape in each language.

Concerning the Japanese data, while *haji* appears more frequently than *zaiakukan*, this result alone is insufficient to conclude definitively about the emotional landscape in the Japanese linguaculture. It is a preliminary observation that suggests a potential avenue for further research rather than a conclusive statement. With these caveats in mind, the following sections investigate and compare the collocational network of our search words in the two samples.

4.2.1. Semantic space of shame and guilt

The semantic space of *shame* and *guilt* as verbalised by their collocates, and the way they relate to each other in our sample, is illustrated in Figure 4. Drawing from Sinclair, we define collocates as 'the occurrence of two or more words within a short space of each other in a text' (1991, p. 170). For the scope of the present paper, such short space corresponds to five words on each side of the node. The statistical measure employed for visualising the results is logDice, which indicates the tendency of two words to occur exclusively in each other's company (Brezina, 2018). This measure is not affected by the size of the corpus, hence can be used to compare co-occurrence across corpora, and is particularly useful to highlight the cumulative forces of discourse representations (Brookes & Chałupnik, 2022, p. 4). We included collocates with a logDice of at least 7, a score that returns statistically significant results (Brezina, 2018) and prevents over-populated graphs.

In the figure, the closer the collocate is to the node word, the stronger their association. The frequency is indicated by the intensity of the colour of the collocate, and the position of the collocate in the semantic space is the actual position where it appears in the texts, either at the right or the left of the word, or a mix of the two. Tables 3 and 4 report the values associated with each collocate in more detail.

https://doi.org/10.1017/langcog.2024.6 Published online by Cambridge University Press

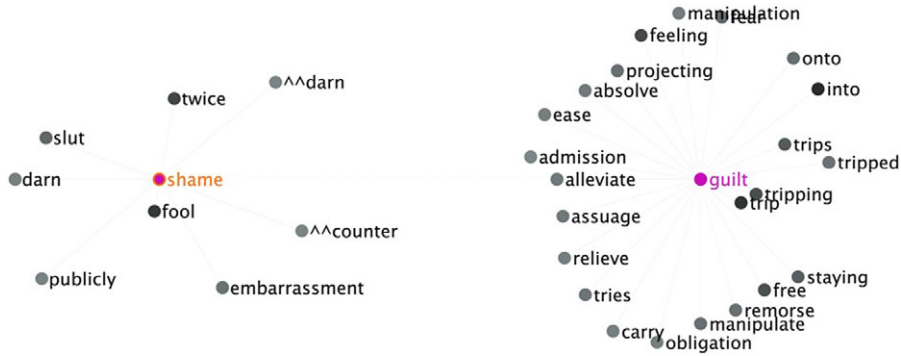

Figure 4. L5R5 collocates of *shame* and *guilt* in the Reddit corpus. LogDice value cut-off <7.

Shame and *guilt* are polysemous words, hence Figure 4 and Tables 3 and 4 include items that are not necessarily related to 'shame' and 'guilt' as emotions. We tracked down the contextual meanings of the collocates by zooming in on the text. In what follows, we focus on those collocates that the close reading of concordances revealed to be relevant to *shame* and *guilt* as emotion terms.

Among the five most typical collocates of *guilt*, we have *shame* and *free*, at the third and fifth position, respectively, while the remaining three are different forms of the lemma *trip. Free* is employed in the idiomatic construction *guilt-free*, which is not as relevant as other patterns for the scope of the paper, whilst the co-occurrence of *shame* and *guilt* is further addressed below. We shall now turn to *trip** in combination with *guilt.* The most typical collocation is *guilt trip* (logDice 10.1, freq. 735) in the construction SBJ *guilt trip* OBJ, as in the following example (all examples are reproduced faithfully to the original, including non-standard spellings, punctuation and so on):

(1) He says you guilt trip him, this is a fucking mess

The second most typical collocation is *guilt tripping* (logDice 9.9, freq. 299), followed by *guilt trips* (logDice 9.1, freq. 197). These values testify to the saliency that this construction and the phenomenon it indexes have among the English-speaking users of Reddit, who conceptualise 'guilt' as something that can be forced on others. Notably, previous studies argued exactly the opposite, saying that while 'shame'

Table 3. L5R5 collocates of *shame* in the Reddit corpus (logDice ≥ 7)

Collocate	LogDice	Raw frequency
fool	10.5	431
guilt	9.4	337
twice	9.3	289
slut	8.3	93
embarrassment	8.2	67
darn	8.2	44
^^counter	7.6	38
^^darn	7.6	38
publicly	7.5	41
fear	7.0	79
there's	7.0	239

https://doi.org/10.1017/langcog.2024.6 Published online by Cambridge University Press

Table 4. L5R5 collocates of *guilt* in the Reddit corpus (logDice ≥ 7)

Collocate	LogDice	Raw frequency
trip	10.1	735
tripping	9.9	299
shame	9.4	337
trips	9.1	197
free	8.2	285
tripped	8.1	86
remorse	8.0	91
projecting	8.0	96
staying	7.9	179
assuage	7.8	69
alleviate	7.8	70
manipulate	7.8	96
absolve	7.7	66
into	7.7	973
onto	7.6	99
feeling	7.5	363
relieve	7.4	56
tries	7.3	80
fear	7.3	115
manipulation	7.2	61
obligation	7.2	56
ease	7.2	57
admission	7.1	44
carry	7.0	56

can be imposed on us by others because it is based on a socially constructed identity, 'guilt', which originates in issues of responsibility, cannot (Bedford & Hwang, 2003, p. 128). Our findings seem to suggest that there may be a gap between academic notions of 'guilt' and how it is conceptualised by the layperson.

As Table 3 shows, *shame* has a more diversified list of collocates, at least if we look at the five most typical items, which are *fool*, *guilt*, *twice*, *slut* and *embarrassment*. *Embarrassment* typically co-occurs with *shame*, but not with *guilt*. This is in line with what was observed in Section 4.1 and with the assumption that *embarrassment* is a 'low level version of shame' (Barrett, 2005, p. 955). In other words, although not conclusive, these patterns of co-occurrence suggest that the two emotions are qualitatively similar but differ in degrees of intensity, with 'shame' being perceived as more intense and destructive than 'embarrassment'.

The use of derogative expressions such as *fool* and *slut* testify to the public nature of *shame* and are worthy of further examination. However, a closer look at the co-text showed that *fool* in collocation with *shame* is used almost exclusively in the idiomatic expression *fool me once*, *shame on you*, *fool me twice*, *shame on me*. It does not refer to the emotional experience under analysis, hence it is of little interest to the study. *Slut* revealed a more interesting usage, that is, *slut shame* (employed in 88 out of 93 concordances):

(2) Men are told all the time on Reddit not to 'slut shame' their girlfriends who used to fuck entire rooms full of guys due to their 'sexually free' nature.

(3) Some guys will slut shame who has had more than two partners.

https://doi.org/10.1017/langcog.2024.6 Published online by Cambridge University Press

Although in these examples we do not have personal accountings of emotional experiences, the salience of this construction in the Reddit corpus indexes a socio-cultural phenomenon that sheds light on potential causes of 'shame' (here 'shame' arising is due to other people's judgments of sexual behaviour) and highlights its social nature.

Going back to the list of collocates, *remorse* typically collocates with *guilt* but not with *shame*. This mirrors the tendencies illustrated in Figure 1, where *remorse* is closer, hence semantically more similar, to *guilt* than *shame*. Finally, another striking feature is the high degree of correlation between *shame* and *guilt*. As illustrated in Tables 3 and 4, with 337 instances of co-occurrence, *shame* is the third most typical collocate of *guilt* and, conversely, *guilt* is the second most typical collocate of *shame*. A closer look at the concordance lines showed that they tend to co-occur in the construction *shame and guilt* (or vice versa). The conjunction *and* suggests that what is projected as shameful is likely to arise (or is functional to the projection of something as worthy of) guilt, and vice versa. It follows that *shame* and *guilt* are indeed related in English but are by no means the same thing – or there would be no need to distinguish between them (Baker et al., 2017, p. 47). This is in line with the vector space representations illustrated in Section 4.1, which demonstrated that, despite overlap between the two, there are components that are closer to either *shame* or *guilt* (e.g., *regret*), and that allow English speakers to differentiate between them.

4.2.2. *Semantic space of haji and zaiakukan*

The collocation analysis of *haji* is visualised in Figure 5. Translations and transcriptions are provided in Table 5, together with the statistical and frequency values associated with each collocate. Note that in the figure, the cut-off value is 8, and not 7 as elsewhere in the paper, otherwise the graph would be very difficult to read. For comparative purposes, however, the cut-off value considered during the data analysis

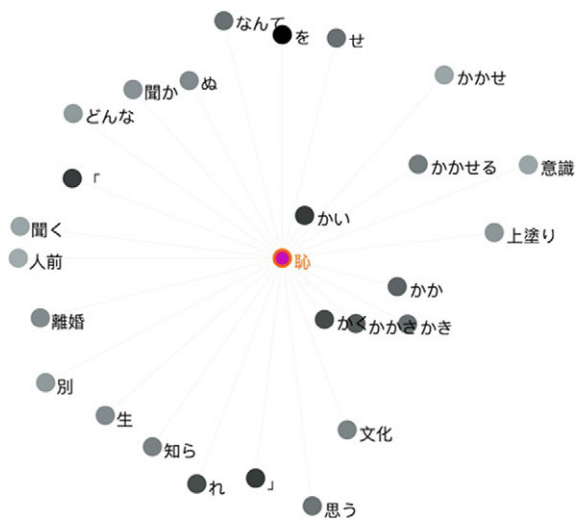

Figure 5. L5R5 collocates of *haji* in the Hatsugen Komachi corpus. LogDice value cut-off <8.

https://doi.org/10.1017/langcog.2024.6 Published online by Cambridge University Press

Table 5. L5R5 collocates of *haji* in the Hatsugen Komachi corpus (logDice ≥ 8)

Collocate	Translation	LogDice	Raw frequency
かい *kai*	suffer	12.1	88
かく *kaku*	suffer	11.6	60
かかかさ *kakasa*	suffer_CAUS_Pass	11.2	42
かか *kaka*	suffer_CAUS	10.8	32
かき *kaki*	suffer	10.3	24
かかせる *kakaseru*	suffer_CAUS	9.9	17
文化 *bunka*	culture	9.6	14
上塗り *uwanuri*	added (shame)	9.0	9
聞か *kika*	listen_CAUS	8.8	10
知ら *shira*	inform	8.7	15
生 *sei*	life	8.5	12
なんて *nante*	what	8.4	23
かかせ *kakase*	suffer_CAUS	8.4	6
思う *omou*	think	8.3	21
離婚 *rikon*	divorce	8.3	12
どんな *donna*	what kind	8.3	8
意識 *ishiki*	awareness	8.1	6
人前 *hitomae*	in public	8.0	5
聞く *kiku*	listen	8.0	6
別 *betsu*	distinction	8.0	8

is 7. The full list of collocates (not reported here for reasons of space) can be accessed through the link in the Data Availability Statement.

The analysis shows that *haji* is indeed a *social* emotion (Shott, 1979), because a number of its collocates can be traced back to interpersonal relationships or society in general. For instance, there is a strong collocation (logDice 9.6) with *bunka* 'culture' in the construction *haji no bunka* 'culture of shame', showing that this second-order classification (Raz, 2002; Sakuta, 1967) is something that laypeople talk about. Of note are also *rikon* 'divorce' and *danna* 'husband' (logDice values are 8.3 and 7.7, respectively), signalling that marital relationships are a recurrent topic among members of the community and, more importantly, that they are often associated with 'shame'. Finally, *hitomae* 'in front of people' (logDice 8.0) mirrors the English *publicly* (see Table 3), showing that in both samples, 'shame' is linked to the self in relationship with others (Tangney & Dearing, 2002). These findings provide linguistic evidence for the assumption that 'shame' lies 'at the intersection of subjective experience of one's own self and inter-subjective sensitivity to the social reality and the self's presence therein' (Krawczak, 2014b, p. 442).

As for *zaiakukan*, its low frequency (23) does not allow for big generalisations and no collocate with a logDice equal to 7.0 or higher was found in the corpus. Such a low frequency, however, may be taken to signal that among the users of Hatsugen Komachi 'guilt' may not be as salient as 'shame', both as a personal psychological phenomenon and as an interpersonal social one. Differently from what was observed in the English data, *haji* and *zaiakukan* are not mutual collocates, a factor suggesting that in the Japanese data 'shame' and 'guilt' do not overlap as they seem to do in English. Another possibility, however, is that *zaiakukan* is not the best candidate to access 'guilt' in texts. Future studies may use the terms listed in Table 2 to corroborate or falsify these preliminary findings.

https://doi.org/10.1017/langcog.2024.6 Published online by Cambridge University Press

4.2.3. Guilty

In the attempt to locate what warrants 'guilt' in the English data, we looked at the adjective *guilty* (11,602 occurrences, available through the link in the Data Availability Statement). First, we restricted the analysis to words that are used to the left of the node (span L1–L5) to identify the emoter, that is, who feels guilty. This time, we looked at raw frequencies of collocations because statistical measures tend to eclipse functional words that are nonetheless relevant to the aims of this section. The findings revealed that in our data *you* is the most frequent pronoun observed in the L1 position (3,100), followed by *I* (1,912), *she* (1,190) and *he* (806). This suggests that the emoter and the receiver of the utterance often overlap. We then asked what people feel *guilty for*, which prompted us to travel to the collocational network and investigate second-order collocates of *guilty for*. We restricted the analysis to positions R1-R2 on the assumption that what people feel *guilty for* is likely to appear immediately to the right of the node. The most frequent collocate found within these parameters is *not* (196) in the construction *guilty for not*, which suggests that people feel guilty for *not* doing/having done something. This is best illustrated with examples:

(4) I felt so guilty for not wanting to be with him anymore.

(5) I feel a little guilty for not making my mind up earlier.

Of note, it is also the fact that users tend to adopt linguistic strategies to distance themselves from the negatively evaluated action. This was achieved, for instance, by using the indexicals *it* (93), *that* (66) and *what* (52) so as not to mention what triggered the emotion, as in:

(6) It made me feel incredibly guilty for what I have done

Attention to the wider co-text is key to identifying what the producer was referring to. Future studies may provide a more comprehensive close reading of concordances to shed further light on what people feel guilty for.

4.2.4. Ashamed

Following the same procedure, a second collocation analysis showed that the personal pronoun most frequently used to the left of *ashamed* is *you* (813), followed by *I* (477), *he* (384) and *she* (368). We can thus assume that in our sample, the internal states verbalised as *guilty* and *ashamed* are more strongly associated with the receiver of the utterance. Triggers of *shame* were tentatively accessed by looking at the words and expressions employed immediately to the right of the construction *ashamed of* (R1–R2). The analysis of second-order collocates of *ashamed of* revealed that *it* (196) and *you* (187) are the two most frequent collocates in this position. *Your* (94) and *yourself* (87) follow in fourth and fifth position respectively. Importantly, these observations provide descriptive insights into the sample but should not be overgeneralised. The patterns observed may be specific to the corpus and not necessarily indicative of broader linguistic or cultural trends. The close reading of concordances revealed an additional reason for caution, showing that these collocational patternings tend to be preceded by a negative clause + *be* or *feel*. For instance, in 31 out of 196 occurrences of *ashamed of it*, the immediately preceding co-text reads *don't be*,

https://doi.org/10.1017/langcog.2024.6 Published online by Cambridge University Press

shouldn't be or *nothing to be*. Similar patternings apply also to *you* and *yourself*. Clearly, then, even when we have *it* or *you* at the right of *ashamed of* it does not necessarily mean that someone is experiencing 'shame', but the utterance may have a supporting function (as in *you shouldn't be ashamed of it*).

Another recurrent pattern at the right of the node and related to the collocate *yourself* is the construction *ashamed of my-/your-/him-/her-/them-selves* (271), as in

(7) I know how terrible that is and I'm so ashamed of myself.

Similarly to what was previously observed for the collocation *guilty for*, here the specific behaviour that triggered shame is not mentioned, and the producer seems to distance themselves from it using the indexical *that*. However, in contrast with example (4), in what follows in the utterance, the producer expresses a negative evaluation whose object is their whole self, not what they have done.

In discussing the causes of emotions like 'shame', as exemplified in phrases like *ashamed of myself*, we recognise that such instances offer valuable insights but are not sufficient for broad generalisations. Despite this caveat, our perspective aligns with the empirical investigations in Krawczak (2017, 2018), which adopt different but complementary methods within corpus linguistics and provide a more comprehensive analysis of these complex emotional constructs across *three* languages and cultures.

4.2.5. Haji

The collocational analysis of the Japanese data mirrors the one carried out in English. In Japanese personal pronouns are often omitted and, in line with this, there were no personal pronouns among the collocates of *haji* or *zaiakukan*, hence collocational analysis did not reveal whether the emoter is the producer of the utterance, other participants in the thread, or general third parties. As for the behaviours and events people are ashamed of, they were accessed by restricting the analysis of concordances to instances where *haji* is immediately followed by the copula *da* (informal) or *desu* (formal) (freq. 39), a construction that can be roughly translated as 'it is shame[ful]'. The choice to focus on concordances instead of collocates is motivated by the relatively low frequency of this two-word collocation, which allows for the manual annotation of the data.

After manually removing instances where *haji* is employed in a negative clause (e.g., *haji da to wa omoimasen* 'I don't think it is shame[ful]') and invalid examples, we annotated the remaining 30 concordance lines according to what triggers the emotion. The coded concordance lines can be accessed through the link in the Data Availability Statement. Based on semantic similarities, six groupings of triggers of *haji* and *zaiakukan* were identified, namely Relationships (11), Identity/Personality (9), (Lack of) knowledge (4), Money (3), Deception (1) and Sexuality/Body (1). In what follows, we illustrate representative examples of the first two.

(8) *Rikon wa haji da to iu ninshiki ga shimetsuite iru.*
'The recognition that divorce is shameful is deeply ingrained.' [Relationships]

(9) *Dokushin ga haji da.*
'Being single is shameful.' [Personality/Identity]

https://doi.org/10.1017/langcog.2024.6 Published online by Cambridge University Press

This taxonomy does not aim to be exhaustive, very often the groupings are blurred and topics overlap; therefore, the chosen categories are subjective and reductive. We are also aware that 30 concordances are way too few to provide considerations that are generalisable beyond the data set at hand. Despite these limitations, however, a more qualitative approach shed light on triggers of *haji*, showing that in a society where cis-heterosexual marriage is recognised as normative and proper, *haji* can be triggered by being, or considering the idea of getting, divorced (as exemplified in (6) and corroborated by the collocational analysis illustrated in Figure 3), or by aspects related to the identity or personality of the emoter.

4.2.6. Zaiakukan

Similarly to *haji da*, the low frequency of *zaiakukan* (23, which drops to 18 after removal of invalid examples) motivates the adoption of a more qualitative approach to its examination. The data were coded according to the same categories identified in the previous section, with the addition of the category Emotions: Relationships (8), Sexuality/Body (4), Deception (4), Money (1) and Emotions (1). By way of illustration, some examples are provided below.

(10) *Sō iu yūjin to no en o kiru no wa tsurai deshō. Yasashii hito nara zaiakukan sae idaite shimaimasu yo ne.*
'It must be difficult to cut that friend out of your life. If you are a kind person, you must feel guilty.' [Relationships]

(11) [*Uwaki*] *shinai to iu yori wa dekinai, to iu no ga tadashii desu ka ne. Moshi bareta toki no shakai teki risuku o osoreru hito ya, zaiakukan ya uwaki e no akuokan ga tsuyosugite dekinai* […] *riyū wa iroiro desu.*
'Rather than not cheating on someone, it's correct to say not being able to cheat on someone. There can be many reasons, people who, should they get caught, are afraid of the societal risks, or that can't do it because the [they feel] guilty and are disgusted by cheating.' [Sexuality]

(12) *Mawari ni uso o tsuiteiru zaiakukan ya aite niwa 'kakushite mōshiwake nai' tte kibun ni naru.*
'I feel guilty to lie to the people around me and sorry for hiding [something] from my partner.' [Deception]

The low numbers preclude any kind of statistical analysis, and these preliminary findings should be viewed merely as pointers for future research. Nonetheless, they show some interesting potential differences with *haji*: explicit references to the emoter identity and/or personality do not trigger *zaiakukan* in the data, which seems to be relatively frequently warranted by sexuality-related matters, such as being in a same-sex couple or cheating on the partner.

5. Discussion and conclusions

In this section, we elucidate patterns from the varied observations reported so far and address our research questions (RQs). We started from the text and the recurrent structures observed in it, based on which we proposed a (tentative) semantic representation of *shame* and *guilt* in English and *haji* and *zaiaku(kan)* in Japanese

https://doi.org/10.1017/langcog.2024.6 Published online by Cambridge University Press

as they emerge in our samples. Rooted in the principles of usage-based linguistics, this approach posits that key elements of meaning representation are acquired through the statistical distribution of words across various texts. Methodologically, this means that empirical observations of statistically significant lexical patterns are crucial in illuminating cognitive and psychological processes. This perspective is grounded in the understanding that language is inherently tied to its use in real-life contexts. Consistent with usage-based theories, it is assumed that individuals' cognitive and linguistic experiences are shaped by the frequency and context of language exposure (Divjak, 2019), a concept known as *priming* (Hoey, 2005). This reflects the core usage-based tenet that language structure and function are deeply interconnected with how language is habitually encountered and processed by individuals. In our hypothesis, 'shame' and 'guilt' can be illustrated as contiguous and overlapping *semantic spaces*, where certain expressions would be closer to certain elements than others or shared across elements. Areas of overlap between semantic spaces can thus be encoded in terms of shared collocates or shared semantic sets.

Semantic vector space representations revealed that *shame* is semantically close to *disgrace, dishonor* and *embarrassment*, whilst *guilt* is more closely related to notions of *fault*, *culpability*, and (*sincere*) *remorse.* Our analysis also suggests that 'shame' is often portrayed as a *public* experience (Figure 4), and 'guilt' as an emotion that encompasses both private elements, akin to sadness, and public aspects, such as the motivation to *openly acknowledge* (Figure 1) a transgression (RQ1). This interpretation, however, requires careful consideration. Specifically, it remains unclear whether *openly acknowledge* is a construction unique to 'guilt' as an emotional state, or if it also emerges in contexts where 'guilt' pertains to legal situations. Further investigation is needed to disentangle these different usages. Areas of overlapping were also observed. For instance, the two emotions can co-occur (as indexed by the construction *shame and guilt)* and both can be imposed on others – despite previous research suggesting otherwise (Bedford & Hwang, 2003, p. 128). Moreover, within our sample, *you* is the most frequent collocate for both *guilt* and *shame.* Similarly, both the adjectives *guilty* and *ashamed* tend to be associated with the receiver of the utterance (RQ2). Our collocational analysis, while illuminating, should be interpreted with caution, particularly regarding the implications for the types of emoters associated with 'guilt' and 'shame', due to the low frequency of data. Nonetheless, our findings suggest a nuanced interplay in the attribution of these emotions in discourse rather than a straightforward assignment of 'guilt' to the receiver and 'shame' to the producer of the utterance.

As for Japanese, *zaiaku*(*kan*) is semantically related to loss of face, trust and regret, whilst *haji* is closer to embarrassment and dishonour. The corpus-assisted close examination of *haji* and *zaiakukan*, whose low frequencies allowed for a more qualitative approach, also showed that they are both triggered by violations of interpersonal norms along the lines of their English counterparts. However, whilst in our data *zaiakukan* is often associated with sexuality-related matters, *haji* links more directly with the identity or personality traits of the emoter. Based on what was observed in our sample, we can then hypothesise that, in English as well as in Japanese, 'shame' and 'guilt' are correlated with different causes and that this, together with the features of meaning representations mentioned above, allows people to distinguish between the two experiences (RQ3). Clearly, however, it is difficult to arrive at firm conclusions, especially for the Japanese data where low frequency is an issue. Moreover, although the two data sources share a number of

https://doi.org/10.1017/langcog.2024.6 Published online by Cambridge University Press

common features that increase their comparability, texts from Reddit focus on relationships, whilst Hatsugen Komachi has traditionally been focused on 'women's issues'. This may skew the results concerning the cause of the emotions and calls for further studies to corroborate, or falsify, the tendencies observed here.

Methodologically, this article illustrated a replicable process to access semantically similar expressions, where we first built language-specific vector space representations and then looked at their meanings in context with corpus linguistic tools. This is an innovative and effective way to further advance our knowledge of the language of emotions, showing the value of merging corpus methods and NLP. Emotions, however, have fuzzy boundaries, and their cognitive and social reality cannot be accessed by looking only at examples of explicit emotional labels. Future studies may employ the hybrid and corpus-based methodology proposed here to further explore 'shame'- and 'guilt'-related expressions both inter- and intra-linguistically.

Data availability statement. All data and materials can be found at https://osf.io/n8d5g/?view_only=f51077116e9846c7aa79f6ac0dd1670e, or at the first author's OSF page.

Acknowledgements. We would like to thank our two anonymous reviewers and the editors of this special issue for providing invaluable comments.

Funding statement. This work was supported by JSPS KAKENHI under grant 22K18154.

Competing interest. The authors declare no competing interests exist.

References

Adoma, A. F., Henry, N. M., & Chen, W. (2020). Comparative analyses of bert, roberta, distilbert, and xlnet for text-based emotion recognition. In *17th international computer conference on wavelet active media technology and information processing (ICCWAMTIP)* (pp. 117–121). IEEE.

Antoniak, M., & Mimno, D. (2018). Evaluating the stability of embedding-based word similarities. *Transactions of the Association for Computational Linguistics*, 6, 107–119.

Arimitsu, K. (2001). The relationship of guilt and shame to mental health. *The Japanese Journal of Health Psychology*, 14(2), 24–31. https://doi.org/10.11560/jahp.14.2_24

Baker, H., McEnery, T., & Hardie, A. (2017). A corpus-based investigation into English representations of Turks and Ottomans in the early modern period. In M. Pace-Sigge & K. J. Patterson (Eds.), *Lexical priming: Applications and advances* (Vol. 79, pp. 41–66). John Benjamins. https://doi.org/10.1075/scl.79

Barrett, K. (2005). The origins of social emotions and self-regulation in toddlerhood: New evidence. *Cognition & Emotion*, 19(7), 953–979. https://doi.org/10.1080/02699930500172515

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science*, 20(5), 286–290.

Bedford, O., & Hwang, K. K. (2003). Guilt and shame in Chinese culture: A cross-cultural framework from the perspective of morality and identity. *Journal for the Theory of Social Behaviour*, 33(2), 127–144.

Bednarek, M. (2008). *Emotion talk across corpora*. Palgrave Macmillan.

Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press. https://doi.org/10.1017/9781316410899

Brezina, V., Weill-Tessier, P., & McEnery, A. (2020). *#LancsBox* v. 5.x. [software]. Available at: http://corpora.lancs.ac.uk/lancsbox

Brookes, G., & Chałupnik, M. (2022). Militant, annoying and sexy: A corpus-based study of representations of vegans in the British press. *Critical Discourse Studies*, 1–19. https://doi.org/10.1080/17405904.2022.2055592

Cavalera, C. (2020). COVID-19 psychological implications: The role of shame and guilt. *Frontiers in psychology*, 11, 571828.

Creighton, M. R. (1990). Revisiting shame and guilt cultures: A forty-year pilgrimage. *Ethos*, 18(3), 279–307.

Culpeper, J., & Tantucci, V. (2021). The principle of (im)politeness reciprocity. *Journal of Pragmatics*, 175, 146–164. https://doi.org/10.1016/j.pragma.2021.01.008

Divjak, D. (2019). *Frequency in language: memory, attention and learning*. Cambridge University Press. https://doi.org/10.1017/9781316084410

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Fabiszak, M., & Hebda, A. (2007). Emotions of control in Old English: Shame and Guilt. *Poetica*, 66, 1–35.

Fabiszak, M., Hilpert, M., & Krawczak, K. (2016). Usage-based cognitive-functional linguistics: From theory to method and back again. *Folia Linguistica*, 50(2). https://doi.org/10.1515/flin-2016-0013

Farese, G. M. (2016). The cultural semantics of the Japanese emotion terms 'Haji' and 'Hazukashii'. *New Voices in Japanese Studies*, 8, 32–54. https://doi.org/10.21159/nvjs.08.02

Garfinkel, H. (1967). *Studies in ethnomethodology*. Polity Press.

Geeraerts, D. (2010). *Theories of lexical semantics*. Oxford University Press.

Glynn, D. (2007). *Mapping meaning: Toward a usage-based methodology in cognitive semantics*. [Doctoral dissertation]. University of Leuven.

Glynn, D. (2010a). Corpus-driven cognitive semantics. Introduction to the field. In D. Glynn & K. Fischer (Eds.), *Quantitative methods in cognitive semantics: Corpus-driven approaches* (pp. 1–42). De Gruyter Mouton. https://doi.org/10.1515/9783110226423.1

Glynn, D. (2010b). Lexical fields, grammatical constructions, and synonymy. A study in usage-based cognitive semantics. In H.-J. Schmid & S. Handl (Eds.), *Cognitive foundations of linguistic usage patterns: Empirical studies* (pp. 89–118). Mouton de Gruyter. http://doi.org/10.13140/rg.2.1.1079.9524

Glynn, D. (2014). The social nature of ANGER: Multivariate corpus evidence for context effects upon conceptual structure. *Emotions in Discourse*, 69–82.

Grefenstette, G., & Tapanainen, P. (1994). What is a word, what is a sentence? Problems of Tokenisation. In *Proceedings of the 3rd international conference on computational lexicography, Budapest* (pp. 79–87).

Guidère, M. (2020). NLP applied to online suicide intention detection. HealTAC2020, Mar 2020. inserm-02521389.

Heinrich, P. (2012). *The making of monolingual Japan: Language ideology and Japanese modernity*. Multilingual Matters.

Higuchi, M. 樋口匡貴. (2002). *Kōchi jōkyō oyobi sichi jōkyō ni okeru haji no hassei mekanizumu - haji no kai jōcho betsu no hassei purosesu kentō* 公恥状況および私恥状況における恥の発生メカニズム-恥の下位情緒別の発生プロセスの検討 [The mediating mechanisms of embarrassment in public and private situations: The process of the subcategories of embarrassment]. *The Japanese Journal of Research on Emotions*, 9(2), 112–120.

Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Routledge/AHRB.

Hultberg, P. (1988). Shame—a hidden emotion. *Journal of Analytical Psychology*, 33(2), 109–126.

Hunston, S., & Thompson, G. (Eds.) (2000). *Evaluation in text: Authorial stance and the construction of discourse* (Reprinted). Oxford University Press.

Imada, H. (1989). Cross-language comparisons of emotional terms with special reference to the concept of anxiety. *Japanese Psychological Research*, 31(1), 10–19. https://doi.org/10.4992/psycholres1954.31.10

Inaba, K. 稲葉小由紀. (2009). *Zaiakukan* 罪悪感 [Guilt]. In Kōki Arimitsu 興記有光 & Akio 章夫菊池 Kikuchi (Eds.), *Jiko ishiki teki kanjō no shinrigaku* 自己意識的感情の心理学 [Psychology of self-conscious emotions]. Kitaōji shobō.

Jack, R. E., Garrod, O. G., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences*, 109(19), 7241–7244.

Kádár, D. Z., & Haugh, M. (2013). *Understanding politeness*. Cambridge University Press.

Kádár, D. Z., & Ran, Y. (2019). Globalisation and Politeness: A Chinese Perspective. In E. Ogiermann & P. G.-C. Blitvich (Eds.), *From Speech Acts to Lay Understandings of Politeness* (pp. 280–300). Cambridge University Press.

Keltner, D., Tracy, J. L., Sauter, D., & Cowen, A. (2019). What basic emotion theory really says for the twenty-first century study of emotion. *Journal of Nonverbal Behavior*, 43, 195–201.

Krawczak, K. (2014a). Shame and its near-synonyms in English: A multivariate corpus-driven approach to social emotions. In I. Novakova, P. Blumenthal, & D. Siepmann (Eds.), *Les émotions dans le discours* (pp. 83–94). Peter Lang.

https://doi.org/10.1017/langcog.2024.6 Published online by Cambridge University Press

Krawczak, K. (2014b). Shame, embarrassment and guilt: Corpus evidence for the cross-cultural structure of social emotions. *Poznan Studies in Contemporary Linguistics*, 50(4), 441–475.

Krawczak, K. (2017). *Contrasting languages and cultures: A multifactorial profile-based account of SHAME in English, Polish, and French.* halshs-01464866v3.

Krawczak, K. (2018). Reconstructing social emotions across languages and cultures: A multifactorial account of the adjectival profiling of shame in English. *French, and Polish. Review of Cognitive Linguistics*, 16(2), 455–493.

Krawczak, K., & Badio, J. (2015). Negative self-evaluative emotions from a cross-cultural perspective: A case of 'shame'and 'guilt' in English and Polish. *Linguistics, Psychology, Sociology*. https://doi.org/10.3726/978-3-653-04976-3/19

Kumamoto, M. (2019). Conceptualization of negative social emotions in French. A Behavioral Profile Approach to honte, honteux, culpabilité and coupable [Conference presentation abstract]. In *15th International Cognitive Linguistics Conference*. hal-04057038.

Laaksonen, S. M., Pääkkönen, J., & Öhman, E. (2023). From hate speech recognition to happiness indexing: critical issues in datafication of emotion in text mining. In *Handbook of Critical Studies of Artificial Intelligence* (pp. 631–642). Edward Elgar.

Langacker, R. W. (1987). *Foundations of cognitive grammar* (Vol. 1). Stanford University Press.

Marchi, A. (2023). Get back! Exploring discourses of nostalgia and nostalgic discourses using corpora. *Elephant & Castle*, 31, 192–211.

Martin, J. R., & White, P. R. R. (2005). *The language of evaluation: Appraisal in English*. Palgrave Macmillan.

Mead, M. (1972). *Blackberry winter: My earlier years*. William Morrow.

Meque, A. G. M., Hussain, N., Sidorov, G., & Gelbukh, A. (2023). *Guilt detection in text: A step towards understanding complex emotions.* arXiv. http://arxiv.org/abs/2303.03510

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). *Efficient estimation of word representations in vector space.* arXiv. https://doi.org/10.48550/arXiv.1301.3781

Miyake, K. 三宅和子. (2022). Mobairu jidai no goyōron: 'uchikotoba' wa nani o kaeta ka モバイル時代の語用論 —「打ちことば」は何を変えたか (Pragmatics in the era of moble phones: What have 'typed words' changed?) [Conference presentation]. In *Pragmatics Society of Japan.*

Öhman, E. (2021). The validity of lexicon-based sentiment analysis in interdisciplinary research. In M. Hämäläinen, K. Alnajjar, N. Partanen, & J. Rueter (Eds.), *Proceedings of the workshop on natural language processing for digital humanities, India* (pp. 7–12). ACL Anthology.

Öhman, E., & Rossi, R. H. (2023). Affect as a proxy for mood. *Journal of Data Mining and Digital Humanities*. https://doi.org/10.46298/jdmdh.11164

Pankratz, E., & Van Tiel, B. (2021). The role of relevance for scalar diversity: a usage-based approach. *Language and Cognition*, 13(4), 562–594.

Papandropoulou, I., & Sinclair, H. (1974). What is a word? *Human Development*, 17(4), 241–258.

Partington, A. (2004). 'Utterly content in each other's company': Semantic prosody and semantic preference. *International Journal of Corpus Linguistics*, 9(1), 131–156. https://doi.org/10.1075/ijcl.9.1.07par

Pavlenko, A. (2008). Emotion and emotion-laden words in the bilingual lexicon. *Bilingualism: Language and Cognition*, 11(2), 147–164. https://doi.org/10.1017/S1366728908003283

Pierrejean, B., & Tanguy, L. (2018). Predicting word embeddings variability. In M. Nissim, J. Berant & A. Lenci (Eds.), *Proceedings of the seventh joint conference on lexical and computational semantics* (pp. 154–159). Association for Computational Linguistics.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. *Theories of Emotion*, 1, 3–31.

Raz, A. E. (2002). *Emotions at work: Normative control, organizations, and culture in Japan and America*. Harvard University Asia Center.

Sabiston, C. M., & Castonguay, A. L. (2014). Self-conscious emotions. In R. Eklund & G. Tenenbaum (Eds.), *Encyclopedia of Sport and Exercise Psychology* (pp. 623–626). Sage Publications.

Sakuta, K. 作田啓一 (1967). *Haji no bunka saikō* 恥の文化再考 [Reconsiderations on the shame society]. Chikuma Shobō.

Scarantino, A. (2016). The philosophy of emotions. In J. M. Haviland-Jones, M. Lewis, & L. F. Barrett (Eds.), *Handbook of emotions* (4th ed.). Guilford Press.

Sheikh, S., & Janoff-Bulman, R. (2010). The "shoulds" and "should nots" of moral emotions: A self-regulatory perspective on shame and guilt. *Personality and Social Psychology Bulletin*, 36(2), 213–224.

Shott, S. (1979). Emotion and Social Life: A Symbolic Interactionist Analysis. *American Journal of Sociology*, 84(6), 1317–1334. https://doi.org/10.1086/226936

Sidorov, G., Balouchzahi, F., Butt, S., & Gelbukh, A. (2023). Regret and hope on transformers: An analysis of transformers on regret and hope speech detection datasets. *Applied Sciences*, 13(6), 3983.

Sidorov, G., Gelbukh, A., Gómez-Adorno, H., & Pinto, D. (2014). Soft similarity and soft cosine measure: Similarity of features in vector space model. *Computación y Sistemas*, 18(3), 491–504.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

Smith, R. H., Webster, J. M., Parrott, W. G., & Eyre, H. L. (2002). The role of public exposure in moral and nonmoral shame and guilt. *Journal of Personality and Social Psychology*, 83(1), 138.

Soares da Silva, A. (2020). Exploring the cultural conceptualization of emotions across national language varieties: A multifactorial profile-based account of pride in European and Brazilian Portuguese. *Review of Cognitive Linguistics*, 18(1).

Suzuki, N. 鈴木直人 (Ed.). (2007). *Kanjō shinrigaku* 感情心理学 [Psychology of emotions]. Asakura Publishing Company.

Tangney, J. P., & Dearing, R. L. (2002). *Shame and guilt*. Guilford Press.

Teodorescu, D., & Mohammad, S. M. (2022). *Frustratingly easy sentiment analysis of text streams: Generating high-quality emotion arcs using emotion lexicons.* arXiv preprint arXiv:2210.07381.

Tissari, H. (2006). Conceptualizing shame: Investigating uses of the English word shame, 1418–1991. In Selected proceedings of the 2005 symposium on new approaches in English historical lexis *(HEL-LEX)* (pp. 143–154). Cascadilla Proceedings Project.

Verschueren, J. (2011). *Ideology in language use: Pragmatic guidelines for empirical research* (1st ed.). Cambridge University Press.

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition*, 1(2), 219–247. https://doi.org/10.1515/LANGCOG.2009.011

Wierzbicka, A. (1999). *Emotions across languages and cultures: Diversity and universals*. Cambridge University Press.

Wierzbicka, A., & Harkins, J. (2001). Introduction. In J. Harkins & A. Wierzbicka (Eds.), *Emotions in crosslinguistic perspective*. Mouton de Gruyter.

Zhou, N., & Jurgens, D. (2020). Condolence and empathy in online communities. In B. Webber, T. Cohn, Y. He & Y. Liu (Eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 609–626). Association for Computational Linguistics.

Cite this article: Diegoli, E., & Öhman, E. (2024). Contrasting the semantic space of 'shame' and 'guilt' in English and Japanese, *Language and Cognition*, 1–23. https://doi.org/10.1017/langcog.2024.6