Extracting useful information from Facebook posts: A Cognitive-Linguistic Analysis

Sudipa Batabyal, Somsukla Banerjee, and Somprakash Bandyopadhyay

Abstract—Social media has emerged as a significant communication platform in recent times where millions of people across the world share thoughts and ideas and interact with each other via social networking sites. Researchers examining communication in social networking sites have stated that communication in these forums has a strong phatic (small-talk) element and there is very little substantive information content in posts. This paper re-examines the view and suggests that posts in social networking sites is not always phatic. We argue that information content of a certain "post" needs to be examined from the perspective of both the sender and the receiver of information. For the purpose of our study, we conducted an empirical study, where users rated posts on their Facebook wall on the basis of their "information content" and "phatic element". Thereafter we undertake a content analysis of posts using tools from pragmatics and demonstrate how users extract implicit information from the content of these posts. We conclude that a detailed and in-depth linguistic analysis based on principal of pragmatics needs to be undertaken to design algorithms for automatic identification of *meaningful posts* from "apparently phatic" social media posts that would help in making useful organizational decisions.

Keywords—Phatic Expressions, Pragmatics, Social Media.

I. INTRODUCTION

SOCIAL media are fundamentally changing the way we communicate, collaborate, consume, and create. They represent one of the most transformative impacts of information technology on any organization, both within and outside organizational boundaries. Nevertheless, a study by Stanford University's Rock Center for Corporate Governance on social media conducted in 2012 revels that, while 90% of respondents from the top management of different companies claim to understand the impact that social media can have on their organization, only 32% of their companies monitor social media to detect risks to their business activities and 14% use metrics from social media to measure corporate performance.

The primary reason behind this is, researchers examining communication in social networking sites have stated that

Sudipa Batabyal is with Cognizant Technology Solutions, Kolkata, India (e-mail: sudipa.batabyal@gmail.com).

Somsukla Banerjee is with Social Informatics Research Group, Indian Institute of Calcutta, India. (e-mail: somsukla2005@gmail.com)

Somprakash Bandyopadhyay is with MIS Group, Indian Institute of Management Calcutta, India (e-mail: somprakash@iimcal.ac.in).
[corresponding author] Phone: +91-92316-83443; e-mail: somprakash@iimcal.ac.in

communication in these forums has a strong *phatic* (small-talk) element where it only serves the purpose of maintaining and defining social relationships. Researchers have overlooked the information content of social networking posts and emphasized mostly on the phatic component of the posts (Miller, 2008). However these phatic posts have important social functions in strengthening the bond between the users (Radovanovic, 2012).

This paper re-examines the view and suggests that posts in social networking sites have a significant information content in addition to the phatic aspect. For the purpose of our study, Facebook posts were classified into seven categories and we conducted an empirical study, where users rated posts on their Facebook wall on the basis of their "information content" and "phatic element". Thereafter we undertake a content analysis of posts using tools from pragmatics and demonstrate how users extract implicit information from the content of these posts. We conclude that an in-depth linguistic analysis based on principal of pragmatics needs to be employed to identify *useful posts* from "apparently phatic" social media posts to generate automatic inferences.

II. PHATIC COMMUNICATION IN SOCIAL NETWORKING

Phatic communication or small talk came into focus with the work of Anthropologist Bronislaw Malinowski (Malinowski, 1923). In Malinowski's Phatic Communion, the term communion highlighted the relational rather than the informational nature of small talk. Linguists later preferred the term Phatic Communication.

Phatic messages create an atmosphere for a communication by acting as an icebreaker, ensuring that exiting communication channels are kept 'open and 'usable'. This is particularly relevant in examining communication in social media. Scholars investigating social media communication using different social networking platforms have often stated that communication in social networking sites is essentially phatic. In other words, in these platforms the shared information or the content of the message is not the key factor but to "say something in order to stay connected". Vincent Miller has argued that phatic communication increasingly dominates conversation in social media culture and the maintenance of the network itself is the primary focus rather than exchanging substantive content (Miller, 2008). It has been argued that Phatic posts in social media communication have important semantic and social functions for interlocutors

in strengthening social bonds and network ties (Radovanovic, 2012).

In this paper we adopt the position that we need to consider a communication event in social media in the entirety of its context to determine its information content and social value. We need to rethink of the statement that phatic posts are informationally empty and have only social functions. When someone posts " Enjoying the new range hungry jacks burger", on a social networking site we need to consider it from the perspective of both the contributor- one who wrote the post and the recipients of the posts. Although the contributor had no intention of disseminating information about the burger and his primary intention was to signal a connected presence but to some of his recipients this may have brought in some information such as a) hungry jacks has introduced some new burgers in their range b) the burgers are tasty. Therefore it will be incorrect to say that posts in social media have no information content. Like all communication, it is a two way process where we have the one who communicates at one end and recipients at the other. The information content or the lack of it needs to be considered from both their perspectives. We need to reexamine the view that communication in social media is essentially Phatic. What may appear Phatic or information less at the surface level may not be so if we attempt a careful and deeper analysis of it.

III. USER RATINGS OF FACEBOOK POSTS: AN EMPIRICAL STUDY

On examining several posts of social networking sites, we found that there was a strong information component present in most of them. We wanted to ask the following questions:

o Is communication in social networking essentially phatic?
o Can we distinguish between phatic and non-phatic in social media communication?
o Can we identify certain types of posts that may appear phatic to some recipients but yet have information content for other recipients ?

We conducted an empirical study and obtained responses from users of social networking sites. Five active Facebook users were asked to assign categories to Facebook posts on their newsfeed. Based on their judgment the posts were assigned the following categories:(Examples in appendix)

1. Generic post, image, video
2. Personal image
3. Experience sharing
4. Advertisement
5. Check in
6. Any other
7. Post by Facebook group or media

The respondents who participated in our empirical study were asked to assign one of the above-mentioned categories to the posts that appeared on their wall. Following that we asked them certain questions to get insight about the function of these posts in their lives. We wanted to validate our claim that social networking posts are not entirely phatic and do have a substantive information content. The following questions were asked to the respondents:

TABLE 1
QUESTIONS TO USERS ON FACEBOOK POSTS

S.L. No.	Questions
Q1	Does the post give you any new information about the contributor?
Q2	Does the post give you any new information about other things (external to the contributor)?
Q3	Does the post make you interested in the things described by the contributor and explore further?
Q4	Does the post make you interested in the life and activities of the contributor?
Q5	Reading the post makes you feel more connected to the contributor?
Q6	How much are you inclined to LIKE on the post?
Q7	How much are you inclined to contribute to the post?

Questions 1-3 in the above list examined the information content of the posts. If a certain post provides information to a user about the contributor or about other things, it will not be fair to say that the post is entirely phatic.

Earlier research has claimed that phatic posts in social networking sites have important social functions for users by enhancing their social relationships (Radovanovic, 2012). It has been claimed that social media tools support people in building, maintaining and benefiting from social relationships (Jung et al, 2013). We wanted to examine the precise role of phatic communication in building social relationships. In order to understand this aspect we included questions 5-7 in our study. If a post elicits a positive response from the user in form of a LIKE or comment, it is possible that he establishes a bond with the contributor.

Posts on social networking sites that have information content may in addition have a phatic function. If a certain post increases the interest level of a user about the contributor and he is inclined to explore further, we would say that this post has both information and phatic content. Question 4 and 5 in our list of questions examined this aspect.

The following section describes the empirical study and the methodology used for data collection.

IV ANALYSIS AND RESULTS

A. Data Collection

For purposes of the empirical study, a questionnaire was prepared and circulated among various active Facebook users. The objective of the questionnaire was to help the users perform a subjective categorization of their Facebook posts and rate those posts in a scale of 1 to 10 against the questions asked in the survey. The rating provided by the users helped to get a great insight in revealing various types of information component out of the daily mundane, from a sender's or receiver's context.

Ten active Facebook users, at the age range of 20 to 32 years participated in our study. The first section of the questionnaire included questions to identify the **category** of Post such as, (i) Generic post, image, video, (ii) Personal image (iii) Experience sharing (iv) Advertisement (v) Check in (vi) Post by Facebook group or media (vii) Any other .

The second section of the questionnaire included a list of questions (Table 1) to identify the importance or influence of

that post to a user. Accordingly they were requested to rate the posts against the set of questions in a scale of 1 to 10, which we have codified: 1 to 3 as LOW, 4 to 6 as MEDIIUM and 7 to 10 as HIGH. The participants were requested to complete the survey and rate approximately 50 to 60 of their latest Facebook posts. While doing so, for any post rated in a scale of 6 to 10 (which we call non-phatic) they were requested to provide a reason of why those posts are important to them. Later, by combining the survey data we were able to get a good representation of the information available in social media and also which categories of posts users found most useful.

B. Aggregated Results

Upon examining several posts of social networking sites, we found that there is a strong information component present in most of them. The following chart is an aggregated representation of survey results across all categories combining the ratings of all users. The result shows that, against every question (Table 1), there are a considerable percentage of Medium and High ratings provided by the users. This shows that users do get potential information content from their day to day mundane in social networking sites.

In the next section, we further analyze the survey results across each category. The charts below emphasize our interpretation of the posts, where the individual pattern noticed from the survey response of each user is aggregated for each category.

C. Generic Post/Image/Video

This seemed to be the most common and frequently occurring pattern across most people's timeline. The responses from the individual users were plotted in a graphical pattern.

As seen from the chart in Figure 2,

- The generic posts (such as, video, images, inspirational messages, newspaper cuttings etc.), does not provide much insights or additional details about the contributor. For high percentage of "LOW" for the 1st question emphasizes the same. This indicates that most of these posts are not related to the user and do not provide any new information about the contributor.
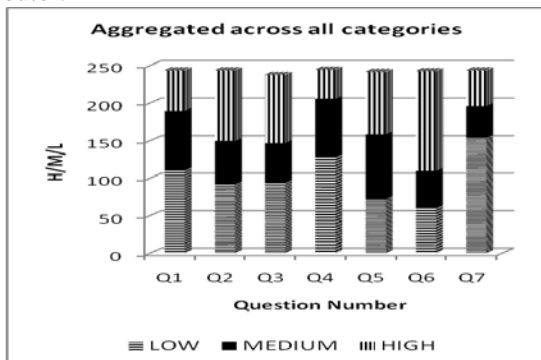

Fig. 1 Aggregated view: More than 50% posts are rated as non-phatic by the Facebook users
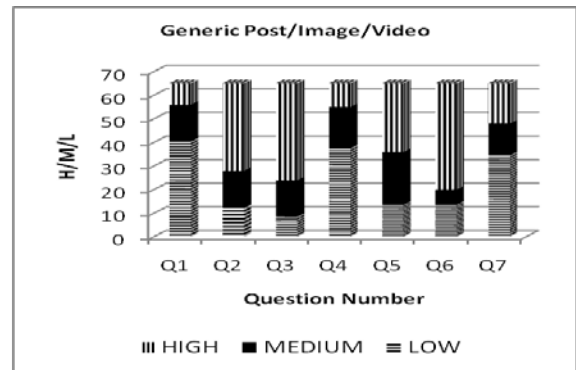

Fig. 2 User Rating of Generic / non-personal Posts

- However when we look at the responses of Q2, it is noteworthy that the user seems to agree that these generic posts provide details on information around aspects other than the contributor. For majority of the users, a predominant percentage of "high" rules the charts.

- The response to Q3 is thought-provoking. Although the users here are sharing information not related to them, they do share generic posts of common interests with their friends. Most of the time, the recipients are keen to explore the shared interests from the contributor. The above observations certainly indicate the strong informational content of these posts.

- The generic posts, certainly, does not provoke the user to explore further about the contributor's activities, as seen in Q4. However, these details certainly enhance the users' bond with the contributor. This is revealed in Q5, where the majority of the users are in agreement that these posts do help them to connect to the users.

- Questions 6 exhibits the typical pattern of social network where a user is more inclined to show his / her support towards the posts with a "like" and add their viewpoints to it. However, users are usually less likely to contribute, as indicated in the responses of Q7.

D. Experience Sharing

These posts deal with sharing photo / caption about something that they have experienced. This varies from a small incident noticed on the roadside when going to office to detailed experience on a vacation. The responses from the users for these categories are depicted in the chart below:
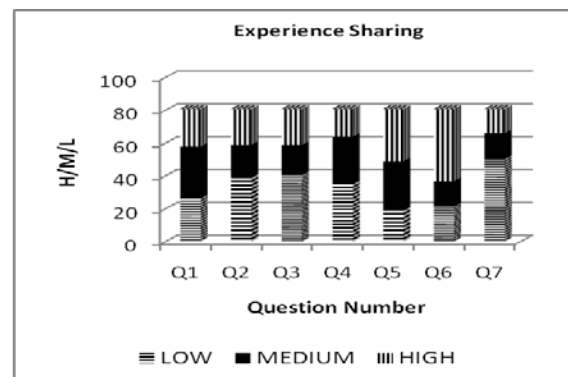

Fig 3. User Ratings of Posts on Experience Sharing

From the chart, we can see that:

- For Q1, most of the users have responded with "MEDIUM" and "HIGH" indicating that these details does provide any additional information about the contributor
- Incidentally, the responses to Q2, Q3 and Q4 depict not-so-high information content. This indicates that most of the information shared in these experience sharing posts are too personal to provide substantial detail for the topics shared.
- However, as a natural consequence of personal experience sharing posts, readers do feel connected to the contributor, enhancing the social relationship among the users. This is depicted in the response of Q5.
- The pattern of Q6 is similar to Q5, where the users are more inclined to 'LIKE' the post. This seems to be following a more social behavioral pattern of complimenting someone even though the information is not surprising to the readers.

E. Personal Image

- These images are generally associated with someone updating their profile pictures or adding a personal album. Analyzing the chart, we see that:
- The pattern for this category tends to exhibit almost similar pattern of "Experience Sharing".
- We notice that for Q1 to Q5, we see a distributed response and are more biased towards low and medium rather than high. From this observation, it can be inferred that, by seeing someone's personal image in the News Feed, the reader does not get much new information about the user or about other things external to the user. Also, from the response of Q5, it can be concluded that personal image posting does not always help the reader to feel connected to the users.
- The pattern for Q6, where the user is more likely to LIKE that post, further enhancing our belief that the social behavior of humans influences the 'Likes', even though their connection / bonding with the contributor does not increase.
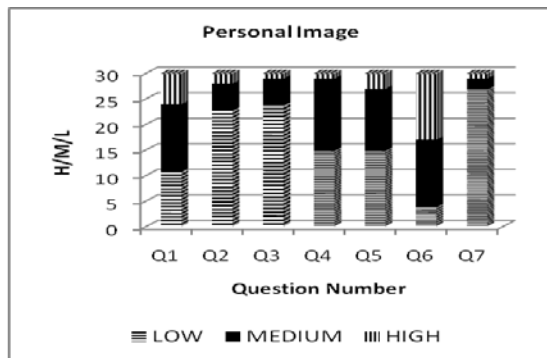

Fig. 4 User Ratings of Posts on Personal Images

F. Check IN

Check-in posts provide information related to the whereabouts of the contributor: a feature in Facebook that is frequently used, especially when on vacation or at a social gathering. Check-ins is common in restaurants, airports or vacation places. Coupled with the "tag-in" feature, this has become a trending pattern in several timelines.
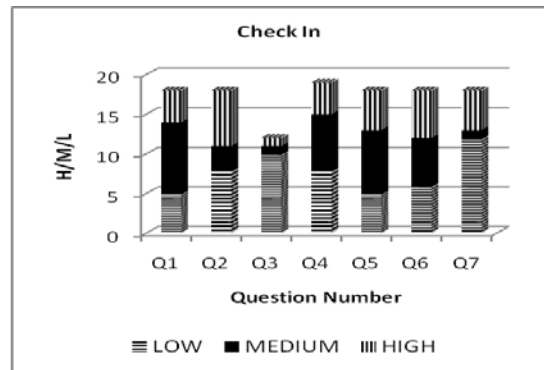

Fig. 5 User Ratings on Check-in posts

- These posts help the readers get new information about the contributor and about other things, such as, the places they are visiting, or activities being done by them. The Medium to High rating provided against Q1 and Q2 do support this hypothesis. We notice that the information provided about the external world in this context of sender usually appears interesting to users (Q2). However, users are not inclined to explore further, as evident from the responses of Q3.
- Q4, Q5 and Q6 exhibit similar pattern of user response. Users tend to rate maximum posts from Medium to High category. The reader of the posts becomes interested about the life and activities of the contributor, feels connected to the contributor and are more inclined to like the posts rather than commenting on the posts. This certainly helps to increase social bonding.

G. Advertisement

These are Posts from corporate sales and marketing teams to provide more insights into a product or service.

The interesting pattern that we notice in this category against Q2 and Q3 (Figure 6) is the user's urge to explore the advertised product further. The users not only get new information about the product / service (Q2), but also feel interested in the things described by the advertisement and like to explore further (Q3). This seems to be the response from majority of the users. This is good news for advertisers and corporate using the Facebook platform for their marketing purpose.
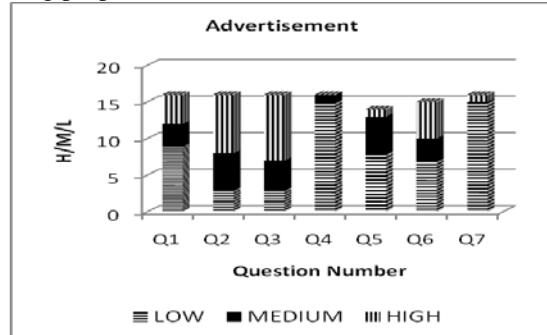

Fig 6. User Ratings on Posts related to Advertisements

H. Post by Facebook Group/ Media/ Personality

This is similar to the "follow" feature in Twitter: the user subscribes to the feed from a brand / personality / product or any special interest groups.

•The chart shows posts from the groups that the users are interested in. So, it is not surprising to see that users connect more to the contributor in these cases, due to common interests. This is evident from the responses of Q2 and Q3, where the user seems to be getting good details about the information provided by the group and also enables to user to explore these further.

•Ratings of Q5 and Q6 indicate that the users tend to like and feel more associated towards a brand or group of their choice.

I. Discussions

Based on the above analysis, we can infer that:

✓ There are lots of conversations with "low" information contents, termed as phatic communication. Social media researchers universally agree upon the fact that phatic communication has a very important social role to perform. Social Networking sites facilitate phatic communication where a user writes obvious statements or even simply signal his presence without engaging in overt communication. Also, we have seen that the social nature of users are exhibited in posts, where there is a tendency to "like" a post from friends / relatives even if it does not provide much additional information.
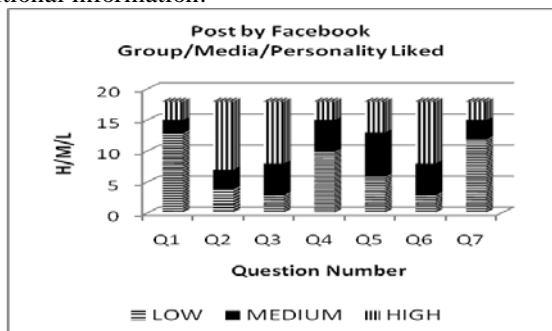

Fig. 7 User Ratings on Posts by Groups / Brands, Subscribed by Users

✓ At the same time, there are large number of information-rich generic posts, including both personal experience and general advertisements, and posts generated by favorite groups / brands. Users are inclined to get new information and usually like to explore them further. These posts are non-phatic in nature, although it serves the function of social bonding as well.

Since posts in social networking posts have multiple consumers, we cannot objectively comment on the phatic/non-phatic-ness of posts. Much depends on the context and the recipient of the post. Therefore our central claim in this paper is that, all phatic posts in social networking sites may not actually be phatic and needs to be examined from the perspective of the recipient. In order to analyze the linguistic data collected from social networking sites, we need to use constructs from pragmatics- the sub discipline of linguistics-that studies language in a context. This will be examined in the next section.

V. PRAGMATIC ANALYSIS OF FACEBOOK POSTS: EXTRACTING IMPLICIT INFORMATION

Pragmatics studies language from the perspective of the language users and the context of language use. Pragmatics focuses on what is not explicitly stated when two people communicate with each other and the interpretation of situational contexts. In real life communication situations, participants often engage in the task of inferencing in order to derive implicit information from what is explicitly stated.

Let us consider the following posts that are typical examples of Facebook Status Updates:
1) " didn't know Maggi tastes go good in fever"
2) "I regret cooking eggplant for dinner today. They taste terrible at this time of the year"
3) "Listening to new album by Air"
4) "My new Macbook Air rocks".

The posts 1-4 exemplify the claim that the creators of these posts have written about their trivial activities just to indicate their online presence. Their purpose is to foster and maintain social bonds and they have no substantive information content. But this is only partially correct and we would like to suggest that despite their predominant social function they do have some significant information content and we should not ignore the informational aspects of what apparently appear to be Phatic posts.

If we examine examples 1-4 cited above, it is quite possible that the creators of the posts did not put much intellectual introspection while posting them. They only wanted to stay connected by updating their network or circle of friends about what they were doing. However the nature of the social media communication is such that one reaches out to many individuals at the same time and depending on who the audience is, it may arouse some reflection on the part of the audience.

In order to further substantiate our argument let us introduce the concept of presupposition. Presupposition is an implicit assumption about the world or background belief relating to an utterance whose truth is granted in discourse.

For example, if one says, "John no longer writes fiction", it presupposes the fact that John once wrote fiction. Stalnaker (1973) has defined presupposition as: "A speaker presupposes that P at a given moment in a conversation just in case he is disposed to act, in his linguistic behavior, as he takes the truth of P for granted, and as if he assumes that his audience recognizes he is doing so".

So if we examine the Status updates 1-4 we will be able to work out a presupposition that is associated with each of these:

"I didn't know that Maggi tastes so good in fever"
Presupposition: Maggi tastes good in fever
 I have fever
"I regret cooking eggplant for dinner." **Presupposition:** I cooked eggplant for dinner
Listening to the new album by Air
Presupposition: There is a album released by Air
My new Macbook Air rocks
Presuppostion: I possess a Macbook Air

If we carefully look at the status updates we have listed in 1-4 and the presuppositions included in them, we will note that each one of them has some information content that is not explicit in the sentences but implicitly stated. Although they appear to be trivial and have been posted by the communicator with the intention of remaining connected to others however they do convey information. Once a post is created on a social networking site, it reaches a vast network of people. Depending on the audience this post may have some value in terms of information content. for example, as an audience, I read (1) and note that the communicator has fever and perhaps will not be able to join me in an outing that I intended inviting her to. In addition, I make a note of her preference for Maggi. (2) will perhaps ward me off eggplants for some time. Similarly all the posts 3and 4 have some information content about launch of a recent album, efficiency of a computer etc.

Social media including social networking sites are an important source from where gather information about people, their attitudes, likes and dislikes as well as what is happening around us in the world. Some of this information is obtained from regular status updates that users happen to post in these sites. Therefore it is not right to say that these posts have only a phatic function to perform. These days a lot of companies are relying on data from social networking sites to evaluate a potential employee. In this context, we believe that a lot of information about an individual, his tastes and preferences and value system can be obtained by looking at these posts on social networking sites. Therefore something that a person has posted without much introspection with the sole purpose of staying connected may have more serious consequences. Depending on the recipients of the post, while some may read it and perhaps press a LIKE button and forget about it, others may give it a more serious thought. As a single post from an individual reaches multiple recipients simultaneously, it is unlikely that it will elicit the same response from all individuals. For some of the recipients there may actually be a strong information content in these posts. It is possible that these posts contain valuable information about the contributor, his state of mind, his tastes and preferences or information about people in his network. Such information is of value to enterprises who strive to gain insights about consumers from posts in social media.

Researchers in Social Media Analytics are striving towards extracting important information content from semi structured and unstructured data. In these approaches, words that denote important concepts are identified, extracted and classified into entity types (Cortis,2013). Microblogs are an important source of information in emergency management as lots of situational information about a crisis situation is shared. Using Machine Learning techniques information is extracted and classified from these microposts (Schulz et al. 2014).

However, researchers have not yet considered the issue of pragmatics in the context of analysis social media data. It will be, therefore, worthwhile to employ analytic techniques to extract the presuppositions from the data that is implicit in social media posts. Although till date we don't have tools to automatically extract all types of presupposition triggers, NLP researchers have been successful in extraction of presuppositions that are triggered by definite descriptions.

VI CONCLUSION

In this paper, we have reviewed the concept of phatic communication as proposed by Malinowski and its significance in the age of digital and social media. We have argued that a significant volume of posts in social networking sites which have been classified as phatic, does actually have substantive information content. Since a post in social networking sites has multiple consumers, we cannot objectively say that the post is phatic unless we consider it from the perspective of the recipient. We have conducted an empirical study where Facebook posts were classified into seven categories and users have rated posts on their Facebook wall on the basis of their "information content" and "phatic element". We have demonstrated that social media posts do have significant information content. Our further analysis suggests that a post may not be phatic if we can derive a presupposition from it. By means of extracting presuppositions from Facebook Status Updates, we have shown one way in which we can look for information content in posts that appear to be phatic.

We are working towards developing more sophisticated linguistic analysis based on principal of pragmatics to identify *useful posts* from "apparently phatic" social media posts to generate automatic inferences. If we have tools at our disposal that can extract presuppositions, it will be of immense value to people relying on social media data for commercial purposes. It will also lead to an understanding of human behavior and communication patterns. This has enormous potential for commercial and social enterprises as the insights about peoples' thoughts, attitudes, opinions and preferences can lead to important insights.

REFERENCES

[1] W. Jung, R. Gray, N. Ellison and C. Lampe. 2013. Dimensions of Social Capital and Facebook Favors. Proceedings on Annual Conference on Human Factors in Computing (CHI), ACM: New York.
[2] S. Levinson. 1983. Pragmatics. Cambridge Textbooks in Linguistics, Cambridge University Press, Cambridge, UK.
[3] Licoppe, and Z. Smoreda. 2005. Are Social Networks Technologically Embedded? How Networks are Changing Today with Changes in Communication Technology, Social Networks,27 (4): 317-335.
[4] B. Malinowski. 1923. The Problem of Meaning in Primitive Languages, in Charles. K. Ogden and Ian Richards (eds.), The Meaning of Meaning. Routledge, London, 146-152.
[5] L. Manovich. 2001. The Language of the New Media, MIT Press, Cambridge, Massachusetts.
[6] V. Miller. 2008. New Media, Networking and Phatic Culture, Convergence: The International Journal of Research in New Media Technologies, Volume 14(4) :387-400.
[7] Radovanovic,and M. Ragnedda. 2012. Small talk in the Digital Age: Making Sense of Phatic Posts," in Making Sense of Microposts (#MSM2011), Matthew Rowe, Milan Stankovic, Aba-Sah Dadzie (eds).10-13.
[8] Sarjanoja., Isomurusu,H.,Hakkila, J. 2013. Small Talk in Facebook-Phatic Communication in Social Media in Proceedings of Academic Mindtrek, ACM.
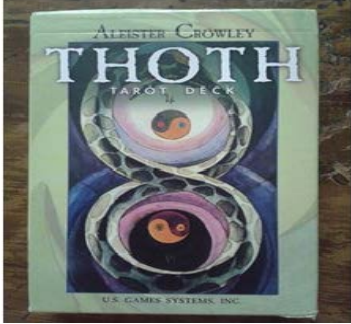
[9] R. Stalnaker. 1973, Presuppositions, Journal of Philosophical Logic. 2: 447-457.
[10] V. Wang, J. Tucker and K. Haines. 2012. Phatic Technologies in Modern Society, Technology in Society, Volume 34 (1-2): 85-92.
[11] V. Wang, J.Tucker and T. E. Rihil. 2011. On Phatic Technologies for Creating and Maintaining Human Relationships, Technology in Society 33 (1-2): 44-51.
[12] Yus, 2011. Cyberpragmatics : Internet Mediated Communication in Context John Benjamins, Amsterdam/Philadelphia.

Appendix: Examples of Facebook Posts

Example: **Generic Post/Image/YouTube Video**

Example: **Experience Sharing**

"Three more days left and it feels so good to have an entire table of food platter spread out before me as parting gift. The heavenly feeling of gorging on delicacies"

NOTE:

Textbook definition of "Information" can be illustrated as: "facts provided or learned about something or someone" or "what is conveyed or represented by a particular arrangement or sequence of things". Facebook users get ample of information content or clue to derive further information from the various posts they perceive in the site. This certainly helps the users to be more involved, more connected, more knowledgeable, thus increasing their social capital and decreasing loneliness.

For example, Post # 1 doesn't tell anything about the book apart from the book cover. However, a general user may gather some knowledge about (i) the interest of the sender, and, (ii) the subject mentioned by exploring further.