



HAL
open science

A corpus-assisted approach to conventionalised apologies in Japanese and English computer-mediated discourse

Eugenia Diegoli

► **To cite this version:**

Eugenia Diegoli. A corpus-assisted approach to conventionalised apologies in Japanese and English computer-mediated discourse. 2022. hal-04172047

HAL Id: hal-04172047

<https://hal.science/hal-04172047>

Preprint submitted on 28 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**A corpus-assisted approach to conventionalised apologies in Japanese and English
computer-mediated discourse**

Eugenia Diegoli

eugenia.diegoli2@unibo.it

University of Bologna

Abstract

This study provides a corpus-assisted contrastive investigation of the apology-like expressions *su(m)imasen* ‘(I’m) sorry’ in Japanese and *sorry* in English. The two expressions are investigated in computer-mediated discourse (CMD, Herring 2005), as represented in the web corpora JaTenTen11 and EnTenTen20, which are provided by Sketch Engine and are seen as comparable. Using collocational analysis and close reading of the extended concordance lines, quantitative (statistical) and qualitative analysis are combined. In this understanding, speech acts are investigated not as single units detached from the context, but as part of longer stretches of text, in an effort to go beyond the merely linguistic level of analysis to access the social and interactional dimension of communication. The analysis revealed that both *su(m)imasen* and *sorry* are recurrent and conventional practices that frequently co-occurs with two types of items, namely elements referring to the object of regret (i.e., the reason for apologising; Coulmas 1981) and intensifiers. However, while *su(m)imasen* is often accompanied by visual elements and metalinguistic descriptors (*henji* ‘answer’, *shitsumon* ‘question’, etc.), *sorry* is frequently observed in conjunction with verbs of perception (*feel*, *hear*). *Sorry* also exhibits more diversified patternings in terms of semantic collocates, a finding suggesting that apologies in CMD may be semantically more conventionalised in Japanese than in English. The close reading of the extended concordance lines also showed that *su(m)imasen* and *sorry* exhibit different pragmatic functions. The findings have both methodological and theoretical

implications for the ways in which we attempt to understand how interactional practices differ (or not) across lingua-cultures.

Keywords: cross-cultural pragmatics, (im)politeness, digital speech acts, apologies, corpus-assisted discourse analysis, computer-mediated discourse

1. Introduction

The present study provides a corpus-assisted contrastive pragmatic analysis of the apology-like expressions¹ *su(m)imasen* ‘(I’m) sorry’ in Japanese and *sorry* in English as observed in two web corpora provided by Sketch Engine. Semantically, both *su(m)imasen* in Japanese and *sorry* in English are not truth-conditional, in that they could not be ‘true’ or ‘false’. For instance, if we assume that in general there is complete agreement in the judgment of colours, the statement “this pen is red”, in a situation where the speaker is holding a pen, can be either true, if the pen is indeed red, or false, if the pen is not red. However, apologies need a different characterisation, as it becomes apparent in the English construction “I apologise”, which can hardly be said to be true or false. Their being not truth-conditional is also related to the fact that apologies do not contribute to the denotational meaning of the utterance and their value in terms of information transmission is secondary at best. Rather, their relevance lays in the social and interpersonal level of language, in that (generally speaking) they contribute to smooth communication and ‘doing what is right’ in a given context. However, in order to do so, their use needs to meet a number of necessary conditions – what Austin’s (1962) refers to as *felicity conditions*. The notion of felicity conditions (although quite problematic as they may be met only loosely) is noteworthy, in that it presupposes an explicit link connecting linguistic expressions, circumstances of use and meanings. In Austin’s words, “it is always necessary that

¹ This terminology emphasises that the use of a *form* usually associated with ‘apologies’ does not automatically determine the illocutionary force of the act.

the *circumstances* in which words are uttered should be in some way, or ways, *appropriate*” (1962: 8; original emphasis). Although the felicity conditions as outlined by Austin (1962) have shown to be too rigid to adequately describe real world language usage, they laid the foundations for an approach to language where the context is key to determine the appropriateness of linguistic choices. This paper deals precisely with the appropriateness – which, as will be seen, blends into conventionalisation – of apology-like expressions in Computer Mediated Discourse (henceforth CMD, see Herring 2004, 2005). By focusing on two quite different languages, the study aims at identifying similarities and differences in the use(s) and function(s) of apology-like behaviour across languages.

The paper is organised as follows. Section 2 presents the theoretical background I draw from. It starts by introducing early works on politeness where apologies are conceptualised as a form of polite language aimed at conflict avoidance, before presenting two alternative and related approaches (namely the frame-based view and the notion of *wakimae*). Section 3 describes the methodology and presents the two corpora employed as data sources. The reasoning behind the choice of written, online data is presented, and the interdisciplinary approach that combines principles and methods of corpus linguistics, discourse analysis and pragmatics is also addressed. The main findings are illustrated in Section 4. Finally, Section 5 concludes the paper by summarising the argument and proposing a number of insights for future research.

2. Theoretical background

A number of important works have focused on the uses and functions of *su(m)imasen* and *sorry* (see Coulmas 1981; Ide 1998; Obana 2015; Oishi 2013 for *sumimasen*; Blum-Kulka & Olsthain 1984; Blum-Kulka, House & Kasper 1989; Deutschmann 2003; Holmes 1990; Leech 2014 for *sorry*, among others). It should be noted that *su(m)imasen* contains the formal verbal suffix *-masu* (*-masen* being its negative form), thus it is slightly more formal than its English counterpart. Nonetheless, among the various apology-like expressions employed in Japanese,

su(m)imasen is the closest one to *sorry* in terms of contexts of use and formality level, as showed in previous comparative studies (Spencer-Oatey 2008; Sugimoto 1998). Their assumed similarities and the large body of works I can draw from to analyse *su(m)imasen* and *sorry* in a comparative perspective motivates the choice of these two lexemes over seemingly analogous terms (e.g., *gomen[nasai]* ‘(I’m) sorry’ in Japanese). Nevertheless, functionally related lexemes may be worth elaborating on in future research (a point I will go back to in the concluding section).

In contrast with previous works, the present study focuses specifically on online contexts where, as will be seen, apologies often deviate from the classical model of speech acts as instances of remedial behaviour aimed at conflict avoidance. It may thus be useful to briefly outline classical approaches to seemingly polite utterances before proceeding with the analysis.

2.1. *Classical approaches to politeness*

Early frameworks on politeness (Lakoff 1973; Leech 1983; Brown & Levinson 1987) explained the use of seemingly polite linguistic resources in terms of tradeoff between being clear and avoiding conflict. Such approaches, and Brown and Levinson’s (1987) seminal work in particular (as its title “Some *universals* in language usage” clearly shows), are based on the assumption that it is possible to apply the same general principles to virtually all languages. These general principles have been referred to as “maxims” (Leech 1983), “politeness strategies” (Brown & Levinson 1987), “pragmatic constraints” (Leech 2007), etc. and are reminiscent of Austin’s (1962) felicity conditions, in that they try to locate a limited number of principles to explain how communication works. A second key element of early approaches to politeness is that they tend to equate politeness with indirectness (see also Searle 1975), on the basis that a less straightforward utterance leaves more freedom of choice to the addressee, as in the classical example “Can you pass me the salt?” as opposed to “Pass me the salt” (Searle 1969). This is what Leech refers to when he talks about “optionality” (2014: 154–155), and can

be put on a par with Brown and Levinson's negative politeness strategies, some of which explicitly include giving options (1987, 172). However, these early works have been subsequently criticised as Anglocentric and, as such, not adequate for describing language use in non-Western societies (see Matsumoto 1989; Wierzbicka 1991; Ide 1989, among others). I adopt Leech's (2007) view that, in so far as we see the politeness principles theorised in these classical works not as absolute but as relative (and thereby subject to cross-cultural variation), they still provide a useful toolkit for describing what is going on when some degree of deviation from optimally efficient communication is observed. However, we will see that the use of apology-like expressions in my data does not seem to support the idea that rationality, indirectness and non-imposition are at the basis of language choices – although to different degrees in Japanese and English. Thus, a different approach which can better account for the data at hand is needed.

2.2. *Frame-based view*

Marina Terkourafi (2001, 2003, 2005, 2015) proposes an alternative approach which associates politeness with conventionalisation – rather than, for instance, indirectness. This is usually referred to as the frame-based approach to politeness (see Culpeper 2011), and it is the framework I primarily draw from in the current study. Terkourafi defines politeness as a perlocutionary effect originated by the producer's utterance meeting the receiver's general expectations about how language is commonly used (and thereby *should* be used) "given the setting [...] and the activity interlocutors are engaged in" (Terkourafi 2001: 26). We can empirically observe the situations in which interactants' expectations are met because the performance of conventionally polite behaviours in the appropriate context often goes unchallenged – i.e., it is not explicitly acknowledged by the other participants in the interaction (see also Watt's [2003] notion of *politic* behaviour, which overlaps with unmarked politeness). To assess whether a linguistic form is (im)polite we have to take into account the interactional

context where it is commonly employed, as it is precisely the co-occurrence of linguistic forms with specific contexts that gives rise to perceptions of politeness in terms of unmarked behaviour. Then, we can use unmarked behaviour as the point of reference to assess whether (a) someone does extra politeness work which results in surplus politeness (positively evaluated marked behaviour); or, (b) someone fails to meet other people's expectations (negatively evaluated marked behaviour). The latter can be further divided in unconscious and conscious deviations from standards of politeness and can be brought about either by "not putting in the work" or by doing something potentially impolite.

To explain the type of knowledge required for recognising what counts as a typical co-occurrence of linguistic forms and contexts of use, Terkourafi draws from the notion of 'frame', namely a structure that "combine[s], as separate components, information about a situation [...] with information about the appropriate use of language therein" (Terkourafi 2001: 25)². Since frames are experientially and socially acquired, they are both grounded on one's individual experience of the world and shared across members of a given group. The frame-based approach to politeness has, in my view, three main advantages. First, it avoids the conceptualisation of politeness in terms of indirectness (as it is now generally assumed that more indirectness does not necessarily equate with more politeness). Second, it emphasises the experiential and societal nature of language use, in line with a discursive approach to communication. Finally, and related to this second point, it encompasses group-oriented politeness while at the same time allowing for individual variability. From a methodological point of view, it also means that we can quantitatively identify sites of potential (im)politeness by counting how many times a given linguistic form is used in a certain context.

2.3. Wakimae

² Among the first to elaborate on the notion of frame is Minsky (1975, 1977) in his work on Artificial Intelligence, but since its introduction the term has been used in many different ways (see Bednarek 2005 for an overview).

Terkourafi's conceptualisation of politeness as the regular co-occurrence of linguistic forms with specific contexts is reminiscent, in some respects, of Ide's notion of *wakimae* 'discernment'. *Wakimae* is defined as "the practice of polite behaviour according to social conventions" (Ide 1989: 230) and, according to Ide, accommodates the use of polite speech in Japanese better than Brown and Levinson's (1987) rational face-saving strategies. Although Ide's analysis emphasises (often hierarchical) relationships between interactants, it shares with Terkourafi's work the idea that social conventions (and not a limited set of rational principles) are at the core of (im)politeness. Importantly, however, the two approaches are not mutually exclusive: when we monitor the appropriateness of our behaviour we can also employ highly conventionalised linguistic resources as a result of a rational choice. Just as there is no divide between strategic and conventional (im)politeness, it seems to me that there is no divide between West and East either and that production and evaluation of (im)politeness in both contexts can be studied through the lenses of the same framework – a point that recurs a number of times in the following sections. We can thus imagine *wakimae* and strategic politeness as two ends of a continuum, along which different behaviours and different lingua-cultures can be located.

Following the assumptions made above, in the next section I present a corpus-assisted contrastive pragmatic investigation of the apology-like expressions *su(m)imasen* '(I'm) sorry' in Japanese and *sorry* in English. I focus specifically on standard situations where *su(m)imasen* and *sorry* regularly co-occur with particular situational contexts (and specific linguistic contexts). As already noted, it is precisely the recurrent combination of linguistic formulae with minimal contexts that justifies their categorisation as conventional (thus potentially polite) practices. Their relevance lies in the fact that they are "interpersonal rituals" (Goffman 1982: 57) that may provide insights into culture-specific behavioural norms and expectations about what counts as (im)polite, which, in turn, mirror and maintain moral orders.

3. Methodology

Methodologically speaking, the present study adopts an approach that falls within the area described as corpus-assisted discourse studies (see Partington, Duguid & Taylor 2013) and combines quantitative (statistical) and qualitative analysis. In this understanding, speech acts are investigated not as single units detached from the context, but as part of longer stretches of text, in an effort to go beyond the merely linguistic level of analysis to access the social and interactional dimension of communication. This methodological approach guided the choice of data and written online texts were selected for two main reasons. First, they allow me to experience the data without missing important extra-linguistic features that are often lost in other types of discourse, for instance when spoken conversations are transcribed. Second, it is a context very different from the face-to-face dyadic conversation traditional studies on speech acts focus on and, as such, may uncover ways of doing that might have been previously overlooked. I thus selected as data sets two web corpora provided by Sketch Engine, namely the JaTenTen11 and the EnTenTen20 for Japanese and English respectively. Although quite different in size, with the English corpus having almost four times the tokens of the Japanese one (40+ billion versus 10+), the two corpora are prepared according to the same criteria and can be regarded as comparable corpora (see <https://www.sketchengine.eu/docwhiumentation/tenten-corpora/> for more information on the TenTen corpus family). These two corpora include to some degree more “traditional” types of discourse (e.g., online newspapers and face-to-face conversations reported and recontextualised in online settings), but an exploratory analysis of the types of website where *sorry* and *su(m)imasen* were retrieved from seems to support the assumption that they are primarily employed in written interactive contexts, such as online forums or blogs. More systematic analysis is needed to confirm this (a point further explored in the concluding section), but preliminary evidence suggests that the TenTen corpus family is representative of the type of discourse I am interested in – namely, written online interaction.

Drawing from big corpora has a number of advantages, closely related to one another. First, it gives me access to information about the frequency and the typicality of the expressions chosen as search items – two aspects in which introspection is particularly unreliable. Second, it allows me to rely on a wider range of tools, such as collocations, which are quite meaningless with very restricted data sets. Third, it generally increases the reliability of my findings, since it prevents me from, consciously or unconsciously, manipulate the data or throwing away inconvenient data (see Partington, Duguid & Taylor 2013: 331). As will be seen, however, a qualitative approach is also required for the study of highly context-dependent pragmatic meanings. It is hoped that the combination of the two may place this study on a sounder footing.

As for the steps taken, first I identified the most typical collocates of *sorry* and *su(m)imasen* (span L5-R5, ranked by logDice) and I grouped the collocates in semantic categories. The grouping in semantic categories is aimed at shedding light on the degree of conventionalisation of the two expressions (with less semantic variation associated with a higher degree of conventionalisation). I then used the concordancer to look at the extended concordance lines of *su(m)imasen* and *sorry* so as to verify the assumptions made by observing the lists of collocates and access their function(s) in context. This qualitative approach to the data is also motivated by the fact that, as already touched upon in Section 2, if we are to adopt a frame-based approach, we need to take into account the co-occurrence of linguistic resources and specific contexts of use, as it is only the combination of the two that allows for the unchallenged realisation of particular speech acts. As Culpeper puts it, “a simplistic quantitative methodology, such as counting up a particular form, is not possible, as we must count up forms in particular contexts that are unchallenged” (2011: 26). It follows that automatic means of investigation alone are not sufficiently accurate for investigating this subject, which requires some degree of manual analysis. Since the line-by-line reading of concordance lines and the manual categorisation that follows is very time consuming, I limited the analysis to 50 occurrences for each apology-like expression, which were downloaded and placed into an Excel spreadsheet.

While I am aware that the limited number of instances investigated does not allow for a high degree of generalisability, it is hoped that the high level of accuracy can at least provides some insights and pointers for future research.

4. Findings

4.1. Collocational analysis of *su(m)imasen*

The collocational analysis revealed that *su(m)imasen* typically co-occurs with four types of items: metalinguistic descriptors (‘answer’, ‘question’, etc.), items related to the ‘object of regret’ (Coulmas 1981, namely the reason for apologising) (‘long’, ‘late’, etc.), intensifiers and visual elements (i.e., visual forms of expression characteristic of computer-mediated communication, see Miyake 2020). The first 20 collocates of *su(m)imasen* are illustrated in Figure 1, while in Table 1 the collocates are transcribed, translated and grouped according to their semantic category.

Word	Cooccurrences ?	Candidates ?	T-score	MI	LogDice ↓
1 □ 遅く	9,765	304,831	98.67	9.34	8.61 ...
2 □ 長文	6,031	89,093	77.60	10.42	8.36 ...
3 □ 長々	4,956	42,711	70.37	11.19	8.19 ...
4 □ すみ	6,536	379,632	80.61	8.44	7.91 ...
5 □ ; ;	5,065	193,481	71.03	9.04	7.88 ...
6 □ すい	5,288	256,422	72.54	8.70	7.82 ...
7 □ ;	8,960	902,117	94.18	7.65	7.70 ...
8 □ でし	49,195	7,403,459	220.15	7.07	7.67 ...
9 □ ><	4,944	300,050	70.10	8.38	7.64 ...
10 □ 質問	9,754	1,140,985	98.19	7.43	7.60 ...
11 □ 遅れ	4,661	411,644	67.97	7.84	7.37 ...
12 □ ほんと	6,857	882,217	82.28	7.29	7.33 ...
13 □ 返事	4,140	334,130	64.09	7.97	7.33 ...
14 □ 長く	4,583	543,274	67.30	7.41	7.15 ...
15 □ 汗	5,067	712,558	70.69	7.16	7.08 ...
16 □ ホント	4,514	696,233	66.67	7.03	6.94 ...
17 □ レス	3,418	410,396	58.12	7.39	6.92 ...
18 □ はい	4,410	729,872	65.86	6.93	6.86 ...
19 □ 本当	13,522	3,352,784	114.86	6.35	6.84 ...
20 □ z	4,986	916,537	69.97	6.78	6.84 ...

Figure 1. First 20 collocates of *su(m)imasen* (L5-R5) in the JaTenTen11

Table 1. Collocates of *su(m)imasen* (L5-R5) grouped by semantic category

Metalinguistic descriptors	Object of regret	Intensifiers	Visual elements	Other
10. <i>shitsumon</i> ‘question’	1. <i>osoku</i> ‘late’	4. <i>sumi[masen]</i> ‘sorry’	5. ; ;	8. <i>deshi[ta]</i> COP.PST
13. <i>henji</i> ‘reply’	2. <i>chōbun</i> ‘long message’	6. <i>sui[masen]</i> ‘sorry’	7. ;	18. <i>hai</i> ‘yes’
17. <i>resu</i>	3. <i>naganaga</i>	12. <i>honto</i>	9. ><	

‘reply’	‘at length’	‘really’	
	11. <i>okure[ru]</i>	16. <i>honto</i>	15. <i>ase</i>
	‘be late’	‘really’	‘sweat’
	14. <i>nagaku</i>	19. <i>hontō</i>	20. <i>z</i>
	‘at length’	‘really’	

The collocation analysis seems to suggest that apologies in Japanese CMD relate almost exclusively to features of the message itself, which seems to be qualified either as a ‘question’ (*shitsumon* ‘question’) or as a ‘reply’ (*henji*, *resu* ‘reply’). For instance, people apologise for writing at length (*naganaga*, *nagaku* ‘at length’, *chōbun* ‘long message’) or for replying late (*osoku* ‘late’, *okure[ru]* ‘be late’) to a previous message. This was confirmed by a subsequent analysis of the co-text surrounding each pair of collocates (namely, *sumimasen* + one of its collocates). For instance, the collocation *osoku* + *sumimasen* with up to 5 interviewing words seems to be usually employed in the construction

- (a) (o-)henji (ga) osoku-(nat)te sumimasen
(HON-)reply (SBJ) (be.)late sorry
‘Sorry for the late reply’

Of note is also the use of intensifiers of the apology, either in the form of an intensifying adverb or of repetition of the apology-like expression. The third column from the left in Table 1 illustrates the two most frequent transcription of *su(m)imasen* and three different orthographical transcription of the intensifying adverb *hontō* ‘really’, namely in *hiragana* and *katakana* (two moraic alphabets), and *kanji* (Chinese characters). Interestingly, the standard transcription *hontō* in *kanji* seems to be the less common among the three. Intensifiers of the apology increase the perceived sincerity level of the utterance (Vollmer & Olshtain 1989: 213; Aijmer 1996: 82) and are associated with what Leech calls positive politeness (1984: 84). Notably, the use of intensifiers is associated in the literature with serious offences whose face-damaging potential is rather high (Vollmer & Olshtain 1989: 198). However, their frequency

in my data suggests that, at least in this context specifically, they may actually be the unmarked form of apology, which is primarily intended to satisfy social needs, and only secondarily associated with some form of (probably little) regret. Finally, the wide range of visual elements is worth commenting on. In fact, within the 20 most typical collocates of *su(m)imasen* we can observe: (1) the semicolon ‘;’ either to visually conveying the act of ‘crying’/‘sweating’ or as part of *kaomoji* (Japanese emoticons) such as m (; ▽ ;) m, where the two ‘m’ characters index the hands to visually conveying the body movement of bowing; (2) the *kaomoji* > <, which stands for the eyes; and (3) the *kanji* 汗, employed in the *kanji* emoticon (汗) where the ideograph 汗 *ase* ‘sweat’ is enclosed between parenthesis³. This is in line with findings from previous works on Japanese online communication (e.g., Miyake 2019: 176) showing that visual elements convey the producer’s attitude (Halliday 1994) towards or feelings about what they are talking about. Hence, they perform an important interpersonal function (in this case, saving face by showing consideration to the receiver’s needs). It also suggests that the use of semantically restricted linguistic collocates may be counterbalanced by the exploitation of visual semiotic resources that, although conventionalised to some degree, are creatively manipulated to communicate feelings and emotions more accurately. In light of these considerations, we can safely assume that, once again, there is no hard line between conventionalisation and creative, strategic choices.

4.2. Collocational analysis of sorry

The collocational analysis of *sorry* reveals a more varied picture than the one observed for its Japanese counterpart (Figure 2). Once again, in Table 2 I attempted to group the collocates in semantic categories so as to make comparison more meaningful.

³ Their use as part of more complex visual elements (and not as isolated punctuation marks) was confirmed by a subsequent analysis of their extended concordance lines.

	Word	Cooccurrences ?	Candidates ?	T-score	MI	LogDice ↓							
1	'm	365,528	16,641,558	603.42	9.02	9.34 ...	11	inconvenience	8,024	127,897	89.52	10.53	7.07 ...
2	hear	59,536	3,938,061	243.32	8.48	8.40 ...	12	missed	13,307	1,442,185	114.83	7.77	7.06 ...
3	am	153,337	13,835,239	390.08	8.03	8.33 ...	13	delay	11,754	1,067,754	108.00	8.02	7.06 ...
4	sorry	18,549	1,168,132	135.83	8.55	7.66 ...	14	say	62,212	14,457,543	246.97	6.67	6.97 ...
5	Hi	19,311	1,479,260	138.51	8.27	7.58 ...	15	i	44,348	9,987,136	208.58	6.71	6.94 ...
6	feel	61,128	8,925,858	245.71	7.34	7.54 ...	16	forgot	8,240	449,414	90.56	8.76	6.89 ...
7	Oh	22,503	2,419,418	149.33	7.78	7.44 ...	17	loss	20,512	4,091,679	142.01	6.89	6.83 ...
8	Im	10,545	390,623	102.53	9.31	7.28 ...	18	felt	21,277	4,431,838	144.58	6.82	6.80 ...
9	im	10,444	541,694	101.97	8.83	7.17 ...	19	ca	31,727	7,826,949	176.26	6.58	6.75 ...
10	guys	17,485	2,369,848	131.47	7.44	7.09 ...	20	I	792,416	241,226,237	878.69	6.28	6.74 ...

Figure 2. First 20 collocates of *sorry* (L5-R5) in the EnTenTen20

Table 2. Collocates of *sorry* (L5-R5) grouped by semantic category

<i>I am</i> variants	Object of regret	Intensifiers	Expressions of perception	Other
1. 'm	11. inconvenience	4. sorry	2. hear	2. Hi
3. am	12. missed	7. Oh	6. feel	
8. Im	13. delay	10. guys	18. felt	
9. im	14. say			
15. i	16. forgot			
20. I	17. loss			
	19. ca[n't]			

As expected, Table 2 shows that *sorry* is typically used in the construction *I am sorry* (or its contracted variants). A second finding is that items associated with the object of regret (i.e., the reason for apologising) are more varied than their Japanese counterparts, as they do not relate only to features of the message itself – in fact, no metalinguistic descriptors referring to qualities of the message were observed within the 20 most typical collocates. As for the use of intensifiers, following Vollmer and Olshtain (1989), I considered as forms of intensification also exclamations ('Oh') and vocatives ('guys'). In English as well as in Japanese, intensification is often achieved through the repetition of the apology-like expression. The main difference with Japanese is illustrated in the fourth column from the left, namely the widespread use in English online apologies of verbs of perception ('hear', 'feel', etc.). With reference to 'feel' (and its past form 'felt'), an analysis of the extended concordance lines revealed that it is virtually exclusively used in the expression of sympathy *feel (really) sorry for* someone, along

the lines of another quite relevant patterning, namely *sorry to hear that*. It should also be noted that a closer look at the concordance lines also showed that the verb ‘say’ is usually employed in the phraseology *(I’m) sorry to say it/this (but)* as a preface for potentially impolite behaviour. These functions are only loosely related to the apologetic meaning of *sorry* and shed light on the multifunctionality of apology-like expressions – a point which is further investigated in the following section.

4.3. Close-reading of concordance lines

The collocational analysis suggests that the quantitative difference in degrees of apology conventionalisation between English and Japanese may not be the whole story, and that qualitative differences in terms of pragmatic functions could also be observed. To shed further light on this matter, I read line-by-line the first 50 extended concordance lines of *su(m)imasen* and *sorry* and I categorised their pragmatic functions. Since examples classified as being invalid were excluded (i.e., duplicates, metareferences to morphological features of the expression itself and ambiguous instances), in a later stage the invalid examples were substituted with valid ones that occurred after the batch of 50 examples in the corpus. Importantly, I didn’t differentiate between instances where the apology-like expression is actually ‘used’ in the context at hand and reported speech (i.e., when the producer of the utterance is reporting an exchange that took place prior to the interaction in question) or metapragmatic comments (i.e., “[s]ituated comments about or evaluations of language use” [Culpeper & Haugh 2014: 241] where the apology-like expression is the object of reflection). It follows that both reported speech and metapragmatic comments were included into the analysis, and, when possible, their function was categorised according to the way in which it was framed by the participants in the interaction. The results are illustrated in Figure 3, while Table 3 gives one example for each function.

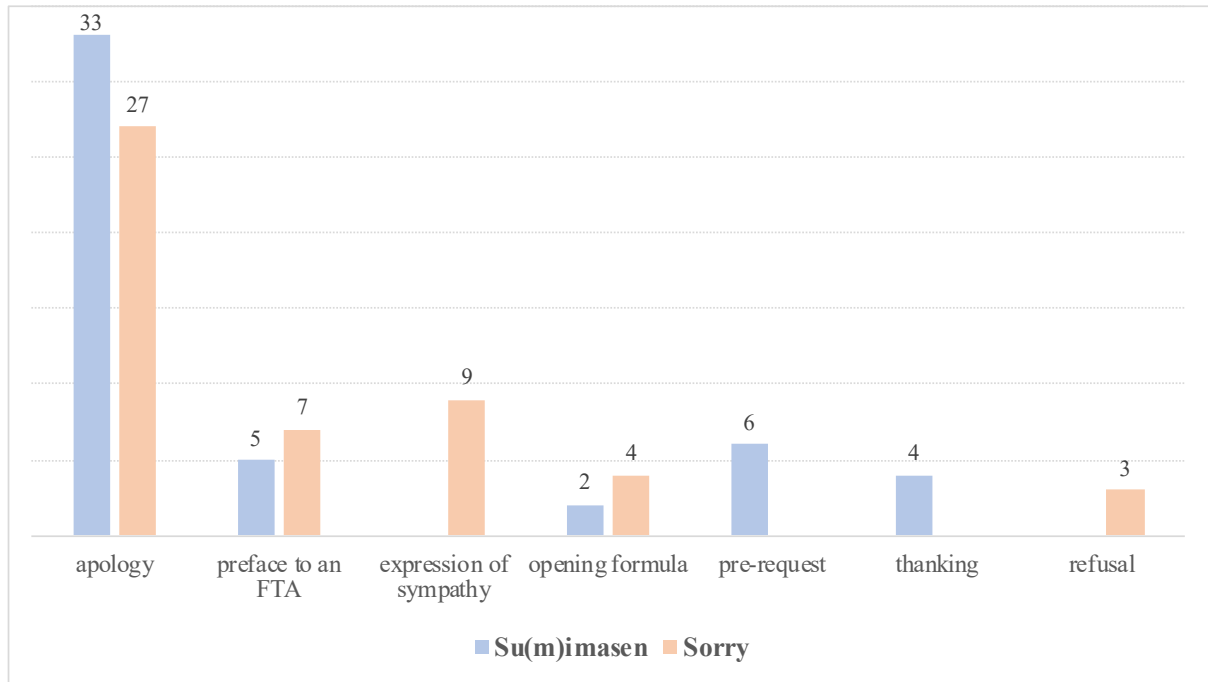


Figure 3. Pragmatic functions of *su(m)imasen* and *sorry*

Table 3: Pragmatic functions with examples

Function	Japanese	English
Apology	<i>Bangō machigaemashita. sumimasen.</i> <i>shitsurei shimasu†</i> ‘I called the wrong number. Sorry. Goodbye †’	Sorry for the inconvenience. Seems to be a few issues here and there. Please mention issues in the comments
Preface to an FTA	<i>Sumimasen ga, o-hanashi shitaku nai n desu.</i> ‘Sorry, but I don’t want to talk’	Sorry to say it, but there's much more of a 'situation' in Gilmer County than most people realize, and it goes MUCH deeper than just the school system
Expression of sympathy	//	I’m sorry to hear that your mother has MDS. You are doing exactly the right thing by seeking information
Opening formula	<i>Shitsumon o shitai toki ni wa</i> <i>‘sumimasen’ to hitokoto kotowatte kara</i> <i>hanashi hajimeru</i> ‘When [the students] want to ask a	Sorry, to clarify, I also meant that there's a 6vs2 forcemod is 233% between the two army last cycle

	question they start talking only after saying “sorry”	
Pre-request	<i>Sumimasen, boku me ga warui n de, ano seki ni shite-kudasai</i>	//
	‘Sorry, I have a low vision, please let me sit there’	
Thanking	<i>Wazawaza sumimasen</i>	//
	‘*Sorry/thank you for taking the trouble’	
Refusal	//	"I want this done in half an hour." "Sorry, the printer needs an hour for this bulk job, at the very minimum."

Figure 3 shows that, although apology is the most frequent speech act associated with both expressions, the implementation of other functions is more culturally defined. For instance, expressions of sympathy such as *I'm sorry to hear that* were quite common in the English corpus, but do not have a counterpart in Japanese. The use of apology-like expressions to open the ground for a potentially impolite behaviour, as in the English construction (*I'm*) *sorry to say this but*, confirmed the assumption already made in the previous section that this is a fairly common use of *sorry*. Drawing from Leech (2014), I have labelled this type of instances as ‘Preface to a face-threatening act (FTA)’, where a FTA (Brown & Levinson 1987) is an act that is more likely than others to indicate that the producer is not taking into consideration the receiver’s wants or desires. For instance, if we assume that agreement between two parties is, generally speaking, the desired course of action (see Leech’s maxim of agreement [1983: 2014]), the English example reported in Table 3 can be considered as a FTA because the ‘apology’ is immediately followed by a statement that seems to contradict what has come before. This supports the already quite established idea that that seemingly polite linguistic devices do not automatically produce polite utterances. It is also in line with Goffman’s (1982: 14) claim that face-work is not limited to a protective orientation towards others’ face, but

includes also a self-defensive orientation and more or less deliberate offensive behaviour. Importantly, the fact that FTAs are conveyed through recurrent patternings also suggests that impoliteness, as well as politeness, has its own repertoire of standardised practices. As for the functions observed in Japanese, in three instances of *su(m)imasen* a feeling of guilt blurs into an expression of gratitude (e.g., *wazawaza sumimasen* ‘*sorry/thank you for taking the trouble’⁴) – something that was not observed in English. Similarly, the use of an apology-like expression as a pre-request was found only in the Japanese data, but this is most probably due to the very limited number of occurrences analysed, as instances such as *sorry say that again* are perfectly plausible in English as well. The same is true for refusals, observed here only in English, but which may be conveyed in Japanese as well, for instance through the wording *sumimasen, chotto yōji ga atte...* ‘I’m sorry, I have something to do... [I can’t come]’.

4.4. A note on conventionalised apologies

With reference to apologies specifically, another important finding is that many instances in my data set could be associated with highly conventionalised behaviour closer to “conversational routines” (Coulmas 1981) than to sincere expressions of regret. Two examples are as follows:

(25) Sorry for the inconvenience. Please see the updates page to see the list of restaurants that need to be added

(53) *Konbanwa. resu arigatō-gozaimasu. o-henji ga osoku natte sumimasen.*

Good evening. Thank you for your reply. Sorry for the late response.

In both examples we do not have any explicit admission of responsibility but only a general specification of the reason for apologising. The impression is that the apology is more a matter

⁴A quick collocational analysis (https://ske.li/wazawaza_collocations) confirmed that *wazawaza* is usually associated with expressions of gratitude, with three different transcriptions of *arigatō* ‘thank you’ within its 100 most typical collocates (at the 3rd, 37th and 60th position; L5-R5). No expression of apology was observed within this range.

of routine, part of what is expected in this specific context, rather than a device indicating that the producer admits their responsibility for the offence. This type of apology makes up for the bulk of my data, a finding that conflicts with the general tendency in the literature to conceptualise apologies as more elaborated and ‘ad hoc’ speech acts, as in:

(43) Sorry, our server choked on your last request. This is most likely our fault, not yours. Please try again in few minutes. If the problem persists, leave a note below and we'll review the issue.

The example above is what I refer to when I talk about a ‘prototypical’ apology, where the remedial function of *sorry* is emphasised by the use of additional apologetic strategies, namely an explicit admission of responsibility (*this is most likely our fault*) and some form of commitment to a future remedial action (*we'll review the issue*). Example (43) illustrates that this type of apology is indeed present in my data, but quite rare (although it should be kept in mind the very limited size of the sample). Its salience in a multitude of contexts (crisis situations, political speeches, legal settings, and so on) makes it the frequent object of reflexive discussions. However, my findings seem to suggest that, although very relevant in terms of markedness and potential real-world consequences, prototypical apologies may not be as frequent as we may think. In other words, we tend to assume that the most representative (i.e., more similar to the prototype as illustrated in traditional speech act theory [Austin 1962; Searle 1969]) instances of the category ‘apology’ are also the most frequent, but that may not be case – or at least it wasn’t in my data. A likely explanation behind the occasionally overlooked role and frequency of conventionalised apologies in everyday interaction is that this type of apologies partly loses its remedial meaning because it is highly ritualistic, functioning primarily as a supportive move. It follows that, if the contextual features allow for such interpretation, they can be considered as instances of unmarked politeness (Watt’s [2003] *politic* behaviour) that often passes unnoticed. This is not to say, however, that they are not an important feature

of discourse. On the contrary, conventionalised apologies play a key role in communication, because their absence may lead to evaluations in terms of impoliteness and, consequently, interactional disruption. On a slightly different note, knowledge about how things are normally done is also required to identify sites where hearers' or readers' expectations of regularity are creatively exploited for strategic purposes.

5. Conclusions

This study examined similarities and differences in the use(s) and function(s) of *su(m)imasen* and *sorry* with the aim to shed light on what is considered (in)appropriate in terms of apology-like behaviour in Japanese and English respectively. A comparative analysis of the two expressions was carried out in two main stages. First, I carried out a collocational analysis to shed light on the contexts and functions *su(m)imasen* and *sorry* are associated with. The results showed that they tend to co-occur with items related to a limited number of semantic categories. More specifically, both *su(m)imasen* and *sorry* are typically associated with items specifying the reason for apologising (i.e., the 'object of regret' [Coulmas 1981]) and intensifiers of the apology. However, a number of differences were also observed. For example, *su(m)imasen* seems to be more restricted in terms of semantic collocates than its English counterpart, with the object of regret almost exclusively related to specific features of the apologetic message itself (e.g., being late, too long, etc.). Thus, if we conceptualise highly conventionalised behaviour (*wakimae*) and rational politeness as two ends of a spectrum, from a semantic perspective online apologies in Japanese are closer to the former than their English counterparts. Interestingly, however, Japanese also exhibits a more varied use of visual elements, suggesting that many factors are at play when communicating online and that it is not possible nor helpful to draw a clear line between conventional and strategic (im)politeness. On the other hand, *sorry* correlates with variants of the construction *I am* and, more interestingly, with verbs of perceptions, hinting at differences on a functional level too. This was confirmed in the second

stage of the analysis by the close reading of 50 concordance lines for each expression. The concordancing showed that, if the context allows for such interpretation, *su(m)imasen* can blur into an expression of gratitude. In contrast, *sorry* is quite frequently employed as an expression of sympathy. Both expressions, although seemingly polite, can also function as a preface for potentially offensive behaviour. Finally, I argued that the role of highly conventionalised apologies may have been overlooked in the literature, which tends to focus on more ad-hoc apologies accompanied by an explicit admission of responsibility.

These findings support two theoretical claims. First, that conventionalisation, which manifests itself through recurrent and standardised practices, is at the basis of linguistic behaviour and evaluations in terms of (im)politeness (Terkourafi 2001): apologies (and related speech acts) often become interactional rituals and are perceived as polite only if they match interactants' expectations given the context at hand. This is in line with recent developments in the field of psychology proving that we tend to associate what is familiar (i.e., regular) with what is safe and true, thus, positively evaluated (Kahnemann 2011). The exploitation of conventionally polite linguistic expressions for opening the ground to potentially face-damaging behaviour adds an additional layer of complexity that calls for further investigation. The second claim is that "there is no absolute divide between East and West in politeness" (Leech 2007: 170), but rather a continuum of behaviours with (often slightly) different degrees of conventionalisation and various pragmatic functions. Provided that we approach the data with the appropriate (combination of) theories and methods and that we take into account the wider discursive context where the linguistic interaction takes place, different languages can be compared. In this view, even if speech acts are necessarily intertwined with the wider socio-cultural system, the underlying principles that motivates (im)politeness in the first place are shared across lingua-cultures.

A number of questions remains to be addressed in future research. For instance, I have mentioned in Section 3 that apology-like expressions seem to be primarily used in blogs and

forums. However, no systematic analysis was carried out and, unfortunately, Sketch Engine does not provide detailed information on where the texts come from. Future studies may attempt to compare apologies across different computer-mediated genres, or test the findings of the present study on more controlled data sets with similar genres, aims and audiences. These constraints would represent a common ground which reduces the variables we have to take into account, hence making comparison more meaningful (see Connor & Moreno 2005: 154). A second limitation is related to the limited number of apology-like expressions analysed. The study could be further developed by considering the set of apology-like expressions in both English and Japanese (e.g., by including *gomen[nasai]* ‘(I’m) sorry’ in Japanese and *excuse me* in English), on the assumption that it is ultimately better to compare fields of related terms in two different languages rather than two single terms. Finally, our knowledge of how different languages employ seemingly similar linguistic resources would be enriched by looking at apology-like expressions in other lingua-cultures, ideally as observed in corpora that can be compared with those employed here. Despite the large number of areas of research that warrant further investigation, it is hoped that the present study, where methods of corpus linguistics and discourse analysis are applied to classical pragmatic phenomena as observed in online discourse, provided a number of scattered insights which may be worth pursuing. Finding a balance between, on the one hand, the assumption that Western practices are universal, and, on the other, a cultural relativism that does not allow meaningful comparison, is a great challenge indeed, but could have a refreshing effect on the theoretical and methodological insights we can produce.

Bibliography

- Aijmer, K. (1996). *Conversational routines in English: Convention and creativity*. London; New York: Longman.
- Austin, J. L. 1962. *How to do things with words*. Oxford: Oxford University Press.

- Bednarek, M.. 2005. Frames revisited—The coherence-inducing function of frames. *Journal of Pragmatics*, 37(5), 685–705. <https://doi.org/10.1016/j.pragma.2004.09.007>
- Blum-Kulka, S., & Olshtain, E. 1984. Requests and Apologies: A Cross-Cultural Study of Speech Act Realization Patterns (CCSARP)1. *Applied Linguistics*, 5(3), 196–213. <https://doi.org/10.1093/applin/5.3.196>
- Blum-Kulka, S., House, J. & Kasper, G (eds.). 1989. *Cross-cultural pragmatics: Requests and apologies*, 155–173. Norwood, NJ: Ablex.
- Brown, P., & Levinson, S. P. 1987. *Politeness. Some universals in language usage*. Cambridge: Cambridge University Press.
- Connor, U., & Moreno, A. I. 2005. Tertium comparationis: A vital component in contrastive rhetoric. In P. Bruthiaux, D. Atkinson, W. Eggington, W. Grabe, & V. Ramanatan (eds.), *Directions in Applied Linguistics: Essays in Honour of Robert B. Kaplan* (pp. 153–164). Multilingual Matters. <https://doi.org/10.21832/9781853598500-015>
- Cook, H. M. 2018. Speech style shift. In Hasegawa, Y. (ed.), *The Cambridge Handbook of Japanese Linguistics* (1st ed.). Cambridge: Cambridge University Press. 628–648. <https://doi.org/10.1017/9781316884461>
- Coulmas, F. (ed.), *Conversational Routine: Explorations in Standardized Communication Situations and Prepatterned Speech*, 69–91. Berlin & New York: De Gruyter Mouton.
- Culpeper, J. 2011. *Politeness and Impoliteness*. Berlin: Mouton de Gruyter.
- Culpeper, J., & Haugh, M. 2014. *Pragmatics and the English Language*. 235–263. London: Bloomsbury Publishing Plc.
- Deutschmann, M. 2003. *Apologising in British English*. Umeå: Umeå Universitet.
- Goffman, E. 1982. *Interaction ritual: Essays on face-to-face behavior*. New York: Anchor Books.
- Halliday, M. A. K. 1994. *An Introduction to Functional Grammar*. London: Edward Arnold.
- Haugh, M. 2013. Speaker meaning and accountability in interaction. *Journal of Pragmatics*,

48(1), 41–56. <https://doi.org/10.1016/j.pragma.2012.11.009>

Herring, S. 2004. Computer-Mediated Discourse Analysis: An Approach to Researching Online Behavior. In S. Barab, R. Kling, & J. H. Gray (eds.), *Designing for Virtual Communities in the Service of Learning*, 338–376. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511805080.016>

Holmes, J. 1990. Apologies in New Zealand English. *Language in Society*, 19(2), 155–199. <https://doi.org/10.1017/S0047404500014366>

Ide, R. 1998. ‘Sorry for your kindness’: Japanese interactional ritual in public discourse. *Journal of Pragmatics* 29(5). 509–529. [https://doi.org/10.1016/S0378-2166\(98\)80006-4](https://doi.org/10.1016/S0378-2166(98)80006-4)

Ide, S. 1989. Formal forms and discernment: two neglected aspects of universals of linguistic politeness. *Multilingua* 8(2,3). 223–248.

Kahneman, D. 2012. *Thinking, fast and slow*. London: Penguin Books.

Lakoff, R. 1973. Language and woman's place. *Language in Society*, 1(2), 45–80.

Leech, G. 1983. *Principles of pragmatics*. London; New York: Longman.

Leech, G. 2007. Politeness: Is there an East-West divide? *Journal of Politeness Research. Language, Behaviour, Culture* 3(2). 167–206.

Leech, G. 2014. *The pragmatics of politeness*. Oxford: Oxford University Press.

Matsumoto, D. 1989. Cultural Influences on the Perception of Emotion. *Journal of Cross-Cultural Psychology*, 20, 92-105.

Minsky, M. 1975. A framework for representing knowledge. In Patrick Winston (ed.), *The Psychology of Computer Vision*, 211–277. New York: McGraw-Hill.

Minsky, M. 1977. Frame-system theory. In Philip Johnson-Laird & Peter Wason (eds.), *Thinking. Readings in Cognitive Science*. 355–376. Cambridge: Cambridge University Press.

Miyake, K. 2019. Mobairu media ni okeru hairyō – LINE no irai danwa no tokuchō – [Expressions of consideration in mobile media – features of request exchanges in LINE –]. In Yamaoka Masaki (eds.), *Nihongo hairyō hyōgen no genri to shosō* [aspects and principles of

expressions of consideration in Japanese]. 163–180. Tokyo: Kuroshio Shuppan.

Miyake, K. 2020. Evolution of emoji and beyond: A diachronic observation of visual representations in Japanese mobile media. *The Bulletin of the Institute of Human Sciences, Toyo University* 21. 1–15.

Obana, Y. 2015. Kansha hyōgen toshite no ‘arigatō’ to ‘sumimasen’ no kyōkaisen: shinborikku sōgo sayō riron o tekiyō [The boundary between ‘arigatō’ and ‘sumimasen’ as expressions of gratitude: application of the symbolic interaction theory]. *Gengo to bunka* [language and culture]. 15–28.

Oishi, E. 2013. Apologies. In Sbisà, M. & Turner, K. (eds.), *Pragmatics of Speech Actions*, 523–554. Berlin ; Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110214383.523>

Partington, A., Duguid, A. & Taylor, C. 2013. *Patterns and Meanings in Discourse: Theory and practice in corpus-assisted discourse studies (CADS)*. Amsterdam. John Benjamins.

Searle, J. 1969. *Speech acts: an essay in the philosophy of language*. Cambridge: Cambridge University Press.

Terkourafi, M. 2001. *Politeness in Cypriot Greek: A Frame-based approach*. Cambridge: University of Cambridge dissertation.

Terkourafi, M. 2003. Generalised and particularised implicatures of linguistic politeness. In P. Kühnlein, H. Rieser, & H. Zeevat (eds.), *Pragmatics & Beyond New Serie*, 149–164. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/pbns.114.09ter>

Terkourafi, M. 2005. Beyond the micro-level in politeness research. *Journal of Politeness Research. Language, Behaviour, Culture* 1(2). 237–262.

Terkourafi, M. 2015. Conventionalization: A new agenda for im/politeness research. *Journal of Pragmatics*, 86, 11–18. <https://doi.org/10.1016/j.pragma.2015.06.004>

Vollmer, H. J., & Olshtain, E. 1989. The language of apologies in German. In Blum-Kulka, S., House, J. & Kasper, G (eds.). 1989. *Cross-cultural pragmatics: Requests and apologies*, 155–173. Norwood, NJ: Ablex.

Watts, R. 2003. *Politeness*. Cambridge & New York: Cambridge University Press.

Wierzbicka, A. 1991. Japanese Key Words and Core Cultural Values. *Language in Society* (20)3. 333–385.