

Emoticonsconsciousness

Carl Vogel* and Jerom F. Janssen

Computational Linguistics Group, O'Reilly Institute,
Trinity College Dublin, Dublin 2, Ireland
{vogel,janssenj}@tcd.ie

Abstract. A temporal analysis of emoticon use in Swedish, Italian, German and English asynchronous electronic communication is reported. Emoticons are classified as positive, negative or neutral. Postings to newsgroups over a 66 week period are considered. The aggregate analysis of emoticon use in newsgroups for science and politics tends on the whole to be consistent over the entire time period. Where possible, events that coincide with divergences from trends in language-subject pairs are noted. Political discourse in Italian over the period shows marked use of negative emoticons, and in German and Swedish, positive emoticons.

Keywords: emoticons, intercultural analysis, extra-linguistic cues.

1 Introduction

It has been noted of conversation that in different linguistic communities, verbal and nonverbal feedback patterns vary. In a comparison of verbal interactions between Swedish and Italian interlocutors [4] it has been recorded that there is far more likely to be overlap of primary dialog contributions in Italian than in Swedish, and conversely longer pauses between turns in Swedish conversations than Italian. With respect to nonverbal communication, it is noted that Japanese and Swedish cultures exhibit less eye contact than typical Greek communications, although perhaps with different associations with eye contact between Japanese and Swedish cultures, and instead employ greater levels of verbal than visual feedback [1]. A question then arises about what communication patterns will emerge in communicative settings that lack an auditory channel, but whose visual channel is still primarily linguistic, through reading.

Other recent work uses emoticons to classify overall sentiment of documents [10,12,8] or to associate emoticons with words in texts [9]. The work reported here constitutes our initial experiments in attempting to analyze inter-cultural differences in emoticon use. One comparable study is a within-culture analysis of emoticon use which addresses the differences in emoticon use in online discussions in Greek, depending upon whether the forum [13] has access to the Greek alphabet or is confined to a Roman alphabet. Other work examines comparable

* This research is supported by Science Foundation Ireland RFP 05/RF/CMS002. We are grateful to anonymous reviewers for constructive critique of an earlier version of this paper, but take full responsibility for the content of this version, errors and all.

online prose [11], and forms demographic generalizations about blog postings (e.g. teenage bloggers tend to be women; older bloggers, men, etc; more personal blogging among women, and more blogging on science and technology by men; etc.) taking into account features like inclusion of URLs (this propensity increases with age), but not emoticon use and not with an inter-cultural dimension. In contrast, other recent work has examined gender presentation in Italian and Chilean chat rooms [5], specifically identifying significant differences in emoticon use cross-culturally. While there are related studies, the analysis of inter-cultural differences in expressing sentiment using emoticons in on-line asynchronous communication is in its infancy.

In this paper, we examine informal written communication in electronic media. We focus on the forums for asynchronous exchange provided by Usenews groups. Emoticons are analyzed as a sort of non-linguistic visual feedback mechanism in written media. We want to know whether intercultural differences in verbal and non-verbal feedback from other media transfer to asynchronous electronic communication. Recently, an analysis of emoticon use in this context has been described [6]. The results presented there considered about 400,000 postings from late September 2006 to the end of December 2007 in four linguistic communities: German, Italian, Swedish and English. Two topic areas were analyzed: science and politics. With respect to politics, the Swedish discussion was more likely to include positive emoticons than negative or neutral emoticons, and the Italian postings were more likely to include negative emoticons than the others. Discussions in science newsgroups showed more positive emoticons than anything else for German, Italian and English, and more neutral emoticons for Swedish. The results presented in §2 summarize the research methods and findings from past analysis [6]. The classification of emoticons as conveying positive, negative or ambiguous sentiment was based on only our own judgments. This is expanded to address group classification of the emoticons using two different methods. The main result persists in that the Italian messages use the least proportion of positive emoticons. That study was based on an aggregation of the data over the 66 weeks sampled. The role of the present paper is to show how the data distributed over time, to demonstrate that the qualitative tendencies named above are not localized to a short time frame within the data.

2 Background

Usenews groups were sampled from a server fed by the HEANET in Ireland. Binaries were filtered by the source. Server configuration issues entail that the entirety of the remainder was not received, but messages were distributed over the monitored newsgroups without evidence of selection criteria. Spam was filtered locally using Spam-Assassin. Data on Swedish and Italian were sought as language sources for which we had *a priori* reason from other communication channels to expect differences, as mentioned above. English and German were included as baseline and contrast sources. The subdomains *.swnet, *.se, *.it, *.de and *.uk provided postings representative of the corresponding languages.

Table 1. Messages per language per topic

Language	Topic	Messages	APPI	Language	Topic	Messages	APPI
Swedish	Politics	18225	23.13	Italian	Politics	173672	32.94
Swedish	Science	814	5.73	Italian	Science	32117	5.97
<i>Sum:</i>		19039		<i>Sum:</i>		205789	
German	Politics	933	3.30	English	Politics	81635	10.90
German	Science	75230	12.72	English	Science	13561	10.66
<i>Sum:</i>		76163		<i>Sum:</i>		95196	

We did not classify or filter data further with a language guesser [3]; further, we do not presume that everyone who posts within the *.de domain is German, or correspondingly for any of the other areas. The topic areas which had coverage for all four languages during the sampled period included those in science and politics. We did not examine topics at any more fine-grained level because of data sparseness. After filtering, 396,187 postings remained, distributed across languages and topics sampled as shown in Table 1. The average number of postings per individual (APPI) provides a coarse metric of newsgroup interactivity. Analysis of emoticon use as a function of interactivity has only just begun [6].

A list of 2,161 unique emoticons with their descriptions was compiled from two web sources.¹ We added three more classes of emoticons consisting of three or more consecutive characters that are all exclamation marks, or all question marks, or a mixture, with prototypical members: “!!!”, “???” and “!?!?”. It is contentious to think of tokens like “???” as emoticons. Some researchers refer to this general category more specifically as containing *emodevices* with coding criteria that overlap with those of emoticons.² One study compares Italian and Chilean chat rooms for a number of textualized conversational moves and found significantly greater use in Italian groups of both emoticons and emodevices [5]. Instances of ALLCAPS or tokens like “!#@\$” are taken to be emodevices. Emoticons are distinguished in that work as iconic. However, emoticons like emodevices clearly also make “irregular” use of characters to form emotive tokens. The distinction between the categories is not so sharp. First of all, the use of the exclamation mark in a token like “!#@\$” is, in fact, rather regular in comic strips (e.g. *Beetle Bailey*). As words have primary and secondary senses, so also letters and punctuation marks have alternative but conventionalized uses. Secondly, if pressed to say what ALLCAPS or “!#@\$” is iconic of, one defending their iconicity might refer to the representation of a sound intensity, and then have to defend the applicability of “icon” to audible as well as visual stimuli.³

¹ One was <http://www.gte.us.es/~chavez/Ascii/smileys.txt> — last verified March 2008; the other was <http://www.windweaver.com/emoticon.htm> — last verified March 2008.

² The *ProjectH Codebook*: <http://www.it.murdoch.edu.au/~sudweeks/projecth/codebook.html> — last verified August 2008.

³ It is relevant that items on computer desktops for invoking software are called “icons” even when they lack both forms of iconicity.

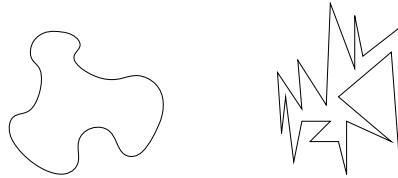


Fig. 1. Which is the “krikatorik” and which is the “souble”?

An argument for the appropriateness of this use of “icon” is in an old example in which image is taken to correspond to sound: reliably, presented with images like those in Figure 1, people will judge the item on the right to be the krikatorik and defend the judgement with a statement like “because ‘krikatorik’ *sounds* sharp”. We do not argue that sound is the same as image, but that iconicity takes in both modalities. Hence, we include emodevices under the category of emoticons.

A within-Japanese qualitative study of *kaomoji*⁴ used by young married women addresses the rather ornate instances used to as extra-textual bonding devices [7]. In contrast, our study focuses on the most frequently occurring emoticons in the texts sampled, which correlates directly with the least ornate. Of the 2,161 tokens considered as emoticons (including some *kaomoji*) only 121 actually occurred; the 12 most frequent are indicated with their raw frequencies in Table 2. Our parsing of the emoticons sought longest possible matches, so that, for example, the frequency of “:-)” is independent of that of “:-))”.

These emoticons were classified as positive, negative or neutral/ambiguous. We analyze the distribution of positive, negative and neutral emoticons as a function of language and subject area. The initial classification of the emoticons is our own, indicated in Table 2 under the heading “Us”. We evaluated our judgments by presenting the 100 most frequent of the 121 emoticons to raters who could choose one of three categories (positive, negative or neutral) with instructions that “neutral” should be selected if the emoticon is deemed either not interpretable or ambiguous between positive and negative.⁵ Emoticons were presented without context and with an order randomized for each participant (who were among our friends and colleagues, and theirs). Forty-eight participants between 18 and 67 years of age (avg. 31.9) each voted for a category appropriate to each emoticon. To assess the level of agreement between our classification of the emoticons and that of the rest of the group, we began with an estimate of the group opinion on the three-way classification. For each emoticon, we assume the default group opinion of it as neutral/ambiguous; we count its classifications in each category, and if the number of positive or negative votes exceeds the number of neutral/ambiguous votes we take the group opinion accordingly.⁶ We

⁴ Consensus characterizes *kaomoji* as having an axis of orientation that is orthogonal to the majority of emoticons in other languages (e.g. (^) vs. :)).

⁵ Each item (or none at all) had to be classified: “neutral” includes “abstain”.

⁶ In case of a tie in votes between “positive” and “negative” votes both exceeding the number of votes for “neutral/ambiguous,” the default was set to somewhat arbitrarily set to “positive”, although such a situation did not actually transpire.

Table 2. Frequencies of the top 12 emoticons

Emoticon	Class		Count	Emoticon	Class		Count	Emoticon	Class		Count
	Us	MK			Us	MK			Us	MK	
!!!	-	?	34313	**	?	?	14108	;)	+	+	3933
:-)	+	+	20375	:)	+	+	11478	8)	?	+	3401
???	-	?	14843	***	?	?	8698	(x)	?	?	2832
;-)	+	+	14531	*(-	-	-	4611	:-))	+	+	2774

measured agreement between this construction of the group opinion and our own with Cohen’s Kappa statistic ($\kappa = 0.37$).⁷ However, a higher level of agreement would be desirable, since the emoticons are not evenly distributed in frequency of use. From this data we also constructed two alternative classifications.

The first alternative is provided by an individual who we dubbed “Max Kappa”: this is the individual whose Kappa statistic computed with respect to the remainder of the raters (including ourselves) had the highest value (0.88). We used that individual as a coherent source of judgements across the 100 emoticons and representative of the group. The Max Kappa ratings on the 12 most frequent emoticons is also provided in Table 2 under the columns headed, “MK”.

Finally, because of a separate intuition that emoticons are generally used in context to express a definite sentiment rather than ambiguous one, we constructed a third classification of the emoticons as either positive or negative based on the plebiscite, taking the neutral/ambiguous votes as abstentions on whether an emoticon was positive or negative. Emoticons where there was a tie between positive and negative votes (three) were eliminated, as were cases in which definite “quorum” (defined as at least one quarter of the votes not being regarded as abstentions). Thus, the plebiscite yielded a definite binary opinion for the remaining 79 emoticons. The top 12 of these in frequency are provided in Table 3. In can be seen that emoticons that were eliminated were not particularly high on iconicity, and were certainly low on polarized sentiment.

Table 2 shows that our own judgements differed from that of Max Kappa (the group consensus) on the most frequent and third most frequent emoticons, amounting to a disagreement between negative polarity and ambiguity that affects 28% of the tokens used. However, this is not to say that the real difference is 28% because we don’t have a statement of the relative number of times the actual context fully settles the intended use of !!! or ??? as positive or negative. The consensus-driven binary classification agrees with the initial scheme for the most frequent emoticons. It is important for later observations that the equivocation between the classification schemes is between “negative” and “neutral/ambiguous” and not between “positive” and “negative”.⁸

⁷ Kappa ranges between -1 (complete disagreement) and 1 (complete agreement).

⁸ For 8), the 8th most frequent on the binary scheme and 10th on the ternary scheme, we thought it neutral, while both Max Kappa and the electorate rated it positive.

Table 3. Frequencies of the top 12 emoticons with plebiscite determined polarity

Emoticon	Class	Frequency	Emoticon	Class	Frequency	Emoticon	Class	Frequency
!!!	-	34313	:)	+	11478	:~)	+	2774
:~)	+	20375	*~)	-	4611	=>	+	2537
???	-	14843	;)~)	+	3933	:~)	-	2439
;)~)	+	14531	8)	+	3401	*~)	+	2294

2.1 Distributions of Emoticons by Language, Overall

Most messages posted did not contain any emoticons, and that was true for each language, even with considerable variation in the number of messages posted. The leftmost columns of Table 4 indicate this. From these frequencies, one may deduce that in comparing the type of emoticon used across the languages, in science and politics, one is comparing distributions involving small numbers with distributions involving substantially larger counts.

The language with the greatest proportion of postings with emoticons was German, and the rightmost six columns in that table indicate that of the emoticons that were used, the German postings included overwhelmingly positive emoticons. This is regardless of the classification scheme based on our own judgements or the consensus judgement provided by Max Kappa. With respect to both three-way classification schemes, we looked at the distribution of positive, negative and neutral/ambiguous emoticons each language versus each other language overall, in politics and science. Each of these two row by three column comparisons was significant using χ^2 , except for German versus English political discussions on our classification (and that comparison provided the least value $\chi^2 = 15.1$, $p < 0.0005$ for Max Kappa's classification).

On our classification scheme, the most significant overall differences were between German and Italian ($\chi^2 = 9828.06$, $p < 0.0001$),⁹ German and English ($\chi^2 = 2613.07$), Italian and English ($\chi^2 = 2059.75$), and Italian and Swedish ($\chi^2 = 1056.87$). On the alternative scheme provided by Max Kappa, a great frequency of the emoticon tokens shifted from the negative category to "neutral". There, German is the most distinctive, the distribution of German and Italian emoticons still providing the greatest difference ($\chi^2 = 10283.69$), followed by German and English ($\chi^2 = 5548.15$) and then German and Swedish ($\chi^2 = 2961.57$). Relative to German on either classification scheme, Italian had the least overall positive emoticons.

Table 5 shows the distribution of emoticons overall across messages when there are two categories.¹⁰ As mentioned above, the procedure for adjudicating the categories on the basis of opinion derived from definite polarities vs. abstentions: 21 emoticons were eliminated. This entails that there are distinct counts of

⁹ All further χ^2 values reported also have $p < 0.0001$.

¹⁰ The values in the "Emoticon Distribution" columns of this table and the values in remaining tables are percentages.

Table 4. Postings With and Without Emoticons & Proportions of Emoticon Types

Language	Message Distribution			Emoticon Distribution					
	Emoticons	No Emoticons	% With	Authors			Max Kappa		
				% +	% -	% ?	% +	% -	% ?
Swedish	4353	14686	22.92	58.34	27.36	14.31	44.26	1.59	54.15
German	23671	52492	31.08	71.42	16.46	12.12	73.37	8.29	18.34
Italian	52179	153610	25.36	39.70	51.70	8.61	41.63	3.59	54.78
English	20099	75097	21.11	50.50	34.67	14.84	46.40	4.08	49.52

Table 5. Group Opinion: Postings With and Without Emoticons & Type Distribution

Language	Message Distribution			Emoticon Distribution	
	Emoticons	No Emoticons	% With	% Positive	% Negative
Swedish	3424	15615	17.98	62.56	37.44
German	25209	50954	33.10	82.86	17.14
Italian	54716	151073	26.59	45.50	54.50
English	18367	76829	19.29	61.23	38.77

messages containing emoticons than for the analyses described above using the three-way classification. Focusing on this binary sentiment classification of emoticons, the only overall language comparison that was not significant was between Swedish and English, and within analysis of messages in politics newsgroups, neither the German-Swedish nor the German-English comparisons were significant. The greatest overall difference was between German and Italian ($\chi^2 = 9842.00$), where German has the most positive emoticons and Italian has the greatest proportion of negative emoticons; German and English ($\chi^2 = 2566.23$), which splits in the same direction but with German positive emoticons representing a greater proportion than for English; Italian and English ($\chi^2 = 1360.34$), where the split is in the same direction as for German and Italian; German and Swedish ($\chi^2 = 788.06$), which is quite like the German and English split; and Italian and Swedish ($\chi^2 = 376.21$). On the binary classification scheme, there is an increased proportion of positive emoticons among the German messages over both the original and comparison category schemes, and a distribution of negative emoticons in Italian which is exactly like the first scheme. This relationship for Italian is to be expected since between our initial classification scheme and the second one which employed a plebiscite, negative judgements shifted to the category of “neutral/ambiguous”; however, when abstentions are disregarded in cases where quorum persists, the original positive and negative classifications are ratified. Clearly, equivocation between “negative” and “neutral/ambiguous” between the categorization schemes makes a difference in the overall distribution of negative emoticons, but in terms of positive emoticons versus non-positive emoticons, the classifications agree.

2.2 Distributions of Emoticons by Language and Topic

Initial Three-Way Classification. Table 6 shows the distribution of emoticons according to our classification scheme across topic area, by language.

Within newsgroups with a science focus, the greatest differences on our classification were between German and Italian ($\chi^2 = 715.17$), where the positive German emoticons concentrated and non-positive emoticons split closely between negative and ambiguous while Italian had fewer positive emoticons and more negative than ambiguous emoticons; Swedish and English ($\chi^2 = 547.96$), where while the English data had a pattern more like the German one, the Swedish data was closely split between positive and negative, and most emoticons ambiguous; German and Swedish ($\chi^2 = 544.37$), for which the German data concentration of positive emoticons is matched by the Swedish concentration of ambiguous ones; Italian and Swedish ($\chi^2 = 428.13$), which split along the lines of the German and Swedish comparison. On this scheme, the Germans, English and Italians have mainly positive emoticons for science discussions (Germans the most positive), and the Swedish messages have notably fewer positive emoticons.

Within newsgroups with a politics focus, the greatest differences on our classification were between Italian and English ($\chi^2 = 3200.29$), where the Italian messages have more than half negative emoticons, about half of that amount of positive emoticons and few ambiguous ones; Italian and Swedish ($\chi^2 = 1647.32$), where Italian has as great proportion of negative emoticons as Swedish messages have positive; and Swedish and English ($\chi^2 = 210.41$), where the English messages have more negative Emoticons than the Swedish ones. On this classification scheme, for the discussion of politics the use of emoticons designated as negative sharply separate the Swedish messages and the Italian ones.

Max Kappa's Categories. Table 7 shows the distributions of emoticons determined by the individual closest to the consensus opinion over the three categories, analyzed according to subject area and by language.

For science newsgroups the greatest difference is between German and Italian ($\chi^2 = 944.31$), where the pattern is in the same direction (mainly positive, some ambiguous, a small proportion of negative emoticons) but with nearly double the proportion of ambiguous emoticons in Italian than German and a greater

Table 6. Our Classification: Percentages of Emoticon Types by Language and Topic

Science				Politics			
Language	Positive	Negative	Ambiguous	Language	Positive	Negative	Ambiguous
Swedish	20.38	14.69	64.93	Swedish	60.09	27.94	11.97
German	71.60	16.29	12.11	German	45.73	40.70	13.57
Italian	59.39	26.57	14.04	Italian	33.81	59.20	6.99
English	65.48	25.27	9.25	English	48.16	36.13	15.70

Table 7. Max κ : Percentages of Emoticon Types by Language and Topic

Science				Politics			
Language	Positive	Negative	Ambiguous	Language	Positive	Negative	Ambiguous
Swedish	20.38	1.90	77.73	Swedish	45.36	1.58	53.06
German	73.55	8.30	18.15	German	48.74	7.54	43.72
Italian	61.90	6.48	31.61	Italian	35.58	2.73	61.70
English	67.89	8.66	23.46	English	43.05	3.37	53.58

proportion of positive emoticons for German than Italian; German and Swedish ($\chi^2 = 490.68$), where the proportion of ambiguous emoticons for Swedish is greater than the large proportion of positive emoticons for German; and Swedish and English ($\chi^2 = 294.91$).

In the politics newsgroups on this classification scheme, the greatest cross-linguistic difference is between Italian and English ($\chi^2 = 371.66$) and Italian and Swedish ($\chi^2 = 180.70$) where in both cases the proportion of positive emoticons in Italian is significantly lower than for the other language and the proportion of ambiguous Italian emoticons is rather higher. Again, this represents the shift of emoticons that had been classified as negative on our initial classification scheme to neutral/ambiguous on the group scheme. The low proportion of positive emoticons in messages posted to politics remains about the same for both classification schemes.

Binary Sentiment Classification. We are not ultimately satisfied with the three way classification involving the category “neutral or ambiguous” described above, partly because of its catch-all status. We have pointed out that some of the emoticons eliminated in this classification were very low on iconicity and also low on polarization of sentiment according to the plebiscite. It is true that emoticons, like any other content bearing expression, can be used ambiguously or ironically, and thus we expect that allowing the abstentions category in the plebiscite filters out those votes about emoticons where context is relatively free to determine the sentiment accented with an emoticon.

Next we describe the comparison outcome by language and subject of distributions within binary category classifications of emoticons. Within science discussions, German provided the most messages and was most distinctive. The German and Italian distributions across the two categories showed the greatest difference ($\chi^2 = 625.25$), where 83% of German emoticons were positive to 72% in Italian messages. For German and English ($\chi^2 = 127.3$) the comparison was to 74% positive. With the German and Swedish ($\chi^2 = 30.74$) comparison involving Swedish’s relatively smaller 58% positive emoticons, it has to be recalled that there were only 814 messages in total posted to the Swedish science newsgroups. For politics newsgroups the greatest differences were shown by Italian and English ($\chi^2 = 1973.02$) and by Italian and Swedish ($\chi^2 = 771.74$) where only Italian had many more negative than positive.

Table 8. Binary Consensus: Percentages of Emoticon Types by Language and Topic

Science			Politics		
Language	Positive	Negative	Language	Positive	Negative
Swedish	58.11	41.89	Swedish	62.66	37.34
German	83.05	16.95	German	56.67	43.33
Italian	71.77	28.23	Italian	38.23	61.77
English	74.43	25.57	English	58.85	41.15

2.3 Observations

This section has indicated how emoticons were distributed by sentiment categorization scheme by language overall and as a function of topic. For Swedish, Italian and English, the distribution of types of emoticons used within discussions of politics closely resembles the overall distribution for the language, while for German emoticon use in science discussions corresponds to the overall use.¹¹ Emoticons in the Swedish discussions of politics were nearly half positive, while for Italian they were more than half negative (depending on the classification scheme) and the remainder split over the other categories. For English and German, a nearly equal distribution across the three types occurred. In discussion of science, emoticon used in Swedish were mainly ambiguous, with an equal distribution of positive and negative, while the other languages used mainly positive emoticons; however, the least number of postings was for science groups in the Swedish news groups. The dependence of the inter-cultural differences upon the classification scheme has also been discussed: the votes supplied by the raters were provided in the absence of influencing context. On one hand, context could resolve an ambiguous emoticon in either direction; on the other hand, the resolving information that was supplied through quorate non-abstentions provides an indication of how context would resolve ambiguous emoticons on average. In this evaluation of the initial classification scheme we have considered two alternatives—the classification provided by a single individual in the pool of raters who most people agreed with and the classification constructed by forcing binary classification. Other alternatives remain to be explored as well, but with the acceptable level of agreement realized between the first scheme and the judgements of the larger pool of raters some consensus among the schemes emerged. German emoticons are overwhelmingly positive with much discussion of science, and little of politics. Swedish emoticons were concentrated in politics rather more than science, and were mainly positive there. Italian messages in politics had a smaller proportion of positive emoticons than the other languages, but a comparable proportion of positive emoticons in science newsgroups. We did

¹¹ This can be explained further from Table 1; the postings for German were concentrated in science newsgroups, while for the other languages, there are more postings in the politics newsgroups.

not expect the cross-cultural analysis to work out the way that it did, expecting in some way that the Swedish discussions might actually have more emoticons proportionately than the other languages for reasons connected to those outlined in the introduction. We sought to determine if the distributions remained constant over time, thinking that politics would be more volatile than science discussion, and that it might be able to track emoticon use peaks and troughs with respect to outstanding events of the time [2].

In §3, we examine the distributions within the three categories using just the initial classification over time. We leave it to further work to analyze the divergences among the trends that would occur with the alternative classifications.

3 Chronological Analysis

The results in §2 are based on the total accumulation of postings. We noted that there was an uneven distribution of postings in each category. Particularly because one of the topic areas is politics, a source of volatile discourse sentiment, it is useful to study the distributions of emoticons over time, in case emoticon use in a particular language and topic is dominated by postings restricted to a short period. Figure 2 shows how the messages available to us over the 66 week period were distributed in each language relative to the total number of postings. The overall figures are represented in the graph on the left, politics in the middle, and science on the right. Italian and English consistently dominate the flow of postings in politics newsgroups, while German and Italian dominate science newsgroups.

In the next graphs, the lines represent the use of positive, negative and neutral emoticons, by week. The values plotted are the number of emoticons relativized to the total number of tokens over time.

Figure 3 shows on the left that emoticons in Swedish political discourse for the first 50 weeks were mostly positive, and thereafter, mostly negative. Shares in Ericsson fell by 25% on October 16, 2007 — this is exactly the week of the

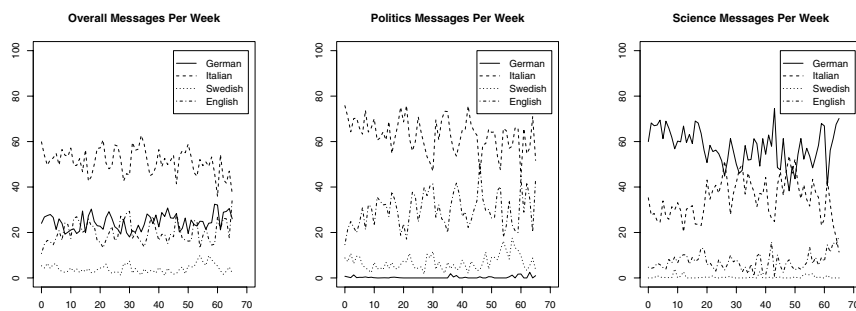


Fig. 2. Relative Message Counts per Week by Language: Overall, Politics and Science

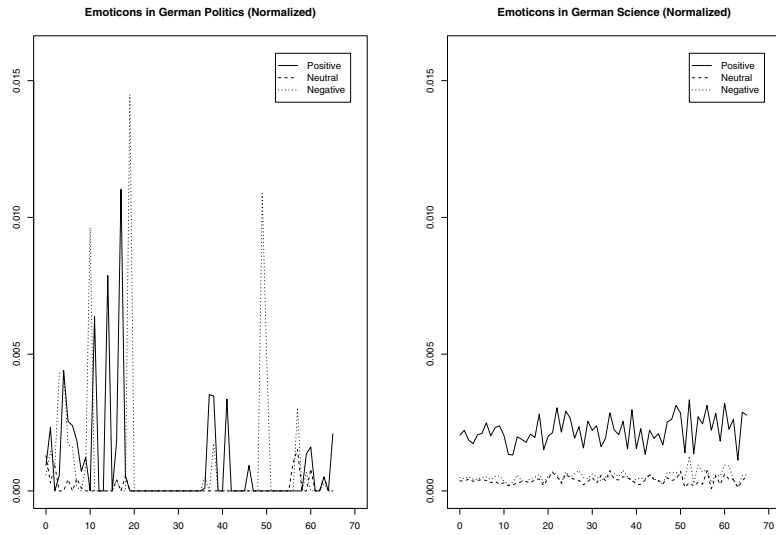


Fig. 4. German Emoticons: Politics and Science

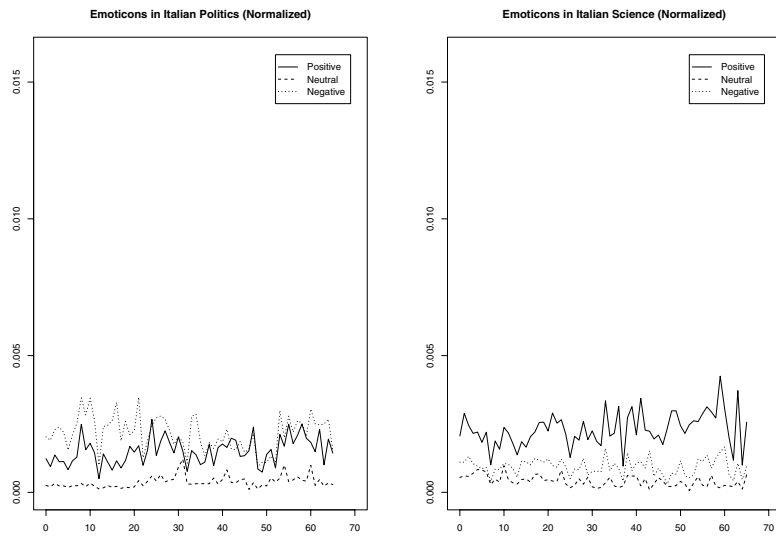


Fig. 5. Italian Emoticons: Politics and Science

August 13-20, in which six young Italians were executed in Duisburg (Germany), evidently in a mafia feud, an event that precipitated an Italian police crackdown on the mafia. The 53rd through the 55th weeks covered the first half of October 2007, and this included in the European Media Monitor summary of dominant news items an announcement of a pending sale of government shares in Alitalia (October 9), “overwhelming” worker approval of pension reform raising

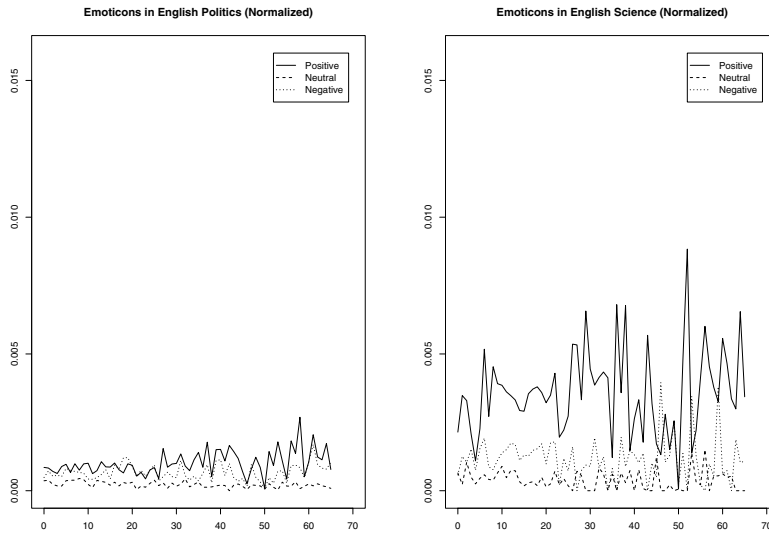


Fig. 6. English Emoticons: Politics and Science

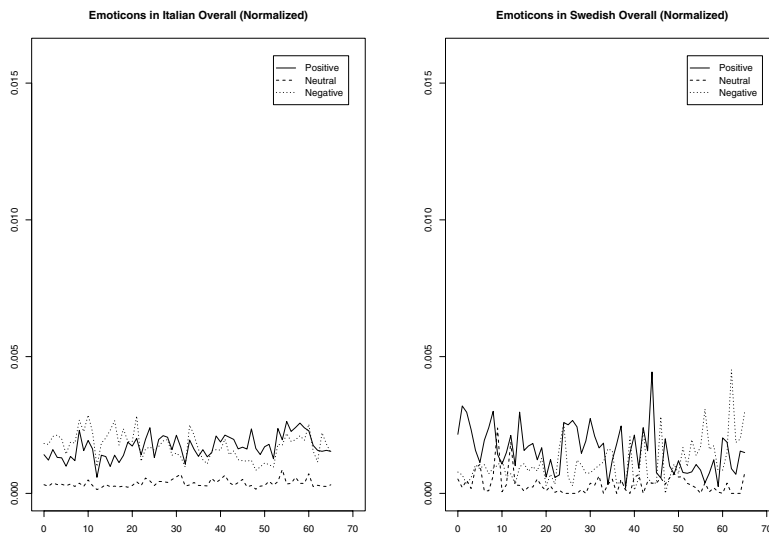


Fig. 7. Overall Emoticons: Italian and Swedish

retirement to age 60 (October 10), an announcement of the state ship building company winning the contract to build the new Queen Elizabeth (October 11), and a landslide victory for Walter Veltroni, Mayor of Rome, in a national election primary (October 14).¹⁵ The graph on the right shows that for discussion in science newsgroups, positive emoticons dominated throughout the period.

¹⁵ <http://press.jrc.it/NewsExplorer/> – last verified June 2008.

Emoticon use in the *.uk newsgroups is shown in Fig. 6. Use of emoticons in politics newsgroups favored positive ones over the entire period except the week which included January 30, the same week that a controversial decision about awarding a super-casino license in Manchester rather than London or Blackpool was announced and Lord Levy, fundraiser for Tony Blair, was arrested, and Blair himself was questioned by police. Emoticons in the science newsgroups are also positive for the period, with the exception of August 19-25.

The aggregate of both subject areas over the 66 weeks (Fig. 7) shows that, coincidentally, the periods in which positive and negative emoticons dominate are in a roughly complementary distribution between Italian and Swedish. More negative than positive emoticons in Italian appear for the first half of the period, and then mainly the reverse. For Swedish, the first two-thirds are positive; the final third are mostly negative.

4 Discussion

A past study demonstrated that aggregate results differentiate Swedish and Italian emoticon use, with more positive emoticons in Swedish politics newsgroups and more negative emoticons in the same context in Italian. It is clear that certain emoticons that we have considered are in the absence of context deemed to be ambiguous by most of the participants in our rating study. The two most important of these emoticons are !!! and ??? because of their rank frequency of first and third in the top ten, respectively. On none of the classification schemes that we considered are they positive. On one they are ambiguous and thus require context to disambiguate (if they can be disambiguated). It may well be that the use of these emoticons accounts for the evident preponderance of negative emoticons in messages posted to politics newsgroups. However, it remains the case that messages posted to Italian politics newsgroups had significantly fewer clearly positive emoticons than the other languages considered.

We observed the way the categorization distributions extend over time from late September 2006 to January 2008. Divergences from those trends were noted and related to contemporaneous external events with presumed impact on public sentiment, regardless of whether they were explicitly mentioned in postings.

We have reported the use of emoticons in four languages and two broad topic areas over a 66 week period. We provide a methodological starting point for interpretive cross-cultural analyses of emoticon use. Further quantitative analysis of emoticon use in terms of levels of interactivity in such discussion groups as sampled here is necessary, as is correlation of emoticon types with accompanying sentiment bearing words. The present study attempts no such content analysis, preferring instead to identify the raw patterns of emoticon use. There is a strong argument to consider use of nearly all but the most clearly negative emoticons (e.g. “!?!?!?”) as actually conveying positive emotions—if a writer has bothered to use an emoticon, then this is a signal of positive affect. Certainly, negative emoticons (e.g. “:-<”) can be used to indicate a sympathetic response to an adverse situation, and equally, a positive emoticon might be used to temper the

content of otherwise negative companion text. These double dissociations may confound any correlations between emoticons and words or phrases. However, this potential is exactly what pragmatic analysis of emoticon use may reveal.

The next step is to conduct a follow-on study like the one we used to assess agreement with our own classification of the emoticons. We have analyzed *inter-cultural differences in the use of emoticons* with the idealizing assumption (modulo individual differences in agreement such as have been quantified here) of *universal agreement on classification* of the emoticons as positive, negative or ambiguous. The results show a dearth of positive emoticons in the Italian discussions of politics. Three possible explanations are available: this distribution might reflect a difference in underlying sentiment in Italian political discourse during the sampled period; the effect might be amplified (or eliminated) with disambiguating contextual analysis; or it might be explained by Italian differences in the classification of certain emoticons as positive that others find negative or ambiguous, such that across the four languages, the underlying sentiment was the same for each during the period, but that there are intercultural differences on the sentiment conveyed by individual emoticons. Future work should evaluate these candidate explanations.

References

1. Allwood, J.: Are there Swedish patterns of communication? In: Tamura, H. (ed.) *Cultural Acceptance of CSCW in Japan & Nordic Countries*, pp. 90–120. Kyoto Institute of Technology (1999)
2. Balog, K., Mishne, G., de Rijke, M.: Why are they excited? Identifying and explaining spikes in blog mood levels. In: *11th Meeting of the European Chapter of the Association for Computational Linguistics* (2006)
3. Cavnar, W.B., Trenkle, J.M.: N-gram-based text categorization. In: *Proceedings of Third Annual Symposium on Document Analysis and Information Retrieval*. UNLV Publications/Reprographics, Las Vegas, NV, pp. 161–175 (1994)
4. Cerrato, L.: A comparative study of verbal feedback in Italian and Swedish map-task dialogues. In: *Proceedings of the Nordic Symposium on the Comparison of Spoken Languages*, pp. 99–126 (2003)
5. Ibarra, F.C., Galimberti, C.: Stereotypes and gender identity in Italian and Chilean chat line rooms. *PsychNology Journal* 4(1), 25–52 (2006)
6. Janssen, J., Vogel, C.: Politics makes the Swedish:-) and the Italians:-(. In: Ahmad, K. (ed.) *Sentiment Analysis: Emotion, Metaphor, Ontology & Terminology*, pp. 53–61 (2008) (EMOT 2008 Workshop at LREC 2008: Poster)
7. Katsuno, H., Yano, C.: Kaomoji and expressivity in a Japanese housewives' chatroom. In: Danet, B., Herring, S.C. (eds.) *The Multilingual Internet: Language, Culture, and Communication Online*, pp. 278–299. Oxford University Press, Oxford (2007)
8. Neviarouskaya, A., Prendinger, H., Ishizuka, M.: Narrowing the social gap among people involved in global dialog: Automatic emotion detection in blog posts. In: *Proceedings International Conference on Weblogs and Social Media*, pp. 293–294. Omnipress, Boulder (2007) (ICWSM 2007: Poster)
9. Nicolov, N., Salvetti, F., Ivanova, S.: Sentiment analysis: Does coreference matter? In: *Symposium on Affective Language in Human and Machine*. AISB, Aberdeen (2008)

10. Read, J.: Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In: Proceedings of the ACL Student Research Workshop, pp. 43–48. Association for Computational Linguistics (2005)
11. Schler, J., Koppel, M., Argamon, S., Pennebaker, J.: Effects of age and gender on blogging. In: Nicolov, N., Salvetti, F., Liberman, M., Martin, J.H. (eds.) AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs, pp. 199–205. AAAI Press, Menlo Park (2006)
12. Suzuki, Y., Takamura, H., Okumura, M.: Application of semi-supervised learning to evaluative expression classification. In: Gelbukh, A. (ed.) CICLing 2006. LNCS, vol. 3878, pp. 502–513. Springer, Heidelberg (2006)
13. Tseliga, T.: It's all Greeklsh to me! Linguistic and sociocultural perspectives on Roman-alphabeted Greek in asynchronous computer-mediated communication. In: Danet, B., Herring, S.C. (eds.) The Multilingual Internet: Language, Culture, and Communication Online, pp. 116–141. Oxford University Press, Oxford (2007)