

ARTICLE

# A linguistic analysis of SHAME and GUILT in English and Japanese – with a focus on emotion co-occurrence

Eugenia Diegoli 

Department of Asian and North African Studies, Ca' Foscari University of Venice, Italy

**Corresponding author:** Eugenia Diegoli; Email: [eugenia.diegoli@unive.it](mailto:eugenia.diegoli@unive.it)

(Received 23 January 2026; Revised 23 January 2026; Accepted 27 April 2026)

## Abstract

SHAME and GUILT refer to compositional emotional experiences, some aspects of which are shared across linguacultures, while others are shaped by socialisation. This study combines questionnaire and corpus data to explore how people talk about and possibly experience SHAME and GUILT in English and Japanese contexts. First, keyword analysis is used to identify English and Japanese expressions frequently used in SHAME- and GUILT-inducing situations. These keywords are treated as potential markers of SHAME and GUILT. The focus then shifts to emotion co-occurrence. Assuming that emotions rarely occur in isolation, I examine patterns of lexical co-occurrence in two general corpora of Japanese and English, using some of the previously identified keywords as nodes. The findings suggest that FEAR and SADNESS have a dominant role in the emotional landscapes of SHAME and GUILT in both languages. In the English data, we find references to ANGER and DISGUST, while in the Japanese data, GUILT is more strongly associated with the semantic sphere of 'responsibility'. Methodologically, the study exemplifies an innovative approach to operationalising emotion through terms that emerge from the data itself and is more generally relevant to scholars interested in questions that revolve around (relative) ineffability and linguacultural variation.

**Keywords:** emotion co-occurrence; emotion network; guilt; Japanese vs English; shame

## 1. Introduction

Everyone knows what an emotion is, until asked to give a definition. Then, it seems, no one knows.

Fehr and Russell (1984, p. 464); (Cited in Wharton and De Saussure (2023), p. 29).



This study is about emotion, about how emotion relates to language and about variability in emotion concepts across languages. But what exactly is emotion? I adopt Scherer's definition of emotion as a 'relatively brief episode of synchronised responses [...] to the evaluation of an external or internal event as being of major significance' (Scherer, 2000, p. 141). For us to experience an emotion, then, there has to be a stimulus that triggers some form of evaluative process and an emotional response, which is in turn normally accompanied by action tendencies (Wharton & De Saussure, 2023, p. 48). These processes, however, often operate below the level of full conscious awareness. Emotions, in fact, are *ineffable*, and in more ways than one: they are often opaque to the very same person who experiences them, as well as to other people, who have to rely on bodily and verbal messages that index only indirectly selected aspects of a private and subjective emotional experience.

Specifically, this study focuses on the language of shame and guilt, which 'originate from the producer's awareness of having failed to be or behave in accordance with the standards recognised as proper by the group' (Diegoli & Öhman, 2024, p. 1298). Shame and guilt are typical examples of (negative) self-evaluative emotions. Self-evaluative emotions (also known as secondary, self-conscious, moral, social and evaluative) (Barrett, 2005; de Hooge et al., 2007; Gu et al., 2019; Krawczak, 2014; Lewis, 2003; Tracy & Robins, 2007, p. 2004) are negatively valenced inner states learned through socialisation, typically (but not exclusively) experienced in interpersonal contexts and involving a negative evaluation of the self, or parts of it. Such (d)evaluation is based on normative standards that a given group or community holds as appropriate, and hence guide our sense of right and wrong. Unlike primary emotions (e.g. Ekman, 1992), which are supposed to be similar across target populations, self-evaluative emotions can vary across linguacultures. *Shame* and *guilt* are two labels that index negatively valenced self-evaluative emotions in English. In the Japanese literature on the topic, their semi-equivalents are normally taken to be *haji* and *zaiakukan* respectively – but we will see that other translations are possible. I will use SHAME and GUILT in capital letters to distinguish emotion labels that apply cross-linguistically from their English realisations *shame* and *guilt*.

SHAME and GUILT can also be described as sequences of events that include specific cognitive processes, causes, repairing or managing behaviours and physiological reactions, which taken together provide an alternative (operational) form of definition. Table 1 illustrates the key components of SHAME and GUILT, identifying both overlapping and divergent features.

SHAME and GUILT, then, are compositional entities (Plutchik, 1980). Like other inner states, they can be observed only indirectly, insofar as they are shown via bodily changes (gestures, facial expressions, movements, etc.) or verbalised through language.

Two underlying assumptions guide my investigation. First, I assume that emotions rarely, if ever, occur in isolation: as documented in previous research (Lange & Zickfeld, 2023; Stella et al., 2022), emotion co-occurrence is the norm, yet it remains overlooked – and is notably absent from Table 1. We need theoretical and methodological tools that can help us account for it. Second, I assume that we can access how people think about and experience emotions by looking at how they habitually verbalise them. Concerning emotion co-occurrence specifically, we can make hypotheses about emotion association by looking at usage-based patterns of lexical co-occurrence in large collections of texts or 'corpora'.

**Table 1.** Key components of SHAME and GUILT

| Component            | SHAME/恥                                                                                                                                                    | GUILT/罪悪感                                                                                                                                                  | Relevant literature                                                                                                           |
|----------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|
| <b>Cognitive</b>     | Negative evaluation of the self                                                                                                                            | Negative evaluation of one's own behaviour                                                                                                                 | Creed et al., 2014, p. 3; Kikuchi & Arimitsu, 2006, p. 140                                                                    |
|                      | <i>Function:</i> enlist us in processes of self-surveillance and self-regulation                                                                           | <i>Function:</i> motivate compensatory action                                                                                                              | Creed et al., 2014, pp. 3–4, 16                                                                                               |
| <b>Social</b>        | <i>Trigger:</i> failure to meet standards of moral self                                                                                                    | <i>Trigger:</i> failure to meet standards of moral behaviour                                                                                               | Creed et al., 2014, pp. 3, 10–11, 17                                                                                          |
|                      | Experienced in front of (real or imagined) others                                                                                                          | Experienced both in front of (real or imagined) others and in private settings                                                                             | Arimitsu, 2015, p. 53; Diegoli & Öhman, 2024; Riezler, 1943, p. 458; Creed et al., 2014, pp. 17–18                            |
|                      | Aggravated if foreseeable/intentional<br><i>Function:</i> preserve social bonds, elicit compliance with prevailing norms, suppress transgressive behaviour | Aggravated if foreseeable/intentional<br><i>Function:</i> preserve social bonds, elicit compliance with prevailing norms, suppress transgressive behaviour | Orizu & He, 2016, p. 4127<br>Arimitsu, 2015, p. 53; Diegoli & Öhman, 2024; Riezler, 1943, p. 458; Orizu & He, 2016, pp. 17–18 |
| <b>Conative</b>      | Flee, hide or withdraw; behave in a way that is aimed at restoring one's standing as a valued person                                                       | Compensatory actions that are aimed at repairing the harm and restoring a balance of payments                                                              | Arimitsu, 2015, pp. 54, 57; Barrett, 2005, p. 954                                                                             |
| <b>Physiological</b> | Flushing, gaze averting, head lowering, accelerated heart rate                                                                                             | Chest tightness, sensation of carrying a heavy burden                                                                                                      | Arimitsu, 2015, pp. 53–57                                                                                                     |

## 2. Previous studies

We have seen that emotions are ineffable, embodied, variable and composite; all these elements make them difficult to be captured by traditional frameworks revolving around propositional meaning alone. Scholars across linguistic fields have proposed creative and interdisciplinary ways to study emotions in interaction that may be more effective than traditional approaches to language in use. Below I summarise those works that are more relevant to this study.

Marchi (2023) investigates nostalgia as a discursive phenomenon in a corpus of contemporary British newspapers. She examines both representations of nostalgia (which make use of explicit emotion terms and hence are generally easier to locate in corpora) and nostalgic content (which is more elusive because it triggers nostalgia without necessarily employing *nostalgia* terms). Marchi begins by looking at the collocational profile of the term *nostalgia*\* in her corpus and grouping the collocates by semantic category to identify dominant patterns of meaning. The collocational analysis is then extended to second-order collocates, which are taken to be potential triggers or markers of nostalgic discourse. Perhaps, more relevant to the current study is the keyword analysis Marchi carries out next: by comparing a sub-corpus of articles containing *nostalgia*\* words with the other texts in the newspaper corpus, she

identifies words that are key in the nostalgia texts. These keywords add to the list of potential markers and/or triggers of nostalgia.

Bednarek (2008) combines systemic and functional approach to explore language about emotion or *emotion talk* (*love, angry, sadness* and so on), as well as language as emotion or *emotional talk*, which includes 'all sorts of linguistic expressions that conventionally signal emotion without the recourse to explicit emotion talk' (Bednarek, 2008, p. 11). Bednarek is not interested in a specific emotion type, but rather in affect more generally. Her data comes from four different registers, as represented in a custom-made corpus compiled from the British National Corpus (BNC). Particularly relevant to my study is Bednarek's analysis of the semantic elements involved in systems of affect. In Chapter 3, she asks what the main elements for describing emotional experience are and, drawing from the FrameNet approach (Baker et al., 1998; <https://www.ninjal.ac.jp/english/>), she identifies three important aspects: Emoter (who feels the emotion), Emotion (the particular emotion involved) and Trigger (what causes the emotion; also called Stimulus in FrameNet). To know the meaning of an emotion term is to know what elements are typically involved in the affective schemas of that emotion (Russell, 1991, p. 39; Vannucci et al., 2026) – for example, we have seen in Section 1 that GUILT is normally triggered by a perceived failure to meet standards of moral behaviour. Emoters, Emotions and Triggers are realised lexico-grammatically through recurrent patterns that can be explored with corpus tools. In the current study, I focus on the category Emotion, although other categories, including Emoter and Trigger, do emerge from the data.

Krawczak (2014, 2017, 2018) employs multivariate corpus methods to investigate expressions of SHAME, EMBARRASSMENT and GUILT across three languages. The underlying assumption is that SHAME, EMBARRASSMENT and GUILT are, at least to some degree, culturally determined concepts that can be accessed by investigating the language people use to describe and express them.

All the studies presented here share a usage-based approach to language and representation: they take (contextualised) frequency of (co-)occurrence as a good predictor of typicality (Hoemann et al., 2019, p. 1837), conventionalisation (Diegoli, 2026, p. 16) and entrenchment (Langacker, 1987; Schmid, 2016, p. 548). Similarly, I take frequency-based patterns of language use to be indicative of conceptual tendencies in experiencing emotions. However, none of the studies reviewed here focus specifically on emotion co-occurrence, and they operationalise emotion through preselected emotion terms (although it is worth noting that the list used by Bednarek is rather comprehensive). The current study adds to the literature by proposing that emotion co-occurrence is a central aspect of emotional experience and an important part of the affective schemas that people acquire through socialisation. Methodologically, it exemplifies a data-driven method for operationalising emotion across typologically different languages and contributes to a growing body of research that combines corpus work with other perspectives and methods (see for example Fuoli & Bednarek, 2022).

### 3. Questions and aims

My two main research questions are as follows:

- What words and expressions are frequently used in prototypical discourses of SHAME and GUILT in English and Japanese data?

- Do SHAME and GUILT correlate with other emotional experiences? If so, what are the emotional experiences that are most frequently associated with SHAME and GUILT?

These two aims are closely interrelated: potential emotional markers identified through keyword analysis in highly specialised samples that are distinctively related to experiences of SHAME and GUILT will be used as search items in larger and more representative corpora. The patterns of lexical co-occurrence that will emerge from the corpus analysis are expected to shed light on key components of SHAME and GUILT, including other emotional experiences SHAME and GUILT tend to be associated with.

#### 4. Methods

Methodologically, the study exemplifies an innovative way of operationalising emotion through terms that emerge from the data itself. To do so, I use two main types of data: responses to a research questionnaire and corpus data.

My questionnaire, which is largely inspired by two measures used in psychology to estimate SHAME- and GUILT-proneness, presents scenarios that are expected to elicit SHAME and/or GUILT and asks respondents to describe how they would feel in that situation. The responses to the questionnaire were searched for (key)words that occur significantly more frequently than expected, some of which were then used as search terms in general corpora to explore linguistic patterns of distribution around emotion labels.

I combine research questionnaires and corpus data because, despite recent developments in corpus linguistics, meanings and functions that are not conventionally signalled by a limited set of linguistic markers are still very difficult to identify using corpus methods alone, as we may not know what to look for. Conversely, questionnaires reduce the number of variables we have to account for and facilitate the collection of data tailored to specific research needs. Questionnaires are also great for comparative purposes, enabling the collection of linguistic expressions across different linguistic cultures within controlled, identical contexts – a task often difficult to achieve with other methodologies (Blum-Kulka et al., 1989). The downside is that results from questionnaires are often difficult to generalise, as they are frequently constrained by small sample sizes and remain highly sensitive to the choice of questions and the language used therein. By integrating questionnaires with corpus methods, this study aims to produce more robust and comprehensive findings.

##### 4.1. Transparency and openness

All corpora, tools and statistical measures used in the present study are illustrated in Section 5. All corpora are publicly available through free online interfaces. Keyword and collocate lists, the Python codes used to generate Figures 1a, b and 2 and other additional materials are available on my [Open Science Framework](#) page. Excerpts from the questionnaire responses can be made available on request. Throughout the paper, I justify all my choices in terms of sample size, data exclusion and manipulation. The Python scripts used to generate Figures 1a, b and 2 were created with assistance from Claude Sonnet 4.1 (Anthropic, 2025).

#### 4.2. Research questionnaires: an introduction

A questionnaire is a research instrument that is made of a set of predefined questions that are expected to elicit appropriate responses from the participants to address the research questions at hand. In the current study, I use two versions (one in English and one in Japanese) of a scenario-based questionnaire composed of 15 questions. Of these 15 questions, 8 were adapted from the Test of Self-Conscious Affect-3 (TOSCA-3) (Tangney & Dearing, 2002), developed based on data from English speakers, and 7 from the *Kikuchi/Arimitsu - Jiko ishikiteki kanjō shakudo shinario-ban 12* 菊池・有光/自己意識の感情尺度シナリオ版-12 ‘Kikuchi/Arimitsu Self-conscious Emotions Scale-12 scenarios’ or KA-JiKoKan-12 (Kikuchi & Arimitsu, 2002), developed based on data from Japanese speakers (see Sections 4.2 and 4.3).

The TOSCA-3 and the KA-JiKoKan-12 were favoured over similar measures because they are based on empirical data, and hence are embedded in contextualised and everyday SHAME- and GUILT-inducing situations that people are likely to relate to. We can then expect the informants to provide descriptions of their reactions that are relatively close to those they would experience in real-life situations. It would have been of course possible to use only the TOSCA-3 in its English and Japanese versions, but a test developed on data from English speakers alone is likely to be biased towards English notions of *shame* and *guilt*. Combining it with a test developed by Japanese scholars on data from Japanese speakers is expected to shed light on emotional aspects that may be backgrounded in the literature in and about English alone.

Other types of questionnaires are also used in the literature – for an overview of the measures used in English-speaking countries, see Tangney and Dearing (2002), pp. 34–36. These include wordlist-questionnaires and diary report data.

In wordlist-questionnaires such as the Personal Feelings Questionnaire-2 (PFQ-2) (Harder & Zalma, 1990), respondents are asked to rate a list of affective descriptors (*remorse*, *embarrassment*, etc.) based on how often they feel that way. Scenario-based questionnaires were chosen over wordlist-questionnaires because they rely less heavily on the subject’s verbal skills and because we cannot take for granted a one-to-one correspondence between emotion labels across linguacultures. Most importantly, for the scope of this study, it is essential that I avoid using preselected emotion labels because they should be the *outcome* of the data collection.

Culpeper (Culpeper, 2011, pp. 62–63) collected diary report data where respondents were asked to report conversations in which someone said something to them that ‘made them feel bad’. Scenario-based questionnaires were favoured over diary reports because they are more appropriate for cross-linguistic studies as they present fewer variables to take into account. Like Culpeper, however, I employ open-ended questions. In the original versions of the TOSCA-3 and the KA-JiKoKan-12, respondents were asked, given a scenario, to rate a list of possible reactions from least to most probable, whereas, considering the purposes of this study, I provided the respondents with the scenarios and simply asked them to imagine themselves in that situation and describe their feelings.

Responses were elicited through the following instruction (English follows Japanese).

以下の文を読んで、自分がもしこういう場面に出会ったとしたら、どう感じたりするかを書いてみてください。もし役に立つと思われるなら、ご自身の気持ちを表すために、自由にあらゆる手段(通常とは異なるつづり、オノマトペ、絵文字、句読点、など)を使ってもかまいません。.

*As you read each scenario, try to imagine yourself in that situation. Then, describe how you would feel. If you deem it helpful, feel free to use any resource at your disposal (for example, non-standard word spellings, onomatopoeias, emojis, punctuation signs, etc.) to describe your feelings.*

Below I present in more detail the two measures on which my questionnaire is based.

#### 4.3. The TOSCA-3

The TOSCA-3 (11 negative scenario version) is a scenario-based questionnaire widely used in psychology to measure the shame- and guilt-proneness of English-speaking individuals, i.e. the degree to which people are prone to experience *shame* and/or *guilt* in certain situations. Each scenario is drawn from written accounts of everyday personal *shame* and *guilt* experiences provided by a sample of several hundred English-speaking college students and non-college adults.

Partly drawing from previous studies (Kitamura et al., 2024), I removed four of the 11 scenarios (numbers 3, 5, 11 and 16) from the TOSCA-3 because they were deemed to be inappropriate across linguacultures and/or target populations. For example, scenarios 5 and 11 were set at the office, a situation that I would expect younger respondents not to be familiar with. As suggested by Kitamura et al. (2024), scenarios 3 and 16, which take place at a dinner between couples and at a wine reception at a private home, were also removed because they were implausible in the Japanese context. Other scenarios (e.g. scenarios 2 and 4) were slightly modified to be appropriate across target populations, for example by removing 'at work', or replacing it with 'at a friend's house'. These modifications enhance applicability but reduce the generalisability of my findings, as discussed in Section 4.4.

As a result, my English questionnaire consists of the following eight scenarios from the TOSCA-3.

1. *You make plans to meet a friend for lunch. At 5 o'clock, you realise you stood your friend up.*
2. *You break something at a friend's house and then hide it.*
3. *You wait until the very last minute to plan a project, and it turns out badly.*
4. *While playing around, you throw a ball and hit your friend in the face.*
5. *You are driving down the road, and you hit a small animal.*
6. *You walk out of an exam thinking you did extremely well. Then you find out you did poorly.*
7. *While out with a group of friends, you make fun of a friend who's not there.*
8. *You are taking care of your friend's dog while your friend is on vacation, and the dog runs away.*

With the help of a native speaker and building on the translation proposed by Kikuchi (2003), these were translated into Japanese as follows.

1. 友人と昼食を一緒にする約束をしました。夕方 5 時に、その約束をすっぽかしたことに気づきました。
2. 友達の家で何かを壊してしまい、そのことをみんなに隠しています。
3. プロジェクトの計画をギリギリまで先延ばしにしたせいで、結局、うまくいきませんでした。
4. ボール遊びをしていて、自分の投げたボールが友達の顔に当たってしまいました。
5. ドライブをしていて、小さな動物をはねてしまいました。
6. テストがとてもうまくいったと思っていました。しかし点数は悪かったです。
7. 友達たちと一緒にいて、そこにいない友人のことをバカにして笑いあいました。
8. 友達が休暇をとっている間、イヌの世話を引き受けていましたが、そのイヌが逃げてしまいました。

The Japanese translation of the TOSCA-3 was verified via retranslation into English.

#### 4.4. The KA-JiKoKan

The KA-JiKoKan-12 (Kikuchi & Arimitsu, 2002), similarly to the TOSCA-3, describes situations that are meant to elicit specific emotions but, in contrast to the TOSCA-3, which is based on experiences provided by English speakers, the KA-JiKoKan-12 was developed on data from 628 Japanese university students.

Specifically, the KA-JiKoKan-12 presents 12 situations to measure six self-conscious emotions: feeling of interpersonal indebtedness, personal distress, shame, guilt, role-taking and empathic concern. All scenarios have been designed to potentially elicit all six emotions, depending on the individual's predisposition. Of the 12 scenarios proposed, I used only those scenarios (seven in total) that I deemed to be appropriate across target populations. Four scenarios that were not included in the current study (scenarios 1, 3, 8 and 9) described situations that only university students are likely to experience in their everyday lives. Scenario 12 was not included in the survey because it already assigned an emotion in the question (FEAR).

As a result, my Japanese questionnaire consists of the following seven scenarios from the KA-JiKoKan-12.

1. 友人が大切にしていたCDを無理をいって借りたのですが、持ち歩いているうちにどこかで落としてしまったらしいのです。
2. 「パソコンのことなら任せて」「何でも聞いてよ」などと言っていたのに、簡単なトラブルを解決することが出来ませんでした。
3. お互いに信頼し合っていると思っていた友人といっしょに買い物に行く約束をしました。その日に友人は地下鉄の駅で 1 時間以上も待っていたというのですが、こちらはすっかり忘れていて、別の友人と映画を見に行っていました。

4. 「ダメな奴だな」「いつも失敗ばかりだね」などと、仲良しの友人がみんなからいじめられているのに、自分は何も言い出せませんでした。
5. 男女いっしょの席でセックスがらみの話をして、「いやーね」「なによ、それ」などと言われ、みんなを不快な気持ちにさせてしまいました。
6. 友人と口論しているうちに、いつもはそういうことはしないのですが、ついカッとなって相手をたたいてしまいました。
7. 「そんなことも知らないの」「それはね」と言ってみんなに教えたことが、後になってまったくの間違いなことが分かりました。

These were translated into English with the help of a native speaker.

1. *You insist on borrowing a CD that your friend values, but you drop it somewhere while carrying it around.*
2. *You said things like 'If it's about computers, leave it to me' and 'Ask me anything', but you are not able to solve a simple problem.*
3. *You make plans to go shopping with a friend who thought they could rely on you. Your friend waits for over an hour at the subway station, but you have completely forgotten and have gone to see a movie with another friend.*
4. *Everyone bullies a close friend of yours saying 'You're useless' and 'You always mess up', but you do not say anything.*
5. *While sitting with male and female friends, you make dirty jokes. Everyone says things like 'That's gross' or 'What the hell?' and you make everyone uncomfortable.*
6. *While arguing with a friend, you lose your temper and end up hitting them – something you normally would never do.*
7. *You tell everyone 'You didn't know that?' 'Actually...' and explain it to them, but later it turns out to be completely wrong.*

The English translation of the KA-JiKoKan was verified via retranslation into Japanese. These seven scenarios from the KA-JiKoKan were then added to the eight scenarios from the TOSCA-3, for a total of 15 scenarios presented in a different random order to each respondent to reduce priming effects.

#### 4.5. Ethical considerations

The English and Japanese versions of the questionnaire were distributed online via Google Forms. I informed the respondents that the questionnaire was anonymous, that the responses would not be shared with any third party or used for purposes other than academic research, and that the questionnaire had received ethical approval from the Ethics Committee at Ca' Foscari University of Venice (report number 6/2025). All respondents were asked to give their informed consent. These practices are in line with the recommendations of the Association of Internet Researchers Ethics Working Committee and the European Commission (2021) (Franzke et al., 2020). I also briefly explained the structure of the questionnaire. I did not elaborate on its aim to avoid influencing responses. My personal information (role, institution, contact information) was also provided.

Then, I asked participants to provide their age, gender, first language and whether they had experiences of study and/or work abroad. The age verification question was the only mandatory question as it ensured that I collected only answers from participants older than 18 years old. Respondents could refuse to provide any other information, as they could refuse to answer any of the questions in the questionnaires. Following the collection of personal information, the respondents were asked to read the 15 scenarios and describe how they would feel in that situation (see Section 4.1).

The questionnaire concluded with the section ご意見やご要望・ご質問等ございましたらご記入ください/*Comments? Questions? Suggestions?*, which allowed respondents to actively engage with me and offer feedback. This final question was particularly helpful in revealing potential issues with the instrument's validity, which are addressed next.

#### 4.6. Limitations of questionnaire data

Despite all the precautions taken, the questionnaire still presents a number of limitations.

First and foremost, introspective judgments and self-reports can differ from actual behaviour: people may not have well-formed ideas about how they would behave in certain circumstances, and self-reports can be used strategically to project a positive self-image (Galasiński & Kozłowska, 2010).

I also appreciate that the TOSCA-3 and the KA-JiKoKan are quite old, which means that some scenarios are not realistic in today's society. This was noted by respondents themselves, some of whom used the final open question (*Comments? Questions? Suggestions?*) to comment that they struggled to imagine themselves in some of the situations proposed, hence challenging the validity of the instrument. The mention of devices that are now obsolete (e.g. the CD in scenario 1 of the KA-JiKoKan), as well as the absence of any mention of the internet, mobile phones and other innovations that are now ubiquitous, further reinforces the impression that these scenarios need updating.

Similarly, both the TOSCA-3 and the KA-JiKoKan are based on a binary construction of gender, while we have now come to have a better understanding of gender in non-binary terms.

Some questions from the KA-JiKoKan seem to be unnecessarily long, which may hinder the data collection and contribute to the perceived artificiality of certain scenarios.

The translation process was not without complexities and different translations are of course possible.

Finally, the choice to focus on scenarios that are expected to be appropriate across target populations led me to remove scenarios that took place at the office. Particularly in the Japanese linguaculture, where social ranks and roles and notions of authority and hierarchy are of central importance and affect the signalling of deference, subordination and affect in complex ways, the exclusion of scenarios where there is a power imbalance between interactants reduces the generalisability of my findings. Future studies may propose an updated version of the scenarios that may offer additional buffers against these limitations. At the time being, and given the comparative scope of my research, I deem my adaptation of the TOSCA-3 and the

KA-JiKoKan to be a sufficiently good resource to elicit SHAME- and GUILT-related expressions in English and Japanese.

## 5. Results: from keywords to collocational networks

### 5.1. Some patterns in the questionnaires responses

To collect as much evidence as possible, all responses given to the two versions of the questionnaire between 16 June and 25 October 2025 were examined, regardless of the difference in size between the English and the Japanese sets of responses. [Tables 2 and 3](#) provide a quantitative overview of the data collected using the questionnaire.

I excluded responses given by minors or non-native speakers, in languages other than Japanese (for the Japanese questionnaire) or English (for the English questionnaire), responses that were clearly unrelated to the question, or given by respondents who filled in only a very small number of questions. As a consequence, responses given by 10 out of 94 respondents to the Japanese questionnaire, and five out of 58 respondents to the English questionnaire were not considered for further analysis. The relatively low number of participants limits the generalisability and representativeness of my findings. Nonetheless, the collected texts are very specific and informative, and I deem them appropriate to carry out qualitative as well as statistical forms of analysis.

The two sets of data were uploaded onto the corpus linguistic software AntConc (version 4.3.1; Anthony, 2024), where I carried out keyword analysis to identify words and expressions that are used statistically more frequently in the two sets of responses, when compared with general use. For this comparison, I used as reference two publicly available corpora from the Leipzig Corpora Collection (Goldhahn et al., 2012): the 2018 10 k English web corpus from .com domains and the 2020 10 k Japanese web corpus. I take these two corpora to be a reasonably good option for the comparative purposes of this study, as they use the same format and comparable sources (<https://wortschatz.uni-leipzig.de/en/download/>) (see also Partington and Diegoli (2025), p. 195).

[Table 4](#) summarises the software, corpora and statistical measures used to compute keywords. [Table 5](#) lists 49 (out of the top 100) keywords that I take to be potential markers of SHAME and GUILT in Japanese and English, with the six most typical

**Table 2.** Overview of Japanese questionnaire data

| Japanese questionnaire data                             | n      |
|---------------------------------------------------------|--------|
| Questions                                               | 15     |
| Respondents/of which not included (as of 25 Oct., 2025) | 94/10  |
| Responses analysed with corpus tools                    | 1,257  |
| Tokens                                                  | 25,247 |

**Table 3.** Overview of English questionnaire data

| English questionnaire data                              | n      |
|---------------------------------------------------------|--------|
| Questions                                               | 15     |
| Respondents/of which not included (as of 25 Oct., 2025) | 58/5   |
| Responses analysed with corpus tools                    | 792    |
| Total tokens                                            | 13,031 |

Table 4. Software, corpora and statistical measures used to compute keywords

|                     | English data                                             | Japanese data                                             |
|---------------------|----------------------------------------------------------|-----------------------------------------------------------|
| Software            | AntConc (version 4.3.1)                                  | AntConc (version 4.3.1)                                   |
| Target corpus       | English questionnaire responses                          | Japanese questionnaire responses                          |
| Reference corpus    | 2018 10k English web corpus (Leipzig corpora collection) | 2020 10k Japanese web corpus (Leipzig corpora collection) |
| Statistical measure | Log-likelihood                                           | Log-likelihood                                            |

Table 5. Selected keywords in the questionnaire responses (statistical measure: log-likelihood)

| Japanese keywords                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | English keywords                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 謝罪 ‘apology’ (385.028); 申し訳‘(to feel) sorry’ (380.957); ごめん ‘sorry’ (372.063); しまっ [verb-ending form indicating unintentionally] (333.931); 謝る ‘to apologise’ (328.806); 恥ずかしい ‘embarrassed/embarrassing’ (188.341); 謝り ‘to apologise’ (180.669); 罪悪 ‘guilt’ (166.016); 後悔 ‘regret’ (152.291); 気持ち ‘feeling’ (151.937); すぐ ‘immediately’ (146.447); 弁償 ‘compensation’ (146.164); 反省 ‘remorse/reflection’ (132.45); 早く ‘quickly’ (112.544); なきゃ ‘must/have to’ (111.124); ... [ellipsis] (108.81); 焦る ‘to panic/to rush’ (107.413); 感じる ‘to feel’ (106.241); 謝っ ‘to apologise’ (102.53); 悪い ‘bad’ (99.162); 訂正 ‘correction’ (98.722); 気 ‘feeling’ (97.225); [気] まずい ‘awkward’ (94.576); バレ [る] ‘to get caught’ (85.005); 正直 ‘honestly’ (82.646); ? [question mark] (82.482); 感 ‘feeling’ (80.768); 間違っ ‘to make a mistake’ (80.227); 埋め [合わせ] ‘to make up for’ (74.075); 仕方 [ない] ‘it cannot be helped’ (73.537); 落ち込む ‘to feel down’ (73.232); 謝ら ‘to apologise’ (68.35); 謝ろう ‘let’s apologise’ (63.467); 恥ずかしく ‘embarrassed/embarrassing’ (63.467); べき ‘should’ (62.5); 詫び ‘apology’ (62.087); 急い ‘to hurry’ (61.281); 約束 ‘promise’ (57.222); 最悪 ‘the worst’ (56.444); 内心 ‘inner thoughts’ (53.702); 大丈夫 ‘it’s okay’ (53.111); バカ ‘idiot’ (53.052); 素直 ‘honestly’ (51.649); 奢る ‘to treat (someone to a meal)’ (48.82); 誠意 ‘sincerity’ (48.82); 馬鹿 ‘idiot’ (47.465); がっかり ‘disappointed’ (47); 許し ‘to forgive’ (45.645); 次 ‘next time’ (45.071). | guilty (714.905); embarrassed (583.186); ashamed (323.739); bad (286.301); sad (185.974); disappointed (184.622); scared (182.821); anxious (180.605); worried (167.016); angry (152.279); apologise (149.768); sorry (133.761); guilt (133.549); laugh (126.344); upset (116.035); apologetic (113.093); shame (105.402); stupid (99.14); frustrated (94.472); very (77.871); shocked (74.261); mistake (73.153); apologise (66.421); hurt (65.792); annoyed (62.194); immediately (52.831); uncomfortable (51.275); 🤔 (50.885); stressed (48.012); confused (45.97); cry (45.284); useless (45.284); profusely (45.23); forgetting (44.505); terrible (43.694); replacement (40.698); awful (39.576); hide (37.755); wrong (37.506); regretful (33.922); extremely (31.692); badly (28.517); devastated (28.268); embarrassed (28.268); silly (26.841); incredibly (25.643); regret (25.643); awkward (25.643); hurting (25.168). |

emotion labels highlighted in bold. The remaining keywords, not listed in Table 5 and not further discussed here, include frequent morphosyntactic elements (e.g. て, た, に, だ; *i, d, and, n't*), certain punctuation signs (e.g. 。) and lexical items that relate to specific elements in the scenarios proposed (e.g. 友人, 友達, 連絡; *friend, dog, animal*). The complete list of keywords is available as additional materials.

The keyword analysis isolates lexical items that characterise language use in the questionnaire responses, and hence addresses my first research question (*What words and expressions are frequently used in prototypical discourses of SHAME and GUILT in English and Japanese?*). The samples are not representative and do not warrant general conclusions, but they do reveal some tendencies in the data. In the Japanese set of responses, it is striking that the top words are almost exclusively apology expressions: *shazai* ‘apology’, *mōshiwake* ‘(to feel) sorry’, *gomen* ‘sorry’, *ayamaru* ‘apologise’; and later in the list we also have other forms of the verb *ayamaru* ‘apologise’ (e.g. *ayamara-*, *ayamarō*), *wabi* ‘apology’ and *yurushi* ‘to forgive’. The high number of apology-related expressions suggests that corrective actions are prototypical reactions in the sequence of events that Japanese speakers describe as SHAME and/or GUILT, and possibly hints at the use of apology expressions as emotion descriptors. Conversely, in the English responses, emotion labels are pervasive (*guilty*, *embarrassed*, *ashamed*, *disappointed*, *scared*, *anxious*, *worried*, *angry*, *guilt*, *upset* and so on), while items indexing corrective action tendencies are fewer and lower down in the list. Moreover, while in Japanese we find elements referring to the sphere of sincerity (*shōjiki/sunao* ‘honestly’, *seishin* ‘sincerity’), English favours adverbs of degree (*very*, *incredibly*). Areas of overlap are also apparent, with shared agglomerations of keywords, including emotion labels (*hazukashii* ‘embarrassed/embarrassing’, *kimazui* ‘awkward’, *ochikomu* ‘to feel down’, *embarrassed*, *sad*, *angry*, *awkward*), corrective action tendencies (*shazai* ‘apology’, *umeawase* ‘to make-up for’, *benshō* ‘compensation’, *sorry*, *replacement*, *hide*) and time references conveying urgency (*sugu* ‘immediately’, *hayaku* ‘quickly’, *immediately*).

In the next section, I explain how I searched for the collocational patterns of the six most typical emotion labels in the keyword lists using the Balanced Corpus of Contemporary Written Japanese (BCCWJ) (Maekawa et al., 2014) and the British National Corpus (BNC) (BNC Consortium, 2007) as target corpora. The collocational analysis is aimed at further developing the keywords into more complex semantic fields of SHAME and GUILT, with a focus on emotion association. The choice to depart from the words that are key in the questionnaire responses and use them as nodes in corpora offers a non-subjective way to access representations, ensuring that the analysis is not restricted only to what we have established a priori counts as SHAME and GUILT. Many of the words listed in Table 5 may simply not have been found via corpus methods alone, and hence the semantic picture that they contribute to establishing is expected to be relatively accurate and comprehensive.

## 5.2. Emotion association in corpora

Bostan et al. (Bostan et al., 2020) note that emotion analysis that considers the full structure of an emotion is still lacking. Corpus analysis, and specifically collocational analysis, which identifies the words that regularly appear together in a corpus, can help us shed light on the elements that compose the conceptual structures of SHAME and GUILT, including other emotional experiences they are regularly associated with.

For both Japanese and English, I investigated the collocational patterns of the top six keywords indexing emotion labels highlighted in bold in Table 5. For Japanese, the focus is on: 申し訳 *mōshiwake* ‘(to feel) sorry’, 恥ずかしい *hazukashii* ‘embarrassed/embarrassing’, 罪悪 *zaiaku* ‘guilt’, 後悔 *kōkai* ‘regret’, 反省 *hansei* ‘remorse/

reflection'<sup>1</sup> and 焦る *aseru* 'to panic/to rush'; for English, on: *guilty*, *embarrassed*, *ashamed*, *sad*, *disappointed* and *scared*. From a prototype view (Fehr & Russel, 1984), some of these are more easily recognisable as emotions than others. While *mōshiwake*, *aseru* and *disappointment* may be viewed as less prototypical members of the category, contextual usage justifies their inclusion as emotion terms. *Mōshiwake*, *zaiaku*, *kōkai*, *hansei* and *guilty* relate more closely to the superordinate category of GUILT; *hazukashii* 'embarrassing/embarrassed', *embarrassed* and *ashamed* to SHAME; *aseru* 'to panic/to rush' and *scared* to FEAR; *sad* and *disappointed* to SADNESS.

The approach I adopt here is data-driven: decisions on which linguistic features should be investigated further are made on the basis of frequency and typicality information extracted from the data itself (Rayson, 2008, p. 521). When used within a comparative study, it has, however, the drawback that, while some of the words that I use as nodes in the collocational analysis are semantically similar across the two languages (*embarrassed* and *hazukashii*, for example), others (including *kōkai* 'regret' and *sad*) appear only in one language.

For Japanese, the L3–R3 collocates (i.e. up to three words to the left or right of the node) of the above six emotion labels were extracted from the BCCWJ using the Chunagon (<https://chunagon.ninjal.ac.jp>), a web application for searching the corpora provided by the National Institute for Japanese Language and Linguistics (NINJAL; <https://www.ninjal.ac.jp/english/>). Within Chunagon, and following an exploratory analysis, I used the Short Unit Word (SUW) search mode.<sup>2</sup> For English, the collocates were extracted from the BNC using SketchEngine (Kilgariff et al., 2014), a web interface for the analysis of corpora. Ideally, one would experiment with different spans for English and Japanese (with a wider span for the latter), considering that the two languages are typologically different, that Japanese in its written form has no spaces between word-like units, and that part-of-speech taggers tend to divide Japanese texts into word-like units that are actually closer to the concept of morpheme in English. However, attempts to search the BCCWJ for collocational spans exceeding three words returned no results, likely due to the complexity of the query. Hence, I stick to an L3–R3 collocational span for both languages.

The Chunagon and SketchEngine are very different in terms of user interface, tools and statistical methods they make available. Two important differences are particularly worth mentioning. First, unlike SketchEngine, which allows for simultaneous searching of collocates on both sides of a target item (e.g. an L3–R3 span), Chunagon requires left (L3–L1) and right (R1–R3) collocates to be computed separately. Because I applied a frequency threshold of three, this separate computation may also have excluded certain (infrequent) collocates that would have met the threshold had they been processed within a single combined span. Second,

<sup>1</sup> *Kōkai* and *hansei* are semantically very similar, and in fact the Genius Japanese-English Dictionary (3<sup>rd</sup> ed.) translated both as 'regret' or 'remorse' depending on the context. Here they are translated as 'regret' and 'remorse/reflection' respectively based on the translations found in the OPUS parallel corpora (Tiedemann, 2012).

<sup>2</sup> Japanese does not use white spaces between 'words', and the same phrase can be segmented in different ways. Chunagon has two main search modes: the Short Unit Word (SUW) and the Long Unit Word (LUW) search modes. The SUW deconstructs compound items into smaller morphological components, while the LUW groups specific compound items into a single unit. In the current study, after experimenting with both modes, I use the SUW because it returned a higher number of collocates, providing for a more comprehensive picture of the semantic landscape of emotion labels.

**Table 6.** Patterns of emotion association in the BCCWJ

| Node                                                          | L3–R3 collocates (freq.; MI score)                                                                                                                                                                                                                                                                                            |
|---------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 申し訳 ‘(to feel) sorry’<br>恥ずかしい ‘embarrassed/<br>embarrassing’ | 嬉しい ‘happy’ (9; 1.762).<br>嫌 ‘to hate/disgusting’ (13; 2.610); 情けない ‘shameful’ (9; 4.895); 嬉しい ‘happy’ (9; 2.272); 照れ ‘to be embarrassed’ (5; 7.148); 怖い ‘afraid’ (5; 1.929); 困る ‘to be troubled’ (5; 1.356); 悲しい ‘sad’ (4; 2.647); 辛い ‘painful’ (4; 1.917); 慌てる ‘to panic/to rush’ (3; 2.186); 悔しい ‘frustrating’ (3; 3.281). |
| 罪悪 ‘guilt’<br>後悔 ‘regret’                                     | 苛む ‘to torment’ (12; 5.458); 羞恥 ‘shame’ (4; 3.256).<br>反省 ‘remorse/reflection’ (14; 5.952); 苛む ‘to torment’ (6; 8.202); 自責 ‘self-reproach/remorse’ (5; 8.820); 悲しみ ‘sadness’ (5; 4.873); 不安 ‘anxiety’ (4; 2.133); 欲しい ‘to want/desire’ (4; 0.720); 責める ‘to blame’ (3; 4.916).                                                 |
| 反省 ‘remorse/reflection’<br>焦る ‘to panic/to rush’              | 後悔 ‘regret’ (14; 5.952); 悔悟 ‘contrition’ (6; 10.749).<br>//                                                                                                                                                                                                                                                                   |

**Table 7.** Patterns of emotion association in the BNC

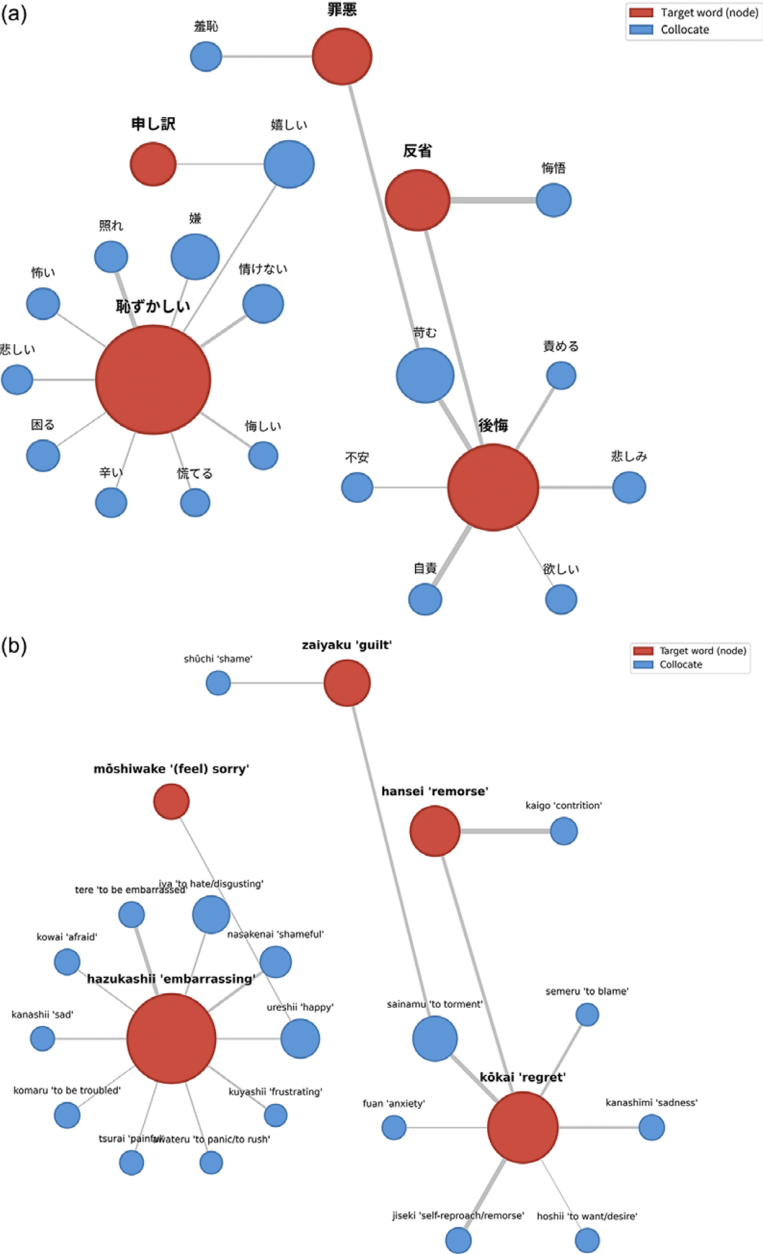
| Node                   | L3–R3 collocates (freq.; MI score)                                                                                                                                                                              |
|------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| guilty<br>embarrassed  | //<br>ashamed (14; 11.366); shy (8; 10.258); angry (8; 8.606); guilty (6; 8.156); awkward (4; 9.018); afraid (4; 7.138).                                                                                        |
| ashamed                | embarrassed (14; 11.366); guilty (8; 7.721); afraid (8; 7.288); proud (5; 7.455); angry (5; 7.079); fear (5; 5.229); sorry (4; 5.328); disgusted (3; 11.902); humiliate (3; 8.774); horrify (3; 8.506).         |
| sad                    | happy (25; 6.054); lonely (23; 8.785); funny (13; 6.574); angry (8; 6.014); pathetic (7; 8.422); bitter (7; 6.577); disappointed (5; 8.813); worried (5; 7.760); disappointment (5; 6.621); pitiful (4; 9.339). |
| disappointed<br>scared | angry (6; 8.867); sad (5; 8.830); frustrated (3; 11.20).<br>//                                                                                                                                                  |

SketchEngine, unlike Chunagon, provides, in addition to the frequency, typicality scores, including the logDice and Mutual Information (MI) score, which are more informative than frequency because what is frequent is not necessarily typical or meaningful. To ensure comparability, I manually computed the MI score of Japanese collocates and ranked both the English and Japanese collocates by MI score.

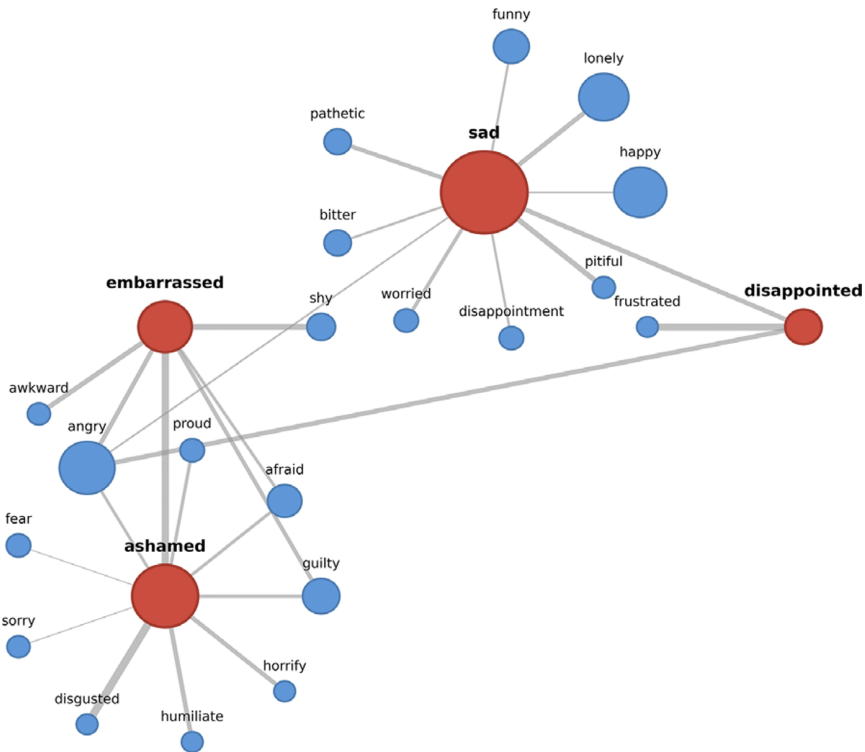
The full list of collocates is available as additional materials. Here, I focus exclusively on collocates semantically tagged as Emotion label to explore how different emotional experiences (as they are verbalised through language) relate to one another in networks of collocations. Tables 6 and 7 list the collocates coded as Emotion label. Between brackets, I indicate absolute frequency and MI score of the Japanese (Table 6) and English collocates (Table 7). As mentioned, I set a frequency threshold of three for both languages. The MI scores of Japanese collocates are overall relatively low, most likely because Japanese POS-taggers deconstruct lexical items into morphological components normally smaller than English ‘words’ (see note 2).

Figures 1a, b and 2 visualise the collocational networks of my Japanese and English nodes.

The nodes are connected according to co-occurrence relationships with the lexical items semantically tagged as Emotion label. The annotation of the collocates as Emotion label is based on the semantic role that they acquire *in context*, which



**Figure 1.** a. Collocational network of Japanese emotion labels in the BCCWJ. Measure of collocational strength: MI score. Created with assistance from Claude Sonnet 4.1. b. Collocational network of Japanese emotion labels (Romanised) in the BCCWJ. Measure of collocational strength: MI score. Created with assistance from Claude Sonnet 4.1.



**Figure 2.** Collocational network of English emotion labels in the BNC. Measure of collocational strength: MI score. Created with assistance from Claude Sonnet 4.1.

requires the analyst to read through multiple (and, when feasible, all) concordance lines for each collocation. This qualitative approach to the data complements statistical analyses and ensures accurate interpretations.

*Aseru* 'to panic/to rush' and *scared* do not typically collocate with any other emotion label, hence these two nodes are not visualised in the figures. Similarly, *guilty* seems to be used in the BNC predominantly in legal contexts, and its collocates (which include *pleaded*, *offence*, *found*, *verdict*; see example 1) are unrelated to *guilt* as an emotional experience. Hence, *guilt* too is not visualised in the figure.

- (1) If this rule is not complied with, the issuer is **guilty** of an offence.

In the collocational networks, the red circles correspond to our nodes, while the blue circles correspond to their collocates. The lines reveal co-occurrence relationships. Circle size index absolute frequencies, with larger circles corresponding to more frequent items. Line thickness varies based on MI score, with stronger collocations indexed by thicker lines. Distance from the node or between collocates was optimised to improve readability and has no particular meaning.

The associations visualised in Figures 1a, b and 2 answer my second research question (*Do SHAME and GUILT correlate with other emotional experiences? If so,*

what are the emotional experiences that are most frequently associated with SHAME and GUILT?). In what follows, I further elaborate on these associations and the understandings that lie behind them.

### 5.3. Zooming in

As mentioned, *aseru* ‘to panic/to rush’ and *scared* do not typically collocate with other emotion labels. There are however two semantically similar items: *awateru* ‘to panic/to rush’, which in the BCCWJ collocates with *hazukashii* ‘embarrassed/embarrassing’ (example 2), and *fear* and *afraid*, which in the BNC collocate with *ashamed* (example 3).

- (2) 「[...]あら、シュンちゃん、なっていたのね。」リンゴちゃんが顔を  
をのぞきこんだので、少年ははずかしそうにあわてて話題を変える。  
‘[...] Ara, shun-chan, naite ita no ne’. Ringon-chan ga kao o nozoki-konda  
node, shonen wa **hazukashiso** ni **awatete** wadai o kaeru.  
“[...] Oh, Shun-chan, you were crying, weren’t you?” As Ringo-chan peered  
into his face, the boy, **embarrassed**, **hastily** changed the subject’.
- (3) ‘Can I stay after you have hit me?’ he replied. ‘You’ve made me **afraid** and  
**ashamed** of you. I will not come here again!’

I will go back to the role of FEAR later in this paragraph. *Ashamed*, as well as *embarrassed* (example 4), collocates also with *guilty*, here used with its emotional meaning.

- (4) Though perhaps because we feel **guilty** or **embarrassed** about the whole area  
of mental health we are not tackling the problems when they come up nearly  
as well as we might.

Likewise, semantically related to GUILT is *sorry* in collocation with *ashamed*.

Similarly, in the BCCWJ, *zaiaku(kan)* collocates with *shūchi(shin)*. However, the low frequency ( $n = 3$ ) and the peripheral position this collocation occupies in the semantic space suggest that *zaiaku(kan)* and *shūchi(shin)* are used in specific situations. A closer look at the concordance lines surrounding the collocation *zaiaku(kan)* + *shūchi(shin)* supports this hypothesis, as two out of the three occurrences found in the BCCWJ are from a 2003 business book. The educational aim of the publication is apparent in example (5):

- (5) 性的欲求を持ったり、性的行為をすることに罪悪感や羞恥心を感じる。  
これは、文化がルールを決め、子供が发育する中で、性的欲求  
について話したり、行為したりすることは悪いことで、恥ずかしい  
ことであることを親や周囲から教えられ、育てられるからである。  
*Seiteki yokkyū o mot-tari, siteki koi o suru koto ni zaiakukan ya shūchishin o*  
*kanjiru. Kore wa, bunka ga rûru o kime, kodomo ga hatsuiku suru naka de,*  
*seiteki yokkyū ni tsuite hanashi-tari, kōi shi-tari suru koto wa warui koto de,*  
*hazukashii koto de aru koto o oya ga shūi kara oshierare, sodaterareru kara*  
*de aru.*

'People feel **guilt** and **shame** for having sexual desires or engaging in sexual acts. This is because culture defines the rules, and as children develop, they are raised being taught by their parents and those around them that speaking about sexual desires or engaging in sexual acts is something bad and embarrassing'.

In other words, in Japanese we find a correlation between *zaiaku(kan)* and *shūchi (shin)* only in specialised discourses, while in English the link between GUILT and SHAME as emotional concepts is much stronger. These findings mirror those observed by Diegoli and Öhman (2024) in a different set of data.

We may also note that in the BNC *embarrassed* and *ashamed* are also (mutual) collocates, and in fact we know from the literature (Barrett, 2005, p. 955) that the two are closely related, as *embarrassment* is often seen as a milder version of *shame*. Conversely, in Japanese, *hazukashii* does not collocate with lexical items directly indexing SHAME, possibly because SHAME and EMBARRASSMENT are continuous under the very same label *hazukashii* (Arimitsu, 2015, p. 57).

With reference to this last point, it is equally interesting to comment on what is *not* there: *haji*, which is often the preferred Japanese translation for *shame* (Arimitsu, 2015; Lebra, 1983), does not appear in our list of collocates. The absence of *haji*, and the peripheral position of *zaiaku(kan)* and *shūchi(shin)*, suggest that there may be a gap between the terms used in the scientific and educational literature on the topic and those we find in naturally occurring data. This is not necessarily detrimental (it may be helpful to distinguish between theoretical notions and first-order terms). Yet, it is important to acknowledge that, while in English the everyday lexical items *shame* and *guilt* are also used to index the broader conceptual notions of SHAME and GUILT, in Japanese it may be appropriate to distinguish between a cluster of everyday terms (*mōshiwake*, *hazukashii*, *hansei*, *kōkai*) on the one hand, and two largely theoretical concepts (*shūchi(shin)/haji* and *zaiaku(kan)*) that laypeople do not really seem to talk (or write) about.

I have already commented on the association, in both Japanese and English, of SHAME- (which here is taken to include EMBARRASSMENT) and FEAR-related lexical items: in Japanese, *hazukashii* 'embarrassed/embarrassing' collocates with *kowai* 'scary' and *awateru* 'to panic/to rush', and in English *ashamed* collocates with *fear*, *afraid* and *horrify*. However, in English *embarrassed* also collocates with *angry* (example 6), pointing to a possible connection between SHAME and ANGER events that we do not find in the Japanese data. Previous psychological research on emotion across cultures (Boiger et al., 2013) found that ANGER is a condoned (desired) emotion in the US but condemned in Japan, and hence is experienced more frequently in the former context but suppressed in the latter. This could explain why we find ANGER words in the English network but not in the Japanese one.

- (6) The rain had smudged my sooty mascara and it was streaked down my cheek. I looked away quickly. [...] they all burst into laughter. I was **angry** and **embarrassed** in equal measure and hated them. I just turned and ran off.

Conversely, in the Japanese, but not in the English, data, GUILT is associated with the semantic sphere of 'responsibility', as both *jiseki* 自責 'self-reproach/remorse' and *semeru* 責める 'to blame' make use of the *kanji* 責, which we find also in *sekinin* 責任

'responsibility'. These patterns align with psychological research on the relatively more prominent role of agency and responsibility in triggering self-evaluative emotions in the Japanese context (Boiger et al., 2013, pp. 549–550).

Finally, and as already noted by Rusch (2004, p. 245), *hazukashii* can acquire a positive connotation, as when used in collocation with *ureshii* 'happy':

- (7) とても嬉しかった●嬉しいような恥ずかしいような笑っちゃう感じ  
*Totemo ureshikatta. Ureshii yōna hazukashii yōna waracchau kanji.*  
 'I was very **happy**. It was a feeling like being **happy** and **embarrassed**, and [that makes you] burst into laughter'.

This positive connotation is not observed in the English data and adds another layer of complexity to the overall picture.

## 6. Conclusion

This study employed a three-stage process to access the lived experiences of SHAME, GUILT and associated emotions, as they are verbalised through language in English and Japanese samples. The first stage entailed the collection of responses in English and Japanese to SHAME- and GUILT-inducing scenarios. These responses make up a sample of texts that are distinctively related to SHAME and GUILT and that we can analyse for linguistic traces of emotion, for example via keyword analysis, which corresponds to our second stage of investigation. By comparing the questionnaire responses with general web corpora, I identified words and expressions typically used in SHAME and GUILT situations. In the third and final stage, I selected six key emotion labels for each language and explored their (contextualised) collocational patterns in general corpora. The results were visualised in the form of collocational networks, with a focus on emotion co-occurrence.

My findings confirm that emotion co-occurrence is the norm in both English and Japanese narratives of SHAME and GUILT, and reveal both similarities and differences in the emotions that are typically associated with SHAME and GUILT. Specifically, FEAR and SADNESS seem to have a dominant role in the emotional landscape of SHAME and GUILT in both languages, although their level of saliency may vary. In English, we also find references to ANGER and DISGUST. It remains to be seen whether emotion co-occurrence typically involves two or more overarching emotions. The associations visualised in Figures 1 and 2 hint at possible correlations between SHAME (*ashamed*), EMBARRASSMENT (*embarrassed*) and ANGER (*angry*) in English, while EMBARRASSMENT (*hazukashii*), GUILT (*mōshiwake*) and HAPPINESS (*ureshii*) appear interconnected in Japanese. These triadic correlations need further investigation.

Finally, the collocational networks also suggest that, for corpus research that aims at investigating clusters of Japanese emotion labels primed for use in SHAME and GUILT scenarios, it may be more helpful and empirically valid to use search items like *hazukashii* for SHAME and *kōkai* or *hansei* for GUILT rather than *shūchi* (*shin*), *haji* and *zaiaku* (*kan*), which are largely used in the Japanese literature on the topic but were peripheral or completely absent in the emotional landscapes that emerged from my analysis.

A view of emotions as compositional and fuzzy entities with more and less prototypical members is promising and may lead to new insights into emotional experiences. Despite the limitations of the current study, I hope that the combination of methods and notions from different disciplines will foster a more integrated approach to the study of emotions across languages.

**Data availability statement.** The research questionnaires, full lists of keywords and collocates, analysis codes and other additional materials have been made publicly available on my [Open Science Framework](https://osf.io/54ghf/overview?view_only=50957184e12b4b2390a146137c073f47) page and can be accessed at [https://osf.io/54ghf/overview?view\\_only=50957184e12b4b2390a146137c073f47](https://osf.io/54ghf/overview?view_only=50957184e12b4b2390a146137c073f47).

**Acknowledgments.** I wish to thank Anna Marchi for reading an earlier version of the paper and Giuseppe Pappalardo for his support with the Chunagon. I am grateful to all the colleagues and friends who helped me distribute the questionnaire, as well as to the respondents for their participation. I also thank my two anonymous reviewers and the editors of this journal. All remaining inaccuracies are my own.

**Funding statement.** This study was carried out within the SELFEE project and received funding from the European Union NextGenerationEU – National Recovery and Resilience Plan (NRRP) – MISSION 4 COMPONENT 2, INVESTMENT 1.1 – CUP N.H73C24001500001. This manuscript reflects only the author's views and opinions; neither the European Union nor the European Commission can be considered responsible for them.

**Competing interests.** None.

## References

- Anthony, L. (2024). *AntConc (Version 4.3.1) [Computer software]*. Laurence Anthony's Software. <https://www.laurenceanthony.net/software/antconc/>
- Anthropic. (2025). Claude 4.5 Sonnet. <https://claude.ai>
- Arimitsu, K. (2015). Jiko ishiki-teki kanjō no keiken-teki teigi no gengo-kan hikaku [Cross-language comparisons of the empirical definitions of self-conscious emotions]. *The Japanese Journal of Research on Emotions*, 22(2), 53–59.
- Baker, C. F., Fillmore, C. J., & Lowe, J. B. (1998). The Berkeley FrameNet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 1* (pp. 86–90). Association for Computational Linguistics.
- 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>
- Bednarek, M. (2008). *Emotion talk across corpora*. Palgrave Macmillan. <https://doi.org/10.1057/9780230285712>
- Blum-Kulka, S., House, J., & Kasper, G. (Eds.). (1989). *Cross-cultural pragmatics: Requests and apologies*. Ablex Publishing Corporation. BNC Consortium. 2007. The British National Corpus, XML Edition. Oxford TextArchive: <http://hdl.handle.net/20.500.14106/2554>.
- Boiger, M., Mesquita, B., Uchida, Y., & Feldman Barrett, L. (2013). Condoned or Condemned: The Situational Affordance of Anger and Shame in the United States and Japan. *Personality and Social Psychology Bulletin*, 39(4), 540–553. <https://doi.org/10.1177/0146167213478201>
- Bostan, L., Kim, E., & Klinger, R. (2020). GoodNewsEveryone: A corpus of news headlines annotated with emotions, semantic roles, and reader perception. *arXiv*. <https://doi.org/10.48550/arXiv.1912.03184>
- BNC Consortium. 2007. *The British National Corpus, XML Edition*. Oxford TextArchive: <http://hdl.handle.net/20.500.14106/2554>.
- Creed, W. E. D., Hudson, B. A., Okhuysen, G. A., & Smith-Crowe, K. (2014). Swimming in a sea of shame: Incorporating emotion into explanations of institutional reproduction and change. *Academy of Management Review*, 39(3), 275–301. <https://doi.org/10.5465/amr.2012.0074>
- Culpeper, J. (2011). *Impoliteness: Using language to cause offence*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511975752>

- de Hooge, I. E., Zeelenberg, M., & Breugelmans, S. M. (2007). Moral sentiments and cooperation: Differential influences of shame and guilt. *Cognition and Emotion*, 21(5), 1025–1042. <https://doi.org/10.1080/02699930600980874>
- Diegoli, E. (2026). The affective meanings of bowing in a web corpus of Japanese. *Journal of Pragmatics*, 251, 14–29. <https://doi.org/10.1016/j.pragma.2025.11.002>
- Diegoli, E., & Öhman, E. (2024). Contrasting the semantic space of ‘shame’ and ‘guilt’ in English and Japanese. *Language and Cognition*, 16(4), 1296–1318. <https://doi.org/10.1017/langcog.2024.6>
- Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6(3–4), 169–200. <https://doi.org/10.1080/02699939208411068>
- European Commission (2021). *Ethics in social science and humanities (Technical Report)*. European Commission.
- Fehr, B., & Russell, J. A. (1984). Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology*, 113(3), 464–486. <https://doi.org/10.1037/0096-3445.113.3.464>
- Franzke, A. S., Bechmann, A., Zimmer, M., Ess, C., & Association of Internet Researchers. (2020). *Internet research: Ethical guidelines 3.0*. Association of Internet Researchers. <https://aoir.org/reports/ethics3.pdf>
- Fuoli, M., & Bednarek, M. (2022). Emotional labor in webcare and beyond: A linguistic framework and case study. *Journal of Pragmatics*, 191, 256–270. <https://doi.org/10.1016/j.pragma.2022.01.016>
- Galasiński, D., & Kozłowska, O. (2010). Questionnaires and lived experience: Strategies of coping with the quantitative frame. *Qualitative Inquiry*, 16(4), 271–284. <https://doi.org/10.1177/1077800409354068>
- Goldhahn, D., Eckart, T., & Quasthoff, U. (2012). Building large monolingual dictionaries at the Leipzig Corpora Collection: From 100 to 200 languages. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)* (pp. 759–765). European Language Resources Association (ELRA).
- Gu, S., Wang, F., Patel, N. P., Bourgeois, J. A., & Huang, J. H. (2019). A model for basic emotions using observations of behavior in *Drosophila*. *Frontiers in Psychology*, 10, 781. <https://doi.org/10.3389/fpsyg.2019.00781>
- Harder, D. W., & Zalma, A. (1990). *Personal Feelings Questionnaire-2*. American Psychological Association.
- Hoemann, K., Xu, F., & Barrett, L. F. (2019). Emotion words, emotion concepts, and emotional development in children: A constructionist hypothesis. *Developmental Psychology*, 55(9), 1830–1849. <https://doi.org/10.1037/dev0000686>
- Kikuchi, A. (2003). Kenkyū nōto: TOSCA-3 (tanshuku-ban) nihongo-ban no kentō [Research note: A study of the Japanese version of TOSCA-3 – short version]. *Iwate kenritsu daigaku shakai fukushi gakubu kiyō [Bulletin of the Faculty of Social Welfare, Iwate Prefectural University]*, 5(2), 35–40.
- Kikuchi, A., & Arimitsu, K. (2002). *KA-JiKoKan-12 (Technical Report)*. <https://jspp.gr.jp/doc/KA-Jikokan-12.pdf>
- Kikuchi, A., & Arimitsu, K. (2006). Construction of self-conscious emotion scale. *The Japanese Journal of Personality*, 14(2), 137–148. <https://doi.org/10.2132/personality.14.137>
- Kilgariff, A., Baisa, V., Bušta, J., Jakubiček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine. *Lexicography*, 1(1), 7–36. <https://doi.org/10.1007/s40607-014-0009-9>
- Kitamura, T., Hada, A., Usui, Y., & Ohashi, Y. (2024). Development of the Japanese version of the Test of Self-Conscious Affect-3 (TOSCA-3): A study among student and parent population. *Behavioral Sciences*, 14(7), 576. <https://doi.org/10.3390/bs14070576>
- Krawczak, K. (2014). Effects of speech rate, phonetic background and gender on vowel reduction in the speech of nonnative speakers of English. *Poznan Studies in Contemporary Linguistics*, 50(4), 441–475. <https://doi.org/10.1515/psicl-2014-0021>
- Krawczak, K. (2017). *Contrasting languages and cultures: A multifactorial profile-based account of SHAME in English, Polish, and French [preprint]*. <https://shs.hal.science/halshs-01464866v3>
- Krawczak, K. (2018). Reconstructing social emotions across languages and cultures. *French, and Polish. Review of Cognitive Linguistics*, 16(2), 455–493. <https://doi.org/10.1075/rcl.00018.kra>
- Langacker, R. W. (1987). *Foundations of cognitive grammar (Vol. 1). Theoretical prerequisites*. Stanford University Press.
- Lange, J., & Zickfeld, J. H. (2023). Comparing implications of distinct emotion, network, and dimensional approaches for co-occurring emotions. *Emotion*, 23(8), 2300–2321. <https://doi.org/10.1037/emo0001214>
- Lebra, T. S. (1983). Shame and Guilt: A psycho cultural view of the Japanese self. *Ethos*, 11(3), 192–210. <https://doi.org/10.1525/eth.1983.11.3.02a00070>

- Lewis, M. (2003). The role of the self in shame. *Social Research*, 70(4), 1181–1204. <https://doi.org/10.1353/sor.2003.0003>
- Maekawa, K., Yamazaki, M., Ogiso, T., Maruyama, T., Ogura, H., Kashino, W., Koiso, H., Yamaguchi, M., Tanaka, M., & Den, Y. (2014). Balanced corpus of contemporary written Japanese. *Language Resources and Evaluation*, 48(2), 345–371. <https://doi.org/10.1007/s10579-013-9261-0>
- Marchi, A. (2023). Get back! Methods for Exploring discourses of nostalgia and nostalgic discourses using corpora. *Elephant & Castle*, 31(31), 192–211. <https://doi.org/10.62336/unibg.eac.31.480>
- Orizu, U., & He, Y. (2016). Detecting expressions of blame or praise in text. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (pp. 4124–4129). European Language Resources Association (ELRA).
- Partington, A., & Diegoli, E. (2025). *Lexical priming: Evolution, evaluation and applications to English and Japanese* (1st ed.) Routledge. <https://doi.org/10.4324/9781003468578>
- Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In R. Plutchik & H. Kellerman (Eds.), *Theories of emotion* (pp. 3–33). Elsevier. <https://doi.org/10.1016/B978-0-12-558701-3.50007-7>
- Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, 13(4), 519–549. <https://doi.org/10.1075/ijcl.13.4.06ray>
- Riezier, K. (1943). Comment on the social psychology of shame. *American Journal of Sociology*, 48(4), 457–465. <https://doi.org/10.1086/219212>
- Rusch, C. D. (2004). Cross-cultural variability of the semantic domain of emotion terms: An examination of English shame and embarrass with Japanese hazukashii. *Cross-Cultural Research*, 38(3), 236–248. <https://doi.org/10.1177/1069397104264274>
- Russell, J. A. (1991). Culture and the categorization of emotions. *Psychological Bulletin*, 110(3), 426–450. <https://doi.org/10.1037/0033-2909.110.3.426>
- Scherer, K. R. (2000). Psychological models of emotion. In J. C. Borod (Ed.), *The neuropsychology of emotion* (pp. 137–162). Oxford University Press. <https://doi.org/10.1093/oso/9780195114645.003.0006>
- Schmid, H.-J. (2016). Why cognitive linguistics must embrace the social and pragmatic dimensions of language and how it could do so more seriously. *Cognitive Linguistics*, 27(4), 543–557. <https://doi.org/10.1515/cog-2016-0048>
- Stella, M., Swanson, T. J., Li, Y., Hills, T. T., & Teixeira, A. S. (2022). Cognitive networks detect structural patterns and emotional complexity in suicide notes. *Frontiers in Psychology*, 13(917630). <https://doi.org/10.3389/fpsyg.2022.917630>
- Tangney, J. P., & Dearing, R. L. (2002). *Shame and guilt*. Guilford Press.
- Tiedemann, J. (2012). Parallel Data, Tools and Interfaces in OPUS. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)* (pp. 2214–2218).
- Tracy, J. L., & Robins, R. W. (2007). The self in self-conscious emotions. In J. L. Tracy, R. W. Robins, & J. P. Tangney (Eds.), *The self-conscious emotions: Theory and research* (pp. 3–20). Guilford Press.
- Vannucci, A., Yu, W., Martin, N., Patel, S., & Tottenham, N. (2026). Affective schemas: Acquisition, updating, and inference. *Emotion*, 26(1), 52–73. <https://doi.org/10.1037/emo0001561>
- Wharton, T., & De Saussure, L. (2023). *Pragmatics and emotion* (1st ed.) Cambridge University Press. <https://doi.org/10.1017/9781108869867>

---

**Cite this article:** Diegoli, E. (2026). A linguistic analysis of SHAME and GUILT in English and Japanese – with a focus on emotion co-occurrence, *Language and Cognition*, 18, e35, 1–23. <https://doi.org/10.1017/langcog.2026.10087>