

# Epistemic Signals and Emoticons Affect Kudos

Carl Vogel<sup>†</sup> and Liliana Mamani Sanchez<sup>†‡</sup>

Trinity College Dublin, Dublin 2, Ireland

email: {vogel, mamanis1}@tcd.ie

<sup>†</sup> Integrated Language Technologies: Text Analytics, Centre for Next Generation Localisation, www.cngli.ie

<sup>‡</sup> Centre for Computing and Language Studies, Computational Linguistics Group

School of Computer Science and Statistics

Trinity College, Dublin

**Abstract**—Our focus is on the interaction between emoticon use and epistemic hedges in the perception of individual contributions to discourse (and posters of those contributions) as deserving of kudos for their input. The communities with English as a *lingua franca* that we explore consist of self-motivated contributors to user-fora supported by a major multinational with a software technology company. User categories are determined by a few orthogonal classifications: employees, novice users, and experts; recipients of kudos vs. non-recipients of kudos; etc. We explore the interaction between social signals and signals of certainty in content. Among the effects reported are the negative influence of epistemic hedges used in posting on propensity for others in the community to accord kudos to such postings, but a positive influence of the same in interaction with the use of emoticons.

## I. INTRODUCTION

Community fora are increasingly used by companies as consumer-communication channels. The data we analyze here is provided by a major multinational software company in which participants discuss technical aspects of products and services by posting messages to a web forum. We think that the result reported will generalize to comparable technical fora where the *raison d'être* is the seeking and sharing of technical information.<sup>1</sup> This data has been analyzed in related work; [1] shows that posters given kudos<sup>2</sup> for their contributions tend to use positive emoticons as social signals, for example. We address these questions: (i) Does emoticon use (in the general sense of “emoticon”) predict likelihood of a message being given kudos? (ii) Are forum lay-leaders more or less likely to use emoticons than employees or novices? (iii) Do the answer to the above questions depend on the polarity of the emoticons, and to what extent? (iv) How are these questions answered with “epistemic signal” in place of “emoticon”? (v) What is the interaction between emoticon and epistemic signal use in predicting whether a message is likely to be accorded kudos?

We explore the usage of emoticons as social signals across user categories. The set of emoticons is described in §III-A, where more detail about the dataset is also given. We wanted

<sup>1</sup>The forum data has been provided to us by the company involved as a research collaborator, but remains unnamed in this discussion. The research is conducted with clearance from the Research Ethics Committee of the School of Computer Science and Statistics at Trinity College Dublin. We do not disclose the userids of any participant. While we are not able to share the corpus, the data is equivalent to what could be obtained by using a web robot to scrape data from the publically visible content.

<sup>2</sup>The act of rating a message as useful is known as giving it kudos.

to know the extent to which emoticon usage is diagnostic of membership in this user classification. Results are described in §IV. We conclude with discussion of our main findings and the directions these suggest for future work in §V.

## II. BACKGROUND

### A. Emoticons

The distinction between emoticons and smilies has emerged over time with the potential for graphical user interfaces to depict them as composed from typographical characters (e.g. “:-)”) or pictorially (e.g. 😊). The former have become known as “emoticons”, and the comparable social signals that take advantage of more advanced graphics and even animations are sometimes designated “smilies”, even though both terms have been in use since the onset of text-based computer mediated social signals. In this work, where this distinction between these two sorts of computer mediated social signal is relevant, we will adopt this usage distinction. Given that we are exploring the expression of perception of merit in postings to technical fora, the distinction can be relevant, since someone using smilies, in the special sense, can be inferred to be making use of aspects of the computer interface that enables the forum. Use of these may be taken to signal at least a minimal level of competence with the interface provided. It may also be taken as a signal of includedness to use the “language” offered by the forum providers. Similarly, the use of text-based emoticons only may be taken as a signal of comfort with computer mediated communication outside the support of friendly interfaces, and thereby, greater levels of technical competence. The open ended nature, through compositionality and novel invention in the text-based emoticons, in contrast to smilies provided by an interface, may also lend to an impression of competence among emoticon users. Such possible interpretations of the different forms of social signals particular to computer mediated communication partly motivate our own exploration of the data using this distinction. However, where the distinction does not matter, we use “emoticons” as a general term for both sorts of social signal.

The perceptual nature of emoticons has been studied and fMRI evidence produced to show that the brain responds to emoticons as dynamic percepts, even when they are static [2]. Functional differences in brain activity related to perception of emoticons strengthens the case that their presence or may have

consequences for perception of other aspects of the messages in which they occur. Through their status as social signals, emoticons have been positively correlated in their perceived sentiment encodings with stereotypical responses to climate patterns [3], [4]. Emoticons have been exploited in sentiment analysis systems as clues for determining sentiment scores [5], [6], and for achieving domain-independent sentiment classifiers [7]. Cross-cultural analyses of emoticon use have been conducted, noting predictable topical relations in discussion of politics, but surprising differences in sentiment of emoticon use in discussion of scientific topics [8], [9].

In treating emoticons as signals of expertise, their informal nature cannot be overlooked. In some contexts, they are considered as potential impediments to vocabulary learning, but not observed as such [10]. In fact, their informal nature is reported as making communications more friendly and personal [11], and this could be a feature in the inclination to rate a posting or poster highly for technical competence, even if it does not have a direct impact on the perception of competence. As the use of emoticons is an aspect of online presentation of self, it is relevant to consider their role in the perception of expertise [12], [13]. In comparing emoticon use in instant messaging and email, [14] found no effect on perception of information content in relation to emoticon use, but did find that users who used more emoticons tended to think instant messaging informationally richer.

We study emoticon usage across three groups of technical forum users communicating in English in relation to explicit signals of epistemic certainty and the joint impact of epistemic signals and emoticons on the likelihood of individual postings including them being accorded kudos (and consequently, the likelihood of the poster receiving kudos).

### B. Epistemics

A linguistic signal of epistemic certainty is an expression such as “I think”, “possibly”, “it seems” or “obviously”, etc. We use a lexicon of such signals due to [15]. Such signals may be deemed meta-linguistic in the sense that they are linguistic forms which comment on the intended interpretation of other linguistic forms that fall within their scope. There are at least two approaches to assessing the semantic contribution of epistemic signals. On one approach, for any sentence  $\sigma$ , and any epistemic marking of the sentence  $[\cdot]$ , the marker supplies an ordering source to the epistemic modal base (using an explication of modal semantics such as provided by [16]):

- (1)  $[[Ithink](\sigma)]^w = 1$  iff  $[[\sigma]]^{w'} = 1$  in some of those worlds at least as similar to  $w$ , the actual world, in terms of my grasp of  $w$  ( $\preceq$ , henceforth).
- (2)  $[[Obviously](\sigma)]^w = 1$  iff  $[[\sigma]]^w = 1$  in all of those worlds  $w' \preceq w$ .

In this approach, the force of the epistemic signal selected corresponds to a kind of quantification over the sorts of worlds in which the scoped content is true, relative to the world in which the overall expression is uttered or evaluated. Thus, “I think” carries a different kind of quantification (1) than “obviously” (2), and so on, for each of the epistemic signals.

An alternative view that we entertain is driven by the manifest unreliability of certain meta-linguistic expressions. For example, it appears to us to be an empirical fact that most uses of the word “literally” indicate not the literal content that the expression in its scope may be taken as literally true, but the opposite, that the expression in its scope must be understood figuratively [17], [18]. On such a view, the unmarked case (3) conveys the greatest certainty, and any form of explicit epistemic signal of certainty has the function of raising doubt about the content in its scope, or at least in yielding a weakened commitment to the truth of the content.

- (3)  $[[\sigma]]^w = 1$  iff  $[[\sigma]]^{w'} = 1$  in all  $w' \preceq w$ .
  - (4)  $[[Ithink](\sigma)]^w = 1$  iff  $[[\sigma]]^{w'} = 1$  in some  $w', w' \preceq w$ .
  - (5)  $[[Obviously](\sigma)]^w = 1$  iff  $[[\sigma]]^{w'} = 1$  in some  $w' \preceq w$ .
- On the latter view, in its strongest form, all explicit epistemic signals synonymously indicate that the content in their scope is not certain. Functional explanations for the use of such signals (e.g. circumspect politeness) may be adduced to explain the presence of the “weaker” signals, like “I think”. Expressions that appear “stronger”, like “obviously”, may also be addressed with functional explanations; however, it is worth considering how few uses of “obviously” embed content that actually is objectively obvious (rather, in analogy with “literally” as a reliable signal of figurative meaning, “obviously” is a good signal of opacity). Support for the synonymy kind of account exists in the fact that even in communicating with generalized quantifiers over quantities, people do not reliably rank natural language determiners (e.g. “few”, “a few”, “not so many”, etc. [19]) in discriminating meaning, and reliable discriminations of meaning between comparable epistemic expressions (e.g. “possibly”, “I’m sure”, etc) can be expected to be as rare.

A technical user forum provides an apt dataset in which to test which of these approaches to epistemic signals is more frequently consistent with usage. This is because a primary premiss of such a forum is the exchange of technical content. Of course, this is not the sole premiss, since another factor in the provision of user fora is the fostering of user communities, and potentially, product evangelists emerging from their midsts; thus, the social element of communication in the forum is also important. Therefore, it can be examined whether forum users are more likely to rate as useful those messages which they take to be informative, without linguistic hedges of the certainty and with socially engagement, or notes that hedge certainty or lack social signals.

### C. Empirical context

Communication in the forum we study is via English, with varying levels of fluency and technical expertise. The forum supports cross-classification of users in terms of professional status (i.e. “employee” vs. “guru” vs. “kudoer”). Individuals using the system have the capacity to give kudos to individual postings, and a record of the number of kudos accorded to each message is available. A kudoer is someone whose messages have received kudos, independent of employment status.

Forum users may be expected to be sensitive to how they are perceived, and given the technical nature of the forum, are

likely to seek to be considered as expert as the employees who contribute. This predicts emulation of employees by the most expert and professional forum “netizens”, and correspondingly, a convergence of linguistic and nonlinguistic features among postings of employees as participation unfolds (cf [20]).

We distinguish three groups of forum contributors in consumer-contributor technical forum scenarios. Firstly, common users (consumers) approach the forum in search of solutions to technical issues arising from the use of software products. In this category are “lurkers”, who by definition, do not make textual contributions to the forum (but may provide kudos to postings), and who therefore are omitted from analysis in what follows. Secondly, gurus are contributors given this “merit badge” partly because of assessments provided through the award of kudos for postings, but also partly through employee assessment of quality of answers and overall engagement. Gurus are not paid for their contributions. Their reward is intangible prestige. Theories from social psychology attempt to account for this sort of dynamic [21]. Thirdly, employees who are current or past workers are also forum members, and *qua* employees, may be paid to contribute.

### III. METHOD

We split the common (non-employee) users group between ranked and unranked users (aka, “not-ranked”). Rank is assigned algorithmically to common users according to levels of contribution. On the other hand, gurus roles are assigned more subjectively, with account of quantitative and qualitative parameters. Here we cross-classify users according to roles: {guru, employee, ranked, not-ranked} users. We wish to know if properties of postings by users give reliable indications of either algorithmically determined or subjectively determined *poster* roles. Additionally, we attend to perceptions of common forum members in rating posts. Therefore, separately, we also classify *posts* – those that received kudos and those that did not receive kudos: {kudoer post, non-kudoer post}. Messages are handled with shallow processing as bags of phrases.

#### A. Data & treatments

The data comprises: message content in XML and HTML format (subject and body) and metadata about posts and users. The message metadata comprises: posting date, user id, thread id, post id, last edition time, last edition author, kudos received, and viewings. Metadata about users comprises: roles and date of promotion to the guru role (where promotion has occurred). Message bodies contain the authors’ text but can also include two sorts of elements not composed by the author first hand: quotes referring to previous posts that are embedded in the text, either in part or entirely, and an edition field.

A set of 98 emoticons selected from previous research [8], [9] is used with an additional set of 45 smilies (e.g. 😊 and 🤔) that are supported by the forum management system. Smilies cover the same sorts of expressions of affect as the emoticons (e.g. smiles versus frowns), but also purport to encode additional distinctions (e.g. gender: 🤔 vs. 🤔). This is the same dataset used by [1], with identical treatment.

From the raw dataset (308,274 posts), we excluded posts and threads where only one or two groups of users could participate, as well as posts from gurus before they were assigned this role. This leaves 208,284 posts. Having isolated texts composed by individual users, we use regular expressions to count emoticon type and token frequencies. Matching true positives is difficult, as text in this genre often contains items that are easily confused with emoticons such shown in Fig. 1. We conservatively match only emoticons well delimited by spaces, tabs or newlines characters.

|  |
|--|
| FAT partition of (C:) and the old (F:) is now (E:)...<br>The storage Destinatioon was a Slave Drive (F:) ...   |
| ... 7) Selected Recover My Computer. 8) In the drop-list-box above the list of recovery points, I selected ... |
| Can add URL or use mask (e.g. with ? or *)   |

Fig. 1. Cases of false matches for the emoticons :, 8) and \*)

#### B. Forum data profile

The scale of the dataset is indicated in Table I, which shows the number of posts, users and average ratio of posts per individual (APPI) across groups. The APPI is a proxy measure of interactivity in each group: gurus show the highest level of interactivity, ranked users show the second highest, employees are next, and not-ranked users show the least posting traffic (c. 1 per user). In a second categorisation, the authors of each kind of post are counted: from 2,302 kudoer users, there were 2007 responsible for some nonkudoer posts and 295 have only kudoer posts. APPI values for kudoer and nonkudoer posts are not significantly different, because of the 2007 kudoer users who also contributed posts that received no kudos. Other measures of interactivity and correlations between emoticon use and interactivity are explored elsewhere [1].

TABLE I  
POSTS & WORDS PER GROUP FOR THE TWO CATEGORISATIONS.

|           | Role      | numposts | numusers | APPI     | Tokens |
|-----------|-----------|----------|----------|----------|--------|
| By role   | employee  | 25,490   | 400      | 63.725   | 58.63  |
|           | guru      | 27,489   | 15       | 1,832.60 | 70.70  |
|           | notranked | 50,456   | 19,462   | 2.593    | 89.23  |
|           | ranked    | 104,849  | 2,273    | 46.128   | 74.03  |
| Post Cat. | kudoer    | 18,540   | 2,302    | 8.054    | 99.38  |
|           | nonkudoer | 189,744  | 21,855   | 8.682    | 73.04  |

### IV. RESULTS

Table II shows the 10 most frequently used emoticons and smilies in each of the three categories of affect considered. Table III shows a sampling of epistemic markers.

#### A. Usage of emoticons across groups

We note the frequency of emoticon use in relation to the number of messages by each group: the number of posts with and without emoticons (by majority decision on polarity) and

TABLE II  
MOST FREQUENTLY USED EMOTICONS AND SMILIES.

| Rank | Positive |       | Negative |       | Neutral |     |
|------|----------|-------|----------|-------|---------|-----|
|      | Symbol   | N     | Symbol   | N     | Symbol  | N   |
| 1    | 😊        | 3,620 | !!!      | 3,272 | 😬       | 851 |
| 2    | :)       | 2,692 | ???      | 2,410 | ()      | 30  |
| 3    | 😄        | 2336  | 😞        | 1057  | \$\$\$  | 21  |
| 4    | 😂        | 1462  | : (      | 597   | (=      | 18  |
| 5    | : -)     | 1,157 | 😏        | 463   | \$\$    | 13  |
| 6    | ;)       | 779   | 😄        | 428   | 😬       | 11  |
| 7    | :D       | 588   | !?!?     | 252   | (D)     | 11  |
| 8    | 😄        | 421   | : - (    | 123   | 😬       | 10  |
| 9    | :P       | 223   | : /      | 48    | 😬       | 10  |
| 10   | ; -)     | 208   | 😬        | 40    | <=      | 5   |

TABLE III  
PROPORTION OF USAGE AND RANKS OF EPISTEMIC MARKERS ACROSS CATEGORIES.

| Uncertainty  | Frequency | Rank | Certainty    | Frequency | Rank |
|--------------|-----------|------|--------------|-----------|------|
| can          | 101,671   | 1    | I know       | 6,745     | 14   |
| would        | 47,678    | 2    | I understand | 2,000     | 36   |
| some         | 31,350    | 3    | I noticed    | 1,808     | 39   |
| could        | 26,684    | 4    | I'm sure     | 1,649     | 40   |
| try          | 25,667    | 5    | I saw        | 1,202     | 46   |
| something    | 15,253    | 6    | I've seen    | 868       | 54   |
| may          | 14,417    | 7    | I remember   | 694       | 59   |
| I think      | 11,908    | 8    | I am sure    | 622       | 63   |
| seems        | 10,819    | 9    | I heard      | 156       | 101  |
| think        | 7,809     | 13   | I've heard   | 143       | 104  |
| I don't know | 3,668     | 23   | I showed     | 32        | 135  |

the percentage of use are shown in Table IV. Emoticon use is low in our dataset; ranked users have the highest use (10%).

Posts from gurus, not-ranked and non-kudoer-posts authors include more emoticons than smilies when compared to the ones authored by employees and kudoers. More than double the number of employees' posts containing emoticons contain smilies, kudoers use almost the same amount of emoticons as smilies. The ratio of usage of emoticons to smilies by gurus and non-kudoer-posts' authors is 1.12 and 1.25 respectively. In not-ranked and non-kudoers' posts this ratio raises to 2.6.

As non-employees, gurus, ranked and not-ranked users can be less formal than users belonging to the company; employees use the forum to make formal announcements where the use of emoticons would be less obviously appropriate than in more engaged posts. This observation is supported by the fact that employees use less emoticons in their posts when compared to the users from the other two role-based groups. Reading Table IV vertically shows that employees and gurus use emoticons less than other user categories and that posts which have kudos are slightly more likely to have used emoticons than those that do not. We show the average frequency and standard deviation of emoticons per post and of tokens per post containing at least one emoticon, for the two user categorizations in Table V.

The ratio of frequency of emoticon type to overall frequencies per group is shown in Table VI. Positive signal usage

TABLE IV  
NUMBER OF POSTS PER POLARITY, WITH AND WITHOUT EMOTICONS AND PERCENTAGE OF POSTS WITH THESE SIGNALS, PER CATEGORY.

| By role   | Category  | pos   | neg   | neut  | > 0    | = 0     | % > 0 |
|-----------|-----------|-------|-------|-------|--------|---------|-------|
|           | employee  |       | 721   | 56    | 33     | 810     | 24680 |
| guru      | notranked | 1,833 | 225   | 118   | 2,176  | 25,313  | 7.9   |
|           | ranked    | 1,729 | 2,383 | 198   | 4,310  | 46,146  | 8.5   |
|           | ranked    | 6,956 | 3,491 | 887   | 11,334 | 93,515  | 10.8  |
| Post Cat. | kudoer    | 1,306 | 340   | 132   | 1,778  | 16,762  | 9.6   |
|           | nonkudoer | 9,933 | 5,815 | 1,104 | 16,852 | 172,892 | 8.9   |

TABLE V  
AVERAGE EMOTICON AND TOKEN FREQUENCY PER POST CONTAINING AT LEAST ONE EMOTICON, ACROSS THE TWO CATEGORISATIONS.

| By role   | Category  | Emoticons = E |             | Tokens = T |             |
|-----------|-----------|---------------|-------------|------------|-------------|
|           |           | $\mu(E)$      | $\sigma(E)$ | $\mu(T)$   | $\sigma(T)$ |
| Post Cat. | employee  | 1.112         | 0.414       | 78.048     | 100.481     |
|           | guru      | 1.402         | 1           | 76.622     | 91.991      |
|           | notranked | 1.235         | 0.673       | 115.004    | 143.601     |
|           | ranked    | 1.321         | 1.446       | 92.002     | 128.743     |
| Post Cat. | kudoer    | 1.31          | 0.98        | 120.83     | 145.34      |
|           | nonkudoer | 1.30          | 1.25        | 92.19      | 125.93      |

exceeds 59% across employees, gurus, ranked, and kudoer and non-kudoer posts, being particularly high in the employees and gurus. More than half of the emoticons of not-ranked users were negative. Employees and gurus used a significant amount of positive (> 80%) compared to negative emoticons ( $p < .05$ ).<sup>3</sup> Comparing those groups with ranked and not-ranked users, their use of positive emoticons is not significantly different ( $p = .127$ ), but the difference is significant for negative emoticons ( $p = .0252$ ). Kudoer posts show more positive emoticon use than non-kudoer posts ( $p < .05$ ).

Table VII shows emoticons use relative to posts per group. Use of positive signals is highest among gurus, followed by ranked users. Employees infrequently use emoticons (cf. Table IV). Not-ranked users have the biggest ratio of negative emoticons compared to positive emoticons (c. 1.5), approximately the inverse of the ratio for ranked users. Neutral emoticons usage is marginal when relativized to number of postings. Combined counts also show greater use of positive emoticons than of negative emoticons by all groups but not-ranked, where negative signals are mostly used. Negative emoticon use by ranked users is higher than among employees and gurus.

<sup>3</sup>The test in this paragraph refer to binomial proportion differences.

TABLE VI  
RATIO OF EMOTICON TYPE TO TOTAL OF SIGNALS PER GROUP.

| By role   | Category  | pos   | neg   | neut  |
|-----------|-----------|-------|-------|-------|
|           | employee  |       | 0.882 | 0.078 |
| Post Cat. | guru      | 0.862 | 0.104 | 0.034 |
|           | notranked | 0.394 | 0.586 | 0.020 |
|           | ranked    | 0.596 | 0.354 | 0.049 |
| Post Cat. | kudoer    | 0.719 | 0.235 | 0.046 |
|           | nonkudoer | 0.583 | 0.377 | 0.040 |

TABLE VII  
RATIO OF EMOTICON TYPE TO TOTAL OF POSTS PER GROUP.

|         | Category  | pos    | neg   | neut  |
|---------|-----------|--------|-------|-------|
| By role | employee  | 0.031  | 0.003 | 0.001 |
|         | guru      | 0.096  | 0.011 | 0.004 |
|         | notranked | 0.042  | 0.062 | 0.002 |
|         | ranked    | 0.085  | 0.051 | 0.007 |
|         | Post Cat. | kudoer | 0.090 | 0.029 |
|         | nonkudoer | 0.067  | 0.044 | 0.005 |

TABLE VIII  
PROPORTION OF USAGE OF EPISTEMIC MARKERS ACROSS CATEGORIES.

|         | Category  | Markers= 0 | Markers> 0 | %Markers> 0 |
|---------|-----------|------------|------------|-------------|
| By role | employee  | 10232      | 15258      | 59.9        |
|         | guru      | 8119       | 19370      | 70.5        |
|         | notranked | 14998      | 35458      | 70.3        |
|         | ranked    | 35803      | 69046      | 65.9        |
|         | Post Cat. | kudoer     | 4758       | 13782       |
|         | nonkudoer | 64394      | 125350     | 66.1        |

B. Usage of markers

As mentioned in §II-B, we used a lexicon of epistemic markers, which is composed by individual words and epistemic phrases that convey either certainty or uncertainty potentially. This is due to the ambiguous and polypragmatic nature of these kind of markers in discourse [22]. We applied a maximal matching procedure to count the occurrence of markers (e.g. occurrences of “I think” are counted first and then occurrences of “think”). As our main focus was put into uncertainty markers, the set of entries of markers of this kind is extensive compared to markers conveying certainty. The lexicon used is by no means exhaustive but is composed by representative markers for certainty and uncertainty (§II-B). Some examples and their raw frequencies in our dataset are shown in Table III. In some cases negated versions of epistemic phrases are in the lexicon, such as “I know” and “I don’t know” whose counts can be observed in Table III. A large proportion of epistemic markers are used, as shown in Table VIII. From the different categories, kudoer contain most epistemic markers.

C. Correlations for presence of signals and markers

Fig. 2 shows the probability of a post being kudoer based on the presence of emoticons and epistemic markers, alongside their confidence intervals. This suggest that posts with no emoticons and with markers are less likely to being given kudos ( $p < 0.001$ ).<sup>4</sup> We need to look at individual types of emoticons to be able to give an account of these correlations. Fig. 3 shows the interaction of emoticon polarity and epistemic certainty marking on kudos. If a post is certain or uncertain it gets more kudos on average than posts without any kind of epistemic marker, independently of the net polarity, except in the case of having neutral sentiment. This reinforces our inference (see Fig. 2) that posts with emoticons and epistemic markers have the highest probability of obtaining kudos.

<sup>4</sup>We used a generalized linear model with a binomial error family,  $p$ -values are adjusted for multiple comparisons [23].

Fig. 2. Probabilities of kudoer post according to signals and markers

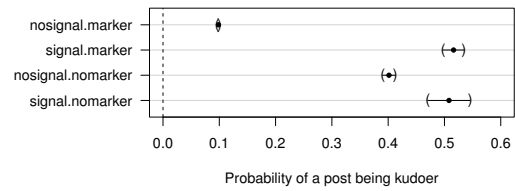
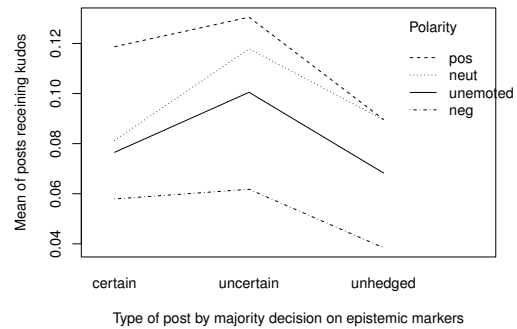


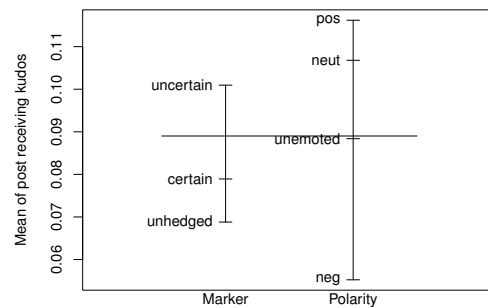
Fig. 3. Interactions between presence of epistemic markers and emoticons by polarity and their effect in the mean of kudos received by post.



The influence of the factors studied on kudoer posts mean is shown in Fig. 4. Posts with positive and neutral emotions, and posts with uncertainty markers yield values above the mean of kudoer posts (0.089). Unemoted posts get values similar to the average, while unhedged posts and posts with markers of certainty or negative emoticons have values below average. Fig. 5 summarizes the co-occurrence of emoticon polarity (or lack) and type of epistemic marker (or lack). This shows only significant comparisons (with confidence intervals) from all the pairwise comparisons of the interaction of these two features and their impact on the probability of a post being kudoer.

Where net uncertainty markers are present, negative emoticon use is the least likely to receive kudos compared to

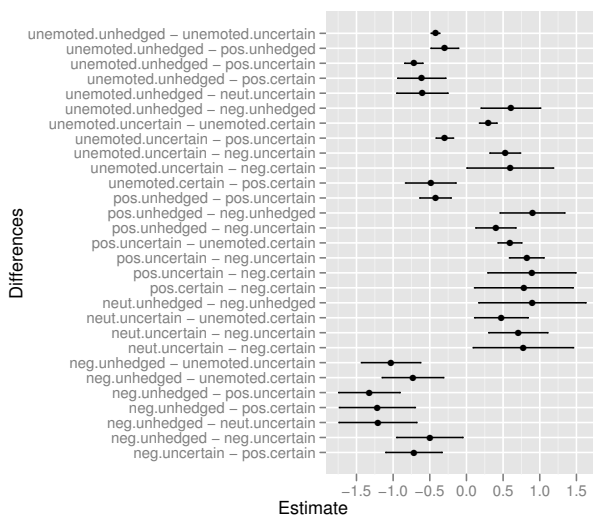
Fig. 4. Independent effect of epistemic markers and emoticons by polarity in the mean of kudos received by post.



Influence of each of the six features in posts

positive, neutral and lack of emoticons. Positive emoticon use is more likely to yield kudos compared to negative or no emoticons, but there is no significant difference with neutral emoticons. There is no significant difference between the use of neutral emoticons and no emoticons. When certainty markers are used, positive emoticon use is more likely to lead to kudo attribution than negative or no emoticons. When no epistemic markers are used, positive emoticons are more likely to lead to kudoer posts than negative or no emoticons, but there is no significant difference when neutral emoticons are used. Also, negative emoticon usage is the less likely to result in kudos compared to positive, negative and neutral emoticons. Postings with net positive emoticons that have no hedges are less likely to get kudos than those with epistemic uncertainty. When there is no emoticon use, there is more chance of a post being kudoer when an uncertainty marker is used compared to certainty markers or lack of epistemic markers. When negative emoticons dominate, there is less chance of kudos when there is no use of epistemic markers than with uncertain marker use.

Fig. 5. Difference of means and confidence intervals of a post being a kudoer by pairwise comparisons of posts with polarity and epistemic markers.



## V. CONCLUDING REMARKS

This exploration of emoticons and epistemic signals has shown trends in expertise-oriented user classifications in a technical community forum. Our hypotheses are confirmed with regard to the positive sentiment shown by kudos-receivers. Negative emoticons are mostly used by common users as they expose technical plights. Epistemic hedges are more likely to yield kudos than markers of certainty or no hedges. People appear to favor pragmatically polite uncertainty. Our findings may enhance automatic identification of likely technical forum leaders.<sup>5</sup> Much remains to analyze, particularly with respect to the scopal extent of both hedges

<sup>5</sup>Our research is funded by a Trinity College Research Scholarship and Science Foundation Ireland's grant (07/CE/11142) to the Centre for Next Generation Localisation. Thanks to Mikhail Timofeev for forum metadata.

and social signals, and also other communication contexts. We have analyzed a dimension in which communicative social signals are exploited in human behavior, even if partly outside conscious cognition. Our interdisciplinary research into these regularities may be applied to augment social intelligence in computational systems.

## REFERENCES

- [1] Mamani Sanchez, L., Vogel, C.: Emoticons signal expertise in technical web forums. In Esposito, A., ed.: Proc. 22nd Italian Workshop on Neural Networks. Smart Innovation, Systems and Technologies. (In press).
- [2] Yuasa, M., Saito, K., Mukawa, N.: Brain activity associated with graphic emoticons. the effect of abstract faces in communication over a computer network. *Electrical Engineering in Japan* **177**(3) (2011) 36–45 Translated from Denki Gakkai Ronbunshi, Vol. 129-C, No. 2, February 2009.
- [3] Logunov, A., Panchenkob, V.: Characteristics and predictability of twitter sentiment series. In: 19th International Congress on Modelling and Simulation. (2011) 1617–1623 Perth, Australia, December 2011.
- [4] Hannak, A., Anderson, E., Barrett, L.F., Lehmann, S., Mislove, A., Riedewald, M.: Tweetin' in the rain: Exploring societal-scale effects of weather on mood. In: Proc. of the Sixth International AAAI Conference on Weblogs and Social Media, AAAI (2012) 479–482
- [5] Bifet, A., Holmes, G., Pfahringer, B.: Moa-tweetreader: real-time analysis in twitter streaming data. In: Proc. of the 14th International Conf. on Discovery Science. DS'11, Berlin: Springer (2011) 46–60
- [6] Kucuktunc, O., Cambazoglu, B.B., Weber, I., Ferhatosmanoglu, H.: A large-scale sentiment analysis for Yahoo! answers. In: Proceedings of the fifth ACM international conference on Web search and data mining. WSDM '12, New York, NY, USA, ACM (2012) 633–642
- [7] Read, J.: Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In: Proc. of the ACL Student Research Workshop. (2005) 43–48
- [8] Janssen, J., Vogel, C.: Politics makes the Swedish :-)) and the Italians :-(. In: Proc. of LREC – EMOT 2008, Sentiment Analysis: Emotion, Metaphor, Ontology and Terminology. (2008) 53–61
- [9] Vogel, C., Janssen, J.: Emoticonsconsciousness. In Anna Esposito, Hussain, A., Marinaro, M., Martone, R., eds.: *Multimodal Signals: Cognitive and Algorithmic Issues*, Berlin: Springer (2009) 271–287
- [10] Ching, G.S.: Blog assisted learning: Experiences in learning business English vocabularies. *International Journal of Research Studies in Educational Technology* **1**(1) (2012) 3–12
- [11] Tu, C.H.: The impacts of text-based CMC on online social presence. *The Journal of Interactive Online Learning* **1**(2) (2002) 1–24
- [12] Tu, C.H., McIsaac, M.: The relationship of social presence and interaction in online classes. *American Journal of Distance Education* **16**(3) (2002) 131–150
- [13] Lui, A.K., Tsang, P., Kwan, R., Ng, S.C., Cheung, Y.H.Y., Choy, S.O.: An evaluation framework of expertise presence in computer conferences. *British Journal of Educational Technology* **38**(6) (2007) 1020–1036
- [14] Huang, A.S., Yen, D.C., Zhang, X.: Exploring the potential effects of emoticons. *Information & Management* **45** (2008) 466–473
- [15] Rubin, V.L.: Identifying Certainty in Texts. PhD thesis, Syracuse University, Syracuse, NY (2006)
- [16] Kratzer, A.: The notional category of modality. In Eikmeyer, J., Rieser, H., eds.: *Words, Worlds and Contexts*. Berlin (1981) 38–74
- [17] Goatly, A.: *The Language of Metaphors*. Routledge (1997)
- [18] Vogel, C.: Dynamic semantics for metaphor. *Metaphor and Symbol* **16**(1 & 2) (2001) 59–74
- [19] Moxey, L.M., Sanford, A.J.: *Communicating Quantities: A Psychological Perspective*. Lawrence Erlbaum (1993)
- [20] Goffman, E.: *The Presentation of Self in Everyday Life*. New York: Doubleday (1956)
- [21] Cheng, R., Vassileva, J.: User motivation and persuasion strategy for peer-to-peer communities. In: Proc. of the 38th Annual Hawaii International Conference on System Sciences - Volume 07. IEEE Computer Society (2005) dx.doi.org/10.1109/HICSS.2005.653
- [22] Hyland, K.: Hedging in Scientific Research Articles. *Pragmatics & Beyond*. John Benjamins (1998)
- [23] Bretz, F., Hothorn, T., Westfall, P.: *Multiple Comparisons Using R*. Chapman & Hall/CRC: Boca Raton, FL. (2011)