Shame and its Near-synonyms in English: A Multivariate Corpus-driven Approach to Social Emotions

Karolina Krawczak*

Résumé

Abstract
This chapter applies multifactorial corpus-driven methodology to the study of social emotions. It focuses on three lexemes designating shame in English, i.e., ashamed, embarrassed, and humiliated. It is hypothesized, in light of Wierzbicka’s findings, that there is a gradation of gravity and temporal boundedness from ashamed, through humiliated, to embarrassed. A large number of occurrences of the lexemes are submitted to a detailed qualitative analysis regarding semantic and sociolinguistic conceptual features. The data thus annotated are modelled quantitatively through statistical exploratory and confirmatory techniques. This reveals usage profiles specific to each emotion term, producing a conceptual map of the semasiological structure of, and onomasiological relations between, the expressions. The results confirm the hypothesis and contribute quantitative and falsifiable evidence for Wierzbicka’s introspection-based analysis.

1. Introduction
This chapter develops a quantitative corpus-driven methodology for the study of abstract social concepts. It is the first attempt to apply multifactorial feature analysis to the study of the conceptual structure of social emotions. It focuses

* Adam Mickiewicz University, Poznan, Faculty of English.
on shame emotions, as designated by three lexemes: ashamed, embarrassed, and humiliated. It is a cross-cultural comparison of how these emotions are conceptualized in present-day British and American fiction. In line with research in Cognitive Corpus Linguistics, it is assumed here that contextualized language structure provides access to conceptual structure, which, in turn, gives us insights into cultural models of reality. The methodology adopted in the present work is referred to as multivariate “usage-feature analysis” (Glynn 2009, 2010a, 2010b), which is also known as a “configurational approach” (Geeraerts et al. 1994) or “behavioural profile analysis” (Gries 2006).¹

Previous linguistic studies of social emotions serve as a theoretical frame of reference in which to formulate research questions and develop a set of conceptual features that play a role in the shame scenario. Shame emotions are genuinely social, as they lie at the intersection of our subjective and intersubjective experience of the world. As Wierzbicka (1999, 46) points out, they originate from the speaker’s sensitivity to others’ actual or potential criticism. In this context, they act as guidelines for social conduct by integrating the subject’s internal and external perspectives on any given situation.

This integration of two possible vantage points from which to consider unfolding reality results, conceptually speaking, in the general common denominator of shame, which has to do with how the subject “does not want other people to think something bad about him/her” (Wierzbicka 1999, 46; 1992, 574). Insofar as this overarching “conceptual component” is likely to be cross-culturally universal, shame can be ranked among basic social emotions (Wierzbicka 1999, 46ff.).

The three lexemes analyzed here are all interrelated semantically by the subject’s inherent wish not to be negatively evaluated. However, they also differ along a number of parameters. As Wierzbicka (1999, 47) indicates, shame is morally underpinned and can be experienced in solitude. This means that the mere awareness of having done something morally wrong or being socially inferior suffices for the subject to feel ashamed. This is different in the case of embarrassment and humiliation, both of which require an audience and which are therefore directly linked to the here and now. The lexical concept of ‘humiliation’, unlike ‘shame’, is always caused by external factors and its experiencer is, therefore, a “helpless victim” (Wierzbicka 1992, 575). That being so, humiliation will significantly damage the experiencer’s self-respect (ibid.). It can also outlast its triggering event.

¹ I wish to express my gratitude to Antti Arppe, Dylan Glynn, Kris Heylen, and Dirk Speelman for sharing with me their expertise in statistics.
and be experienced afterwards without any witnesses, thus affecting the subject for longer. Embarrassment, on the other hand, is instantaneous and fleeting, and, unlike either shame or humiliation, can also be linked to instances of positive “attention focused on us”.

At the heart of these “shame-like” (Wierzbicka 1999, 46) emotions lies their intersubjectivity. We can gain access to this dimension by employing a quantitative corpus-driven analysis. Corpora make available to us a large number of contextualized examples, which we can analyze for a range of formal, semantic, and social variables that determine meaning. Quantitative methodology, in turn, permits two things. Firstly, it allows us to visualize the results of the feature analysis, thereby revealing patterns of use which would not be retrievable with a purely qualitative approach. Secondly, it establishes the descriptive accuracy of the feature analysis. As a result, we can formulate probabilistic descriptions of, and identify statistically significant tendencies in, the linguistic behaviour of the expressions.

2. Method, data, and aims

In this chapter, we examine Wierzbicka’s introspection-based findings in the light of natural language use. In accordance with her assumptions, we make three hypotheses. First, the lexical concept of ‘shame’, related to moral vices, misdeeds, and perceived inferiorities, is expected to be a more grievous emotional state, one which is grounded on serious causes. Second, ‘embarrassment’ is anticipated to be more ephemeral and superficial. And, third, ‘humiliation’, which victimizes the experiencer, is hypothesized to lie between the gravity of ‘shame’ and the momentariness of ‘embarrassment’. Thus, we posit that there will be a continuum from ashamed, through humiliated, to embarrassed along the parameters of emotional weight and temporal boundedness. In what follows, we test the descriptive accuracy of Wierzbicka’s results by comparing them with the findings of a corpus-driven analysis of the use of the three shame lexemes. We will also see whether it is possible to establish any cross-dialectal differences between the concepts.

With regard to the data, this study is based on approximately 400 instances of ashamed, embarrassed, and humiliated. The samples were extracted randomly in equal numbers from the fiction sections of the British National Corpus (Davies 2004–) and the Corpus of Contemporary American English (Davies 2008–). The usage-feature method of corpus-driven semantic analysis, applied here, produces results that are entirely and exclusively determined by what is
found in the corpus. It also involves the meticulous manual annotation of a large number of examples for a spectrum of usage features. The number of conceptual factors according to which the uses of the lexemes were tagged was reduced to the semantic and sociolinguistic factors relevant to the questions posed. The first variable, coded automatically, distinguishes between British and American English.

The most critical factor in the study concerns the cause of the emotion. The specific causes recognized here are based on prior research conducted by Kövecses (1986, 44; 1990, 93f.), Tissari (2006, 150), Fabiszak and Hebda (2007, 29f.), and Krawczak (submitted). It subsumes the following levels: (1) bodily causes; (2) dubious social status; (3) inadequacy; (4) insecurity; (5) social failure; (6) social norm violation: (a) decency, (b) politeness, (c) emotional; (7) social status loss: (a) financial, (b) mistreatment, (c) rejection, (d) belonging to an unprestigious group. Let us illustrate these uses:

(1) She’d felt too humiliated about being heavy to buy maternity clothing when she was pregnant.
(2) “You said whatever you had to say to get elected. When I recall your memories, I feel ashamed.” She looked directly at him. “Don’t you?”
(3) “You can ask me anything you’d like to,” she says, then looks down, embarrassed by her brashness.
(4) And for all his bulk, he was gentle in that olive tree bed, quiet, shy, embarrassed by his own needs.
(5) Mom and Aunt Lily would be embarrassed at my bad playing, and I would be too.
(6a) She became addicted to drugs when she was thirteen and left home at seventeen, living on the streets, begging for money to support her habit, and I’m ashamed to say selling her body.
(6b) Gabriel […] said in a low voice, almost as if he was embarrassed, “Don’t say that. We don’t call them that anymore.” “Don’t say demon? I sure as hell am not saying ‘diabolically challenged’. That’s retarded.” “No,” Gabriel said. “[…] And you really shouldn’t say ‘retarded’ either. It’s not PC.”
(6c) I’m surprised to see tears glisten in the girl’s eyes and she quickly turns away, as though ashamed to reveal weakness.
(7a) She felt humiliated by her poverty.
(7b) “They treat you like – like shit”, Gloria said. “I’ve never been so humiliated.”
(7c) “I think he feels humiliated as much as grief-stricken. For an Arab to be refused by his amour…”
(7d) I took it pretty hard. My parents did too. After I got expelled from camp, they were humiliated. That kind of stuff doesn’t happen to Solomons.

Two more distinctions have been drawn with respect to the cause. First, regarding its temporal reference: (a) present (simultaneous with the emotion, see 3); (b) past
(prior to the emotion, see 7d); (c) general (inherent quality, atemporal). And, second, regarding its type: (a) internal (coming from within the experiencer or related to his/her actions); (b) external (coming from the outside world). The following sentences exemplify these categories:

(8) He was so humiliated by my performance that he couldn’t even look at me. (external)
(9) They knew why Vice stopped playing Little League […] because he was ashamed of the scars […] on his back. (internal, general)

Once all the examples had been manually annotated for the factors enumerated and illustrated above, two types of quantitative methods were used in order to reveal conceptual and cross-dialectal differences in use between the three shame expressions. Firstly, the data were treated with an exploratory statistical technique, in the form of Binary Correspondence Analysis (Glynn 2014). This method allows us to identify the “patterns of use” (Glynn 2009, 2010a, 2010b) for the lexemes on the basis of associations of conceptual usage features. This yielded three distinct usage profiles for ashamed, embarrassed, and humiliated, relative to dialect and the other conceptual features. The next step was to see how reliable these patterns of use are. With this in mind, the data were submitted to confirmatory statistical modelling in the form of Polytomous Logistic Regression (Arppe 2008), which is a method that establishes the descriptive accuracy of the analysis and its predictive power. The results of these statistical analyses are presented in Section 3.

3. Results

This section reports the findings of the cross-dialectal study of the semasiological and onomasiological structure of the three shame expressions. The results are, first, visualized by means of the exploratory method of Correspondence Analysis (Glynn 2014). This graphic representation shows the clustering of usage features associated with each lexeme. Next, the reliability and accuracy of the profiles thus obtained are determined with the confirmatory method of Polytomous Logistic Regression modelling (Arppe 2008). This shows how much confidence we can have in the identified conceptual associations and the manual feature analysis.

Let us look at the Binary Correspondence Analysis plot (Figure 1). It visualizes the clustering of features for the three lexemes relative to dialect, the cause of the emotion, as well as the type and temporal reference of the cause. In this technique, proximity is indicative of the degree of association. The accuracy of the graphic representation is evaluated on the basis of how much variation is explained in the first two dimensions. This is indicated by the percentages
given along the x and y axes. Naturally, the higher the sum of the two, the more reliable the emerging profiles are. Taking this into consideration, we can see that the plot in Figure 1 accounts for over 86% of the variation in the behaviour of the data in the first two dimensions. The patterns thus revealed are, therefore, stable.

Figure 1: Binary Correspondence Analysis

Three distinct clusters can be identified in Figure 1, each corresponding to one of the lexemes. The contribution of dialect to the distribution of the data points in this visualization is negligible, as the lexemes denoting the same emotion cluster together, with hardly any dialectal differences. The first group, located on the right, associates *ashamed* with eight usage features, five of which designate the actual causes of the emotion, while the remaining three specify the type and temporal scope of the triggering factor. With regard to the cause, this lexeme is distinctly correlated with (i) the violation of the social norm of emotional reaction (see 6c) and (ii) dubious social status (see 2). The other three causes linked to the lemma are (iii) social status loss due to financial reasons (see 10); (iv) bodily characteristics (see 9); and (v) failures, either personal or professional (see 11). These three usage features are placed further away from the centre of the cluster, constituted by the lexeme *ashamed* relative to dialect. This means that they are less pronounced in conceptual structure, being
also associated with *humiliated* (<Cause Failure> and <Cause Bodily>) and *embarrassed* (<Cause Financial>).

(10) Ain’t you ashamed, makin’ me wear things out of a pawnshop?
(11) I am ashamed. I let myself down vilely.

The cluster reveals that *ashamed* is also prominently linked to (vi) atemporal <Cause Time General> and (vii) internal <Cause Type Internal> causes (see 9 for both). The latter usage feature, however, is situated slightly to the left, thus also tending towards *embarrassed*. The final conceptual feature for *ashamed*, located on the periphery of this grouping, relates the lexeme to (viii) the past <Cause Time Past> (see 12). Its position shows clearly that it is shared between *ashamed* and *humiliated*.

(12) She rebuked herself, feeling deeply ashamed, for having given way earlier to despair and self-pity.

These findings confirm the hypothesis put forward in Section 2 in accordance with Wierzbicka’s (1999) claim – namely, that the lexical concept ‘shame’ has a much more grievous and deep-rooted basis than either ‘embarrassment’ or ‘humiliation’. Causes (ii)–(v) are serious and are likely to be long-lasting in their effect. Cause (i), although seemingly related purely to the here and now, may, in fact, be indicative of the subject’s inherent weakness, as in example (6c). If so, it may well be related to an underlying intrinsic character flaw, which increases its gravity. Given the nature of the factors triggering the emotion and the underlying character of shame, it is also intuitively interpretable that *ashamed* should be associated closely with internal and atemporal causes. Being inherent in the experiencer’s actions and unconstrained by time, they are, naturally, much more serious.

The second exponent of shame, *embarrassed*, forms a conceptual grouping of usage features in the lower left-hand quadrant. Irrespective of dialect, it is distinctively correlated with (i) social norm violation of politeness (see 6b) and (ii) insecurity (see 4). It also corresponds, but less closely, to (iii) inadequacy (see 3) and (iv) the violation of the social norm of decency (see 13).

(13) She was acutely embarrassed at having to share her cousin’s bed, especially as they would be squeezed so closely together in the small space.

As the plot indicates, *embarrassed* is also closely linked to (v) causes that are immanent in the event that gives rise to the emotion <Cause Time Present>. Causes (i)–(iv) exemplify this tendency. Arguably, all these causes, except personal
insecurities, are clearly the motivation for what Wierzbicka (1992, 576) refers to as “short-lived” emotional states. Hence, the hypothesis made in Section 2, that embarrassment is a passing emotion whose endurance is limited to the interactive situation that has engendered it, is substantiated. Insecurity as a cause of embarrassment upsets the consistency of this picture, however, since it is more ingrained and so may seem to be an enduring basis for the emotion. Yet, it is always awoken by specific circumstances and, when the triggering factor stops operating, the insecurity may well become dormant again, as evidenced in example (4). This lexeme is also attracted, though less closely, to three other usage features: (vi) internal (see 3 or 4), (vii) external (see 13), and (viii) financial (see 14) causes, which it shares with both ashamed (vi and viii) and humiliated (vii).

(14) The Parsons and the Carters were incapable of conceiving that anyone could be financially embarrassed by a lunch bill.

It stands to reason that embarrassment can be triggered by internal factors that inhere in the experiencer him or herself or in his or her actions. Likewise, it does not come as a surprise that we can also become embarrassed as a result of external causes, as illustrated in (13) or (14).

The last clustering, for humiliated, is situated in the top left-hand quadrant. The lexeme, regardless of dialect, is associated distinctly with (i) the loss of social status due to rejection (see 7c) and (ii) mistreatment (see 7b). It is also related, though less saliently, to the feature that identifies the source of the cause as (iii) external <Cause Type External> (see 7b). The placement of feature (iii) in the plot, nonetheless, is indicative of its simultaneous association with embarrassed, as pointed out above. Humiliated is also related to (iv) failures (see 8) <Cause Failure>. The first two causes are external and the experiencer in both cases is definitely a victim, which corroborates Wierzbicka’s point (1992) with respect to humiliation. The situation is slightly different when it comes to failure, though, which is not necessarily caused by external factors. However, similarly to the two other causes, it is characterized by helplessness, which, as stated in Section 1, is typical of this emotion (Wierzbicka 1992, 575). The feature of loss of social status due to lack of prestige is placed almost in the middle of the plot, which means that it is distributed equally between the lexemes and is, therefore, indistinct in its effect on the conceptual structuring of the data.

The above Correspondence Analysis has established three very clear conceptual profiles for the three lexemes, which correspond to the hypotheses made in Section 2 and to Wierzbicka’s claims (1992, 1999). The next step is to determine
the descriptive accuracy and predictive power of these findings and of the usage-feature analysis. This is done by using Polytomous Logistic Regression Analysis, as developed by Arppe (2008). This technique is designed for response variables that have more than two features, as here, where the variable lexeme subsumes three levels: ashamed, embarrassed, humiliated. A response variable is the output whose behaviour in the sample is to be predicted on the basis of the explanatory variables. Since the Correspondence Analysis plot demonstrates that dialect does not contribute much to the emerging usage profiles for the three lexemes, this factor, for the sake of simplicity and due to space limitations, will be disregarded in the confirmatory analysis. Table 1 establishes which factors are significantly associated with which lexeme.

Formula: Lexeme ~ Cause_Type + Cause + Cause_Time

Log-odds:

<table>
<thead>
<tr>
<th></th>
<th>ashamed</th>
<th>embarrassed</th>
<th>humiliated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cause Time: General</td>
<td>0.4832</td>
<td>1.058</td>
<td>-1.173</td>
</tr>
<tr>
<td>Cause Time: Present</td>
<td>-1.591</td>
<td>2.697</td>
<td>(-0.6417)</td>
</tr>
<tr>
<td>Cause Type: Internal</td>
<td>0.9401</td>
<td>-0.505</td>
<td>-0.4819</td>
</tr>
<tr>
<td>Cause: Dubious Social Status</td>
<td>0.8243</td>
<td>1.743</td>
<td>-2.685</td>
</tr>
<tr>
<td>Cause: Failure</td>
<td>-0.919</td>
<td>1.018</td>
<td>0.1341</td>
</tr>
<tr>
<td>Cause: Inadequacy</td>
<td>-0.5589</td>
<td>2.028</td>
<td>-0.7558</td>
</tr>
<tr>
<td>Cause: Insecurity</td>
<td>-0.8584</td>
<td>3.77</td>
<td>-17.47</td>
</tr>
<tr>
<td>Cause: Norm Violation - Decency</td>
<td>-0.2288</td>
<td>2.009</td>
<td>-1.114</td>
</tr>
<tr>
<td>Cause: Norm Violation - Emotional</td>
<td>0.3946</td>
<td>1.701</td>
<td>-1.646</td>
</tr>
<tr>
<td>Cause: Norm Violation - Politeness</td>
<td>-0.9253</td>
<td>3.311</td>
<td>-2.329</td>
</tr>
<tr>
<td>Cause: Status Loss - Financial</td>
<td>-0.1655</td>
<td>1.686</td>
<td>-0.8655</td>
</tr>
<tr>
<td>Cause: Status Loss - Mistreatment</td>
<td>-1.827</td>
<td>-0.1759</td>
<td>1.375</td>
</tr>
<tr>
<td>Cause: Status Loss - Rejection</td>
<td>-1.589</td>
<td>0.454</td>
<td>0.9152</td>
</tr>
<tr>
<td>Cause: Status Loss - Unprestigious</td>
<td>-0.9302</td>
<td>1.957</td>
<td>-0.3031</td>
</tr>
</tbody>
</table>

R2.McFadden: 0.3072687
R2.Nagelkerke: 0.5522443
C: 0.8392158

Table 1: Polytomous Logistic Regression Analysis

The column on the left lists all the predictors. The remaining three columns, each corresponding to one of the three lexemes, specify the degree of correlation between the given lexeme and the explanatory feature. This is done by means of log-odds. It is only the parenthesized scores that are significant. Among these, the higher the log-odds, the greater the effect of the feature on the lexeme, with positive values indicating association and negative values, dissociation. Bearing
this in mind, we can see that the important predictors for ashamed are the lack of attraction of the lexeme to (i) present causes, i.e., those that are simultaneous with the emotion; (ii) its relatedness to internal causes, inherent in the experiencer or resulting from his/her actions (see 9); and (iii) its dissociation from the loss of social status due to mistreatment. The second expression, embarrassed, is strongly correlated with (i) insecurities, (ii) the violation of the social norm of politeness, and (iii) present causes, coexisting with the emotion (see examples 4 and 6b). Finally, humiliated is shown to be unrelated to (i) dubious social status, (ii) the violation of politeness rules, and (iii) general or atemporal causes. These tendencies confirm some of the patterns revealed at the exploratory stage in Figure 1, thus supporting the hypotheses put forward in Section 2. The stability and predictive power of the model are determined by the C statistic, while the strength and effect of the explanatory variables are evaluated by the pseudo $R^2$ scores, all provided at the bottom of Table 1. The rule of thumb is that a value for the C statistic of 0.8 and more means that the model has “predictive capacity” (Baayen 2008, 204). The Nagelkerke $R^2$ should be at least 0.3 to be satisfactory (Lattin et al. 2003, 486), whereas the McFadden $R^2$ should be at least 0.2 (McFadden 1979, 307). The results in Table 1 are, therefore, sound, which means that the profiles identified here are reliable and the model is strong in its explanatory potential.

4. Summary and conclusions

This study has demonstrated that the conceptual architecture of lexemes designating social emotions can be determined by applying quantitative corpus-driven methodology. More specifically, it has been established that identifying semantic-pragmatic features of use embedded in co-text reveals the semasiological structure of, and onomasiological relations between, near-synonyms. The multivariate approach adopted here has allowed us to distinguish three clear usage profiles for the three lexical exponents of shame. Ashamed forms a conceptual cluster with internal and atemporal causes concerning a dubious social status, emotional or bodily problems, and social failures. Embarrassed is closely related to factors triggered by the interactive situation such as personal insecurities or a breach of politeness norms. Humiliated, in turn, is associated mainly with external factors causing loss of status. These statistically determined conceptual tendencies substantiate the hypotheses formulated in Section 2 and add quantitative evidence to Wierzbicka’s claims (1999, 1992). The lexical concept ‘shame’ is shown to be a serious, potentially destructive and long-lasting state,
while ‘embarrassment’ is a more ephemeral emotion linked to the situation engendering it. The lexical concept ‘humiliation’, in turn, lies between the two other categories, hinging on the here and now, but possibly outlasting the immediate instigating event.

References


Davies, Mark (2004–). BYU-BNC. (Based on the British National Corpus from Oxford University Press). <http://corpus.byu.edu/bnc/>


