

Investigating Linguistic Indicators of Generative Content in Enterprise Social Media

Elisavet Averkiadi¹, Wietske Van Osch², Yuyang Liang³

¹ Michigan State University, East Lansing, USA
averkiad@msu.edu

² HEC Montreal, Montreal, Canada

² Michigan State University, East Lansing, USA
wietske.van-osch@hec.ca

³ Michigan State University, East Lansing, USA
liangyuy@msu.edu

Abstract. Teamwork is at the heart of most organizations today. Given increased pressures for organizations to be flexible, and adaptable, teams are organizing in novel ways, using novel technologies to be increasingly agile. One of these technologies that are increasingly used by distributed teams is Enterprise Social Media (ESM), web-based applications utilized by organizations for enabling communication and collaboration between distributed employees. ESM feature unique affordances that facilitate collaboration, including interactions that are generative: group conversations that entail the creation of innovative concepts and resolutions. These types of interactions are an important attraction for companies deciding to implement ESM. There is a unique opportunity offered for researchers in the field of HCI to study such generative interactions, as all contributions to an ESM platform are made visible, and therefore are available for analysis. Our goal in this preliminary study is to understand the nature of group generative interactions through their linguistic indicators. In this study, we utilize data from an ESM platform used by a multinational organization. Using a 1% sub sample of all logged group interactions, we apply machine-learning to classify text as generative or non-generative and extract the linguistic antecedents for the classified generative content. Our results show a promising method for investigating the linguistic indicators of generative content and provide a proof of concept for investigating group interactions in unobtrusive ways. Additionally, our results would also be able to provide an analytics tool for managers to measure the extent to which text-based tools, such as ESM, effectively nudge employees towards generative behaviors.

Keywords: Enterprise Social Media, Generative Interactions, Text Classification, Virtual Teams, Team Collaboration, Corporate Innovation.

1 Introduction

ESM are web-based applications that offer users various features to enable them to effectively communicate with each other, network, organize, leverage information available on the platform, and collaborate. ESM have a set of affordances [11] that promote collaborations to occur. By extension, it therefore seems to have the potential to foster group generative collaborations - group exchanges that involve the creation of innovative ideas and solutions. One of these unique affordances of ESM, namely visibility, allows all contributions to the platform to become visible to anyone who has access to the system. Not only has this affordance been shown to enhance collaboration, and thus possibly generative collaborations, but also offers a unique opportunity to study such group behaviors. Given the visibility of text-based interactions between users and within groups, server-side data from ESM can be used for research purposes, thus eliminating the bias of self-reporting methods and allowing researchers to explore important antecedents to behaviors in unobtrusive ways. This gives us an opportunity to improve the existing theoretical understanding of the nature of group interactions that occur on ESM platforms, yet also to improve such interactions on ESM, and other similar collaboration tools.

Our objective in this preliminary study is to understand the nature of group generative interactions through their linguistic indicators. There is copious server-side data to be leveraged from ESM, in particular the text-based asynchronous, and synchronous, messages that are exchanged within groups, specifically as this information pertains to the antecedents of effective creative collaboration. To conduct this research, we used an ~1% subsample of all group interactions from data provided by an ESM platform used by a multinational organization, and applied machine-learning models to classify the text data as generative or non-generative interactions and extracted the linguistic antecedents for the classified generative content.

2 Theoretical Background

2.1 Generativity and Group Generative Interactions

Generativity was first conceptualized in 1950, in work on the stages of psychosocial development, by psychoanalyst Erikson (1950) [6]. It has since been leveraged repeatedly in the social science and humanities disciplines. These disciplines have utilized this concept to refer to the creative progress and social change; a meta review of the major theories of generativity are presented by Van Osch (2012) [17] and Van Osch & Avital (2010) [16]. Generative interactions in virtual teams are a process of creating new knowledge, reconceptualizing a problem and/or a solution. In essence, generativity is defined as creating, originating, or producing [2, 21]. Generative interactions have further consequences, such as revealing tensions among users that were otherwise unknown, cross-boundary differences are highlighted, new perspectives are shared, and various other forms of creativity stimulants are exposed to an online team [3, 9]. By

focusing on these interactions among employees, we could investigate a critical stimulant for innovations in organizations [16].

Generative interactions are conversations that aim to generate novel concepts, ideas, or solutions [16]. Rather than a single type of interaction, Tsoukas (2009) [15] inferred from creative cognition research [5] three distinct forms of conceptual change, which have received a great amount of attention. These typologies of generative interactions can help us understand the different ways in which novel concepts emerge in the context of generative interactions. One form of generativity, expansion, involves recycling and expanding the use of existing concept from its core use, in order to match a new situation. Reframing, a second form of generativity, is a type of generative collaboration that frequently involves creatively deconstructing an existing concept and reconstructing it to fit a new situation. The third type, combination, involves combining two or more already existing concepts in new ways.

Generativity can thus stem from combining existing concepts in new ways [22], expanding the use of an existing concept from its core use to match a new situation (i.e., expansion), or by creatively deconstructing an existing concept and reconstructing it to fit a new situation (i.e., reframing) [16]. Reframing is a much more disruptive form of generativity, as it often challenges the status quo [16]. We operationalize these three types of conceptual change to identify generativity in text data.

2.2 ESM and Generative Interactions

Research thus far has accumulated evidence that ESM are an appropriate tool to facilitate information exchanges within teams, and thus, by extension may facilitate group generative interactions [12, 18, 20]. ESM platforms enable an information contribution process that results in an eco-system for supporting the generation of innovative concepts [4, 10]. However, it is not clear how, why, when, and to what extent these benefits occur. The scarcity of evidence provides the impetus for this investigation with the aim of finding ways to identify occurrences of generative interactions as a first step toward enabling improved such interactions in ESM.

Users of ESM platforms are able to communicate with other users through synchronous and asynchronous communication. Given increased pressures for organizations to be flexible and adaptable, teams need to organize in increasingly agile ways, using technologies such as ESM to facilitate more flexible communications and collaborations. ESM, as an integrated social media platform for internal communications [13], allows both synchronous and asynchronous communication (e.g., posts and threads). However, despite the mode of communication selected within the ESM, all communications are text-based thereby allowing team members to curate and edit messages between each other. These messages also persist – they are there to refer back to at a later time, and accessible to all team members. Within these text-based messages between employees, there is copious information that could be analyzed to understand the nature of these interactions, what makes them effective, and identifying the antecedents of successful creative interactions.

Generative interactions are a critical antecedent for innovation to occur [2]. They are an important component of group collaborations, as a company’s ability to innovate is

closely linked to their chances to survive and thrive [1, 7, 8, 14]. ESM have a lucrative impact on companies and the economy worldwide. Four out of five companies use ESM, and an estimated \$100 billion is invested on ESM worldwide [19]. Companies investing in implementing ESM as their collaboration tool are particularly interested in generative interactions. All types of generative interactions (i.e. expansion, reframing, and combining) result in some form of new knowledge, which overtime, could become competitive value for an organization [8]. Breakthrough solutions are more likely to occur through generative interactions; they increase the likelihood of innovation [15].

3 Method and Results

3.1 Data

The data used for this study is provided by a multinational organization that researches and consults in the domain of human-computer interactions. Additionally, the organization builds technology and develops office space solutions for a variety of client domains: corporate offices, healthcare, educational institutions, and government institutions. The organization has over 80 locations around the world, and more than 11,00 employees across these locations. The organization launched an ESM tool with the objective of enabling connections, communication, and collaboration, among employees, in an effective way across its locations around the world. The ESM platform had accumulated 10,000 users over the course of five years. Of these 10,000 users, 91% (9,000 users) of its users are members of teams, who actively participate in group discussions.

Using data from this ESM, with permission from the multinational organization, offers a relevant object of study: its employees are distributed across locations and time zones, the users have been utilizing the platform for five years, and the data includes active employee teams. These criteria make the data relevant for our exploration of the linguistic indicators of group generative interactions. The data included 20,000 threads, of which 219 (~1%) were used for our exploratory study.

3.2 Method

Data Preparation. Before implementing a machine-learning classifier, the data was prepared by labelling text from the group threads with a code for the presence or absence of generative activity. Given the small sub-sampled used in this study, the three types of generative activity aforementioned were collapsed into one category. The coding scheme used for labelling the data can be seen in Table 1.

Table 1. Code Scheme for Labelling.

Type	Code	Description
Generative Activity	1	Presence of any of the three typologies of generativity (expansion, reframing, combination).
Non-Generative Activity	0	Absence of any of the three typologies of generativity (expansion, reframing, combination).

We trained human coders to identify the text that contained elements of one of the three types of generative activity (reframing, expanding, combining), with the use of a coding manual that included definitions and examples of each.

Subsequently, the text was lemmatized – a method of reducing a word to its base form. We also extracted features from the text using the ‘bag of words’ method, which represents the text as a numerical description of its occurrence in the data (the number of times it appears). TF-IDF was also implemented at this stage, in order to vectorize the text.

Model Implementation. In order to identify the linguistic indicators of generative interactions, we used a machine-learning approach. We implemented several machine learning models, including Random Forest, AdaBoost (Adaptive Boosting), Naïve Bayes (Multinomial), Support-Vector Machine (SVM), and Logistic Regression, to find the one that was best suited for classifying the data as generative or non-generative. Using performance measures such as f-1 score, accuracy, and Area Under the Curve, we were able to compare the models implemented and identify the best performing one. Once we identified the best performing model, we were able to use it to extract the top 20 important words for distinguishing generative activity.

3.3 Results

The results of the models we implemented can be seen in Table 2. Due to the contrast in performance, we can conclude that Random Forest was the best performing model with a 76% accuracy score, a score of 80% for AUC, and 83% for the f-1 score. These are satisfactory results for a ~1% sub-sample. Adaptive Boosting (AdaBoost) was the second-best performing model, with 71% accuracy, but lower AUC and f-1 scores. The worst performing model was Naïve Bayes with 44% accuracy, 59% AUC score, and 53% f-1 score.

Table 2. Model Performance: All Measures.

Model	AUC	Accuracy	f-1
Random Forest	0.80	0.76	0.83
Adaptive Boosting	0.70	0.71	0.81
Naïve Bayes	0.59	0.44	0.53
SVM	0.67	0.69	0.78
Logistic Regression	0.72	0.66	0.72

Table 3. Model Performance: All Measures.

Model	f-1	
	0	1
Random Forest	0.90	0.67
Adaptive Boosting	0.88	0.67
Naïve Bayes	0.55	0.49
SVM	0.83	0.64
Logistic Regression	0.76	0.61

In more detail, the f-1 score (seen in Table 3) for the two categories displays the performance of the models at correctly classifying either one. At a more granular level, Random Forest still seems to be the best performing model as it was correct 90% of the time at classifying the instances of non-generative text and correct 67% of the time at classifying generative content. In contrast, the Naïve Bayes model was correct 49% of the time at classifying non-generative content and correct 55% of the time at correctly classifying generative content. Due to the results above, we used the Random Forest model to produce the top 20 important features in the data, which are the linguistic indicators that help us identify instances of generative interactions. These terms are significant for the machine-learning model; they aid with distinguishing the generative and non-generative activity indicators in the text data.

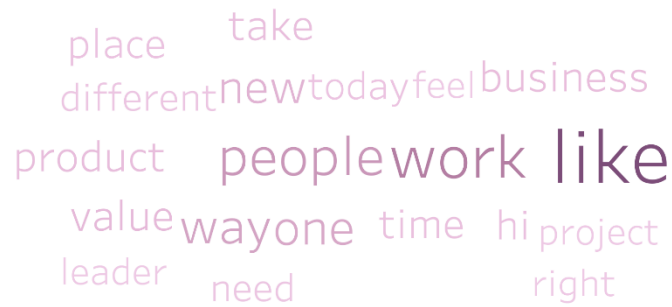
**Fig. 1.** Word cloud with top 20 important terms.

Table 4. Top 20 Important Features.

Term	Score
like	0.0601
work	0.0403
people	0.0313
way	0.0268
one	0.0254
new	0.0214
value	0.0194
product	0.0183
business	0.0181
take	0.0179
time	0.0172
hi	0.0171
place	0.0167
today	0.0162
different	0.0158
need	0.0158
feel	0.0144
right	0.0144
leader	0.0144
project	0.0143

Table 5. Sample Generative and Non-Generative Interactions.

Type	%	Sample
Generative Activity	28%	“Two factors which I feel either hinder or help engagement, are ‘autonomy’ which Rob & Bob mentioned, and ‘change’ or ‘impact’. There is a management concept of ‘leading’ with a light touch’, people want to understand the limits, the outside boundaries of the work they are asked to do.”
Non-Generative Activity	72%	“The [organization name] interns had the opportunity to participate in Chicago yesterday. It was great to see the [organization name] show so full and have such an exciting buzz around it.”

4 Discussion

Terms such as ‘work’, ‘business’, ‘product’, ‘project’, and others, are essential linguistic indicators of generative interactions. These indicators are important in distinguishing team exchanges that involve generativity from those that do not. Our findings showed that 28% of the interactions in the data were generative, while 72% were non-generative content, indicating that indeed ESM is a source of generative interactions.

Though our preliminary study used a small portion of the data corpus available, thereby allowing us to only differentiate generative versus non-generative interactions, it shows promise of using machine learning to reliably discern not only when team exchanges in ESM are generative in nature—and thus identify potential root-causes of breakthrough innovations—but also possibly in distinguishing between the different types of generative interactions, namely combination, expansion, and reframing.

Being able to identify the linguistic indicators of distinct types of generative interactions would allow us to not only theorize the nature of generative interactions occurring through ESM, but also develop theoretical models of the precursors that result in distinct types of ESM-based generative interactions. For instance, the ways in which groups interact with each other and with the ESM in the context of these interactions might be different when groups are engaged in combination, expansion, or reframing. Such insights are theoretically important to obtain holistic understandings of the boundary conditions for different types of generative interactions as well as practically important to provide managers guidance for eliciting different types of generative interactions in an attempt to encourage productive uses of ESM. Hereto, more data will have to be labelled, and further experimentation with machine learning algorithms will be needed to produce an accurate classifier for multiple categories of generative interactions.

References

1. Abernathy, W. J., and Clark, K. B. 1985. Innovation: Mapping the winds of creative destruction. *Research policy*, 14(1), 3-22.
2. Avital, M., and Te'eni, D. 2009. From generative fit to generative capacity: exploring an emerging dimension of information systems design and task performance. *Information Systems Journal* 19(4), 345-367.
3. Burke, M., Marlow, C., and Lento, T. 2009. Feed me: Motivating newcomer contribution in social network sites. *CHI 2009: Proceedings of the 27th International Conference on Human Factors in Computing Systems*, 945-954.
4. Beck, R., Pahlke, I., and Seebach, C. 2014. Knowledge exchange and symbolic action in social media-enabled electronic networks of practice: A multilevel perspective on knowledge seekers and contributors. *MIS quarterly*, 38(4), 1245-1270.
5. Dunbar, K. 1997. How scientists think: On-line creativity and conceptual change in science. T. N. Ward, S. M. Smith, J. Vaid, eds. *Creative Thought*. American Psychological Association, Washington, DC, 461-494.
6. Erikson, E.H. 1950. *Childhood and Society*. New York: W.W. Norton and Company
7. Hambrick, D. C. 1983. Some tests of the effectiveness and functional attributes of Miles and Snow's strategic types. *Academy of Management journal*, 26(1), 5-26.
8. Henderson, R. M., and Clark, K. B. 1990. Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative science quarterly*, 9- 30.
9. Harvey, S. 2014. Creative synthesis: Exploring the process of extraordinary group creativity. *Academy of Management Review* 39(3), 324-343.
10. Kane, G. C. 2017. The evolutionary implications of social media for organizational knowledge management. *Information and organization*, 27(1), 37-46.
11. Leonardi, P. M., Huysman, M., and Steinfield, C. 2013. Enterprise social media: Definition, history, and prospects for the study of social technologies in organizations. *Journal of Computer Mediated Communication* (19:1), 1-19.
12. Leonardi, P.M. 2014. Social media, knowledge sharing, and innovation: Toward a theory of communication visibility. *Information Systems Research* 25(4), 796-816.
13. Leonardi, P M., and Vaast, E. 2017. "Social Media and Their Affordances for Organizing: A Review and Agenda for Future Research," *Academy of Management Annals* (11:1), pp. 150-188.
14. Lieberman, M.B., and Montgomery, D.B. 1988. First- Mover Advantages. *Strategic Management Journal* 9(1), 41-58.
15. Tsoukas, H. 2009. A dialogical approach to the creation of new knowledge in organizations. *Organization Science* 20(6), 941-957.
16. Van Osch, W., and Avital, M. 2010. Generative Collectives. *ICIS 2010 Proceedings*, http://aisel.aisnet.org/icis2010_submissions/175.
17. Van Osch, W. 2012. *Generative Collectives*. Ipskamp Publishers, Netherlands.
18. Van Osch, W., and Steinfield, C.W. 2013. Boundary Spanning through Enterprise Social Software: An External Stakeholder Perspective. *Proceedings of the International Conference on Information Systems (ICIS)*, Milan, Italy.
19. Van Osch, W. 2015. The Business Side of Social Media. *International Innovation*, 195, 27-29.
20. Van Osch, W., and Steinfield, C. W. (2018). Strategic visibility in enterprise social media: Implications for network formation and boundary spanning. *Journal of Management Information Systems*, 35(2), 647-682.

21. Webster-M. 2009. Generativity. In Merriam-Webster Online Dictionary.
22. Wisniewski, E. J. 1997. When concepts combine. *Psychonomic Bulletin and Review*, 4(2), 167-183.

ACKNOWLEDGEMENT

This material is based upon work supported by the National Science Foundation under Grant No. 1749018. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.