Paper:
Take their word for it: The symbolic role of the Linguistic Style Matches in User Communities

1) How can the quantity and the quality of an user be related to an LSM in a community?

2) They developed hypothesis for multiple categories:

   INDIVIDUAL LSM (degree to which a member LSM the overall community style should reflect his/her level of identification with the group and influence his/her participation).

   H1: *A member's LSM is positively related to subsequent (a) participation quantity and (b) participation quality.*

   LSM TRENDS (degree of convergence toward or divergence from a collective style)**:**

   H2: The greater the rate of convergence (divergence) in member's LSM, the higher (lower) their subsequent (a) participation quantity and (participation) quality.

   LSM REVERSALS (members with a high degree of reversals in LSM development should be less motivated to continue to provide high-quality argumentation and participate.

   H3: Increasing amounts of reversals in member LSM relate negatively to (a) participation quantity and (b) participation quality.

   GROUP-LEVEL LSM (Users communities exhibing hogh level of LSM across members should reflect a more prominent collective identity, which affects member participation effort)

   H4: Community-level cohesiveness in members LSM relates positively to individual members (a) participation quantity and (b) participation quality.

   METHOD

   Setting

   Study sample: 37 similar structured user communities

   All the users were hosted by the same market research consultancy in different industries (finance & assurance, information services…)

   4 reasons to that:

   -Members community participation is due to motivation (topics they like…)
   -To motivate and sustain participation it's crucial for members to develop a sense of community identification
   -As with most online communities, nonverbal social cues and personal member information are not available, with text-based posts serving as the sole means by which to develop and assess social identification.
   -Homogeniety across user communities (setup,duration,structure and purpose supports)

   Final sample: 2208 members across 37 communities with a total of 74 246 posts.

   Data & Measures

   They divided the observation period into an:

   - "Initiation period" (T1) => member's first two weeks in a community.

   - "Active participation period" => eight week of membership (T2)

Dependent Variables

Participation Quantity:

$$PQuant_i^{T_2} = \sum_{t=1}^{8} Posts_{it}^{T_2} \qquad (1)$$

Absolute count of posts by each individual community member summed over the eight weeks in the active participation T2.

- *i* => individual community member
- *t* => week in the eight-week time period after the initial two-week initiation period (T2).

Cognitive effort:

$$CE_{ip}^{T_2} = CognitiveWords_{ip}^{T_2} + CausalWords_{ip}^{T_2} \qquad (2)$$

They constructed a composite measure of the text-mined cognitive effort (CE) for each individual (i) by post (p) by summing the total amount of cognitive words and causal words used in a post where T2 indicates the time period (two week initiation period)

Participation Quality:

$$PQual_i^{T_2} = \sum_p \frac{CE_{ip}^{T_2}}{PQuant_i^{T_2}} \qquad (3)$$

- i => Individual
- p => Post
- T2=> time period after
- CE Cognitive effort per post

Independent Variables:

They assess the independent variables in this study according to communications behavior in the first two weeks (T1).

Function Word Usage Intensity (FWC):

$$FWC_{ij}^{T_1} = \frac{\sum_p FWC_{ijp}^{T_1}}{\sum_p TotalWords_{ip}^{T_1}} \qquad (4)$$

They constructed a measure of the FWC for each community member (i) and for each function category (j) by dividing the number of words belonging to the particular function word category across all posts by the total number of words per the post(p) across all posts.

$$FWC_{cj}^{T_1} = \sum_{j \in c} \frac{FWC_{ij}^{T_1}}{PQuant_i^{T_1}} \qquad (5)$$

They determined the specific community level function word category usage intensity (FWC) for each community © and for each category (j) across all the posts that were present before the current post was submitted.

LSM (Linguistic Style Match):

$$LSM_{ij} = 1 - \frac{\left(|FWC_{ij}^{T1} - FWC_{cj}^{T1}|\right)}{\left(|FWC_{ij}^{T1} + FWC_{cj}^{T1} + 0.0001|\right)} \qquad (6)$$

LSM => ratio of overlap between the usage intensity by each individual member (i) for each function category (j) and the cumulative average usage intensity of the same function word category (j) by all community posts that were posted prior to the current post in the community (c). They added 0.0001 to prevent empty sets.

They used 2 levels to describe the dynamics of communicative behaviors:

-Member-level aspects of communicative behavior (average, trend and reversal).
-Overall community-level cohesiveness in communicative behavior.

First LSM ratio is bounded by 0 and 1 for each of the 2261 members

Second a member's LSM trend is establishing by regressing the sequential post incidences by member (i) on the respective LSM of each of his/her post using the least squares method. They the regressed the LSM of member i on the posts incidents of member i. The beta coefficient ($\beta 1$) of the post incidence variable in the equation LSMi=β0+β1.PostIncidents+εi , for members i signifies the rate of change in his/her LSM trend over time.
A rate of change near zero represents a stable LSM trend throughout the initial period of person's community membership.

Third they measure the frequency of change in a member's LSM across subsequent posts by counting the number of slope changes (that exceed one standard deviation) compared with the number of posts.

Fourth to establish community level cohesiveness in communicative style (indicate the degree of equality in communication styles within a community) they used the coefficient of variation adjusted for group size. They constructed a measure of cohesiveness in communication style within communities as one minus the within-group variability of LSM.

To justify the multilevel perspective and validate the aggregation procedure. They calculated the the within-group agreement r wg for single items measures. Because this measure is designed for scales variables only, they discretionized the original ratio measure of LSM into a scale consisting of 10 categories and estimated the rectangular distribution.

They see after that the mean of the r wg coefficients is 0.87. These findings demonstrate that the common communication style of individual LSM within communities is highly consistent.

Table 1. Sample Conversations in a User Community: Recipe Discussions		
Member	**Text**	**LSM of Member**
User 1:	Hi, my family love cake - particularly a nice light victoria sponge. I use …, it makes the cake really moist. My Mum taught me this when I was a nipper.	0.69
User 2:	Well me and my hubby used to go eat in … I kept thinking i used to make that many years ago, so i set about it didn't tell hubby. Just put at front of him, well he loved it, said it is better than … its good but fattening	0.71
User 3:	Lost weight at … and devised a lot of convenient low fat / fat free recipes when I was on the plan. The ideal way to make … is to use …, season appropriately, no fat other than the substances in the milk.	0.32
User 2:	Thanks … I am going to look out for some of that - I do like … but I have to lose quite some weight so this is just the job! Have you discovered … It really does fill me up honest and it tastes wonderful too!	0.57
	…	
User 4:	I LOVE … Hollondaise sauce but you can't buy it in England. I stocked up big time when I was working in the US but it's all gone now - all gobble up and my family are suffering withdrawl!! …. Now I need to make my mock hollandaise every time we have cauliflower with our dinner!	0.73
User 5:	Hi … if you like Mock Hollandaise sauce for Cauliflower without the … this one may be for you. Just use any White sauce recipe and add the nutmeg, lemon juice and egg yolk this makes all the difference and done really quickly.	0.62
User 3:	Talking about …, I also like the range of …. I use the ITALIAN GARLIC one all the time, so easy to shake and add … using this one mixed up with butter is very easy to do and is ready in a flash.	0.56
	…	
User 3:	When roasting a joint of beef use …. Learned this years ago working as a waitress at a night club that had a top notch restaurant. The chef was Croatian and his food was devine!	0.34
User 6:	Hi everyone, for an interesting twist I have a tip for you, use peanut butter as a topping on a beefburger. It sounds odd but tastes really good! I saw this on … and thought I'd give it a try and have used up many jars of peanut butter since!	0.72

They focused on user 3.Her LSM scores are .12, .56, .34 across her 3 posts, compared o the cumulative style of all the community.

We can see that the second post matched the community linguistic style best in terms of functions words. There is a trend toward a greater LSM from her first to her second post, but this trend is reversed with her third post.

Control Variables

Participation Quantity (PQuant)
Participation Quality (PQual)
Number of pages views within the community (PViews)
Timestamp (two weeks of membership T1).

Participation Quality is a count measure they needed to control the average word length (PLength) of a member's post.

Community Size (CSize) => Total amount of members.
Community Participation Quantity => (CPQuant)
Community Participation Quality => (CPQual)

Table 2. Descriptive Statistics and Correlations[a]															
	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13
Member-Level Variables															
1. PQuant^{T2}	14.79	23.86	1.00												
2. PQual^{T2}	1.84	0.05	0.02	1.00											
3. LSMM^{T1}	0.63	0.17	0.35	0.25	1.00										
4. LSMT^{T1}	-0.02	0.09	0.12	0.09	0.01	1.00									
5. LSMR^{T1}	0.27	0.28	-0.16	-0.14	-0.15	-0.03	1.00								
6. PQuant^{T1}	10.19	14.66	0.47	0.43	0.10	0.09	-0.14	1.00							
7. PQual^{T1}	0.05	0.05	-0.03	0.08	0.18	0.01	-0.24	0.01	1.00						
8. PView^{T1}	615.00	223.39	0.26	0.20	-0.02	0.01	-0.03	-0.03	-0.05	1.00					
9. PLength^{T1}	41.99	10.48	-0.06	0.29	0.29	-0.01	-0.29	-0.05	0.20	0.03	1.00				
Community-Level Variables															
10. CLSM^{T1}	0.91	0.15	0.14	0.27	0.17	0.02	0.02	0.18	0.13	-0.03	-0.01	1.00			
11. CSize	178	33.43	0.12	0.11	-0.02	-0.08	0.10	-0.06	-0.06	-0.01	-0.08	0.19	1.00		
12. CPQuant^{T1}	9.03	6.44	0.03	0.01	0.13	-0.01	0.13	0.42	-0.01	0.03	-0.01	0.15	-0.09	1.00	
13. CPQual^{T1}	0.05	0.04	-0.12	0.62	0.01	-0.02	-0.03	-0.07	0.54	-0.01	0.21	0.14	-0.24	-0.01	1.00

Prestudy

They conducted a prestudy to assess the validity of the LSM empirically as a symbol for social identification.

They developed four community measure identification:

- “I am very attached to this community”
- “Other members and I share the same objectives”
- “I see myself as part of the community”
- “The friendship I have with others community members means a lot to me”

They sent an online survey twice (including one follow up survey) by email to all members of a subset of 18 communities 8 to 12 weeks after their had begun their community participation.

Of the 3211 members, 622 completed the survey (response rate = 19,37%)

The result showed a significant correlation between members LSM ratio and their community identification (0.71, $p<.01$). Thanks to this result they can see that the degree of members LSM is a significant positive symbol of their community identification.

Data Analysis

To capture the influence of the explanatory variables they specified two multilevels poisson models also often referred to as hierarchical linear models (HLMs).
With multiple members nested in each user community the HLM modeling approach also controls appropriately for the possibility that communication behaviors from members in the same community may be more similar than they are for members in another community.

Regarding participation quantity the estimated median IRR is 1.45 which implies that the expected participation will range to 0.68 to 1.45.

Similar for participation quality the estimated median is 1.38 which implies that the expectation should lie between 0.73 and 1.38.

This finding provides convincing evidence that community characteristics can have a direct influence on members participation quantity and argumentations quality.

$$PQuant_{ic}^{T_2} = \exp(\beta_{0c} + \beta_1 \cdot LSMM_{ic}^{T_1} + \beta_2 \cdot LSMT_{ic}^{T_1} + \beta_3 \cdot LSMR_{ic}^{T_1} + \beta_4 \cdot LSMC_c^{T_1} + \beta_5 \cdot PQuant_{ic}^{T_1} + \beta_6 \cdot PQual_{ic}^{T_1} + \beta_7 \cdot PViews_{ic}^{T_1} + \beta_8 \cdot PLength_{ic}^{T_1} + \beta_9 \cdot CSize_c + \beta_{10} \cdot CPQuant_c^{T_1} + \beta_{11} \cdot CPQual_c^{T_1} + \zeta_{0c}) \quad (7)$$

and participation quality (Model 2b)

$$PQual_{ic}^{T_2} = \exp(\beta_{0c} + \beta_1 \cdot LSMM_{ic}^{T_1} + \beta_2 \cdot LSMR_{ic}^{T_1} + \beta_4 \cdot LSMC_c^{T_1} + \beta_5 \cdot PQuant_{ic}^{T_1} + \beta_6 \cdot PQual_{ic}^{T_1} + \beta_7 \cdot PViews_{ic}^{T_1} + \beta_8 \cdot PLength_{ic}^{T_1} + \beta_9 \cdot CSize_c + \beta_{10} \cdot CPQuant_c^{T_1} + \beta_{11} \cdot CPQual_c^{T_1} + \zeta_{0c}) \quad (8)$$

Where:

- i => individual member
- c =>community
- PQuant => Participation Quantity
- PQual => Participation Quality
- Timestamp=> T2
- LSMM => average degree of LSM
- LSMT => rate change in LSM trend
- LSMR => frequency of reversals in LSM in a member's post in period T1
- LSMC => group-level cohesiveness in members LSM at the community level at period T1

Table 3. Multilevel Poisson Regression Analysis

	Participation Quantity^{T2}			Participation Quality^{T2}		
Constructs	**Model 1a**	**Model 2a**	**IRR (95% CI)[a]**	**Model 1b**	**Model 2b**	**IRR (95% CI)[a]**
Member-Level Variables						
LSMM^{T1}	0.71**	0.23**	1.26 (1.23–1.30)	0.63**	0.33**	1.39 (1.29–1.50)
LSMT^{T1}	0.09**	0.06**	1.06 (1.05–1.08)	0.20**	0.17**	1.19 (1.14–1.24)
LSMR^{T1}	–0.04**	–0.07**	0.93 (0.97–0.95)	–0.02	–0.01	0.98 (0.94–1.03)
PQuant^{T1}	0.35**	0.31**	1.36 (1.35–1.37)	0.02	–0.03	0.97 (0.93–1.01)
PQual^{T1}	0.08**	0.07**	1.07 (1.04–1.10)	0.27**	0.23**	1.26 (1.19–1.34)
PView^{T1}	0.12**	0.11**	1.12 (1.11–1.13)	–0.02	–0.03	0.96 (0.92–1.01)
PLength^{T1}	–0.02**	–0.01**	0.99 (0.98–0.99)	0.01**	0.02**	1.02 (1.01–1.02)
Community-Level Variables						
CLSM^{T1}		0.90**	2.48 (2.39–2.56)		0.72**	2.03 (1.84–2.29)
CSize		–0.20	0.81 (0.63–1.06)		–0.27**	0.74 (0.62–0.94)
CPQuant^{T1}		–0.19	0.83 (0.67–1.01)		–0.10	0.91 (0.77–1.07)
CPQual^{T1}		0.11	1.05 (0.84–1.32)		0.36**	1.74 (1.46–2.11)
Intercept	3.05**	2.54**		–0.17	–0.38**	
N (members)	2208	2208		2208	2208	
N (communities)	37	37		37	37	
Log likelihood	–13498.30	–12104.40		–3042.21	–2932.29	
Wald χ^2 (df)	19634.35(7)	21486.63** (11)		1231.19 (7)	1376.78 (11)	
Deviance(–2LL)[b]	26996.59	24208.79		6084.43	5864.58	

Results

Table 4. Hypotheses Testing				
Hypothesis	**Relationships[a]**	**Direction**	**ß**	**Result**
H_{1a}	LSM Mean$_{(t1)}$ → Participation Quantity$_{(t2)}$	Positive	.23**	Supported
H_{1b}	LSM Mean$_{(t1)}$ → Participation Quality$_{(t2)}$	Positive	.33**	Supported
H_{2a}	LSM Trend$_{(t1)}$ → Participation Quantity$_{(t2)}$	Positive	.09**	Supported
H_{2b}	LSM Trend$_{(t1)}$ → Participation Quality$_{(t2)}$	Positive	.17**	Supported
H_{3a}	LSM Reversals$_{(t1)}$ → Participation Quantity$_{(t2)}$	Negative	-.04**	Supported
H_{3b}	LSM Reversals$_{(t1)}$ → Participation Quality$_{(t2)}$	Negative	n.s.	Not Supported
H_{4a}	LSM Cohesiveness$_{(t1)}$ → Participation Quantity$_{(t2)}$	Positive	.90**	Supported
H_{4b}	LSM Cohesiveness$_{(t1)}$ → Participation Quality$_{(t2)}$	Positive	.72**	Supported
Fit Measures	**Outcome Construct**	**Model**		**N(members)**
LL	Participation Quantity	–12104.40		2208
	Participation Quality	–2932.29		2208

Almost all hypothesis are supported expected that the 3b hypothesis is not supported (Beta unknown).

We can know thanks to the different type of LSM (Mean,Trend,Reversals,Cohesiveness) see how Quantity and the Quality are linked (Beta).

Contributions

this study contributes to contemporary research on the role of communication in user communities by showing that the degree of members' LSM symbolically reflects their level of identification in online groups

Extend research on CAT to a multilateral setting by demonstrating its multilevel implications for members' participation.

The LSM's focus on function words enriches the conventional view on text analysis, which assumes that the semantics of nouns and verbs are key in understanding text. Instead, our study identifies the use of function words as a subtle, implicit way of rendering other symbols as meaningful and interpretable for community participants.

The current study extends CAT by adopting a temporal perspective, viewing LSM trends and reversals as essential, distinct symbols in the ongoing socialization process through which user community members produce, reproduce, and change their community identification.

The current study shows that members who often alter their degree of LSM are less likely to excel in subsequent participation quantity. Yet contrary to the hypothesized relationship, reversals in members' LSM are not significantly negatively related to their subsequent participation quality.

They demonstrate empirically the substantive symbolic nature of two temporal parameters (trend and reversals) in CAT, while controlling for behavioral (e.g., passive reading behavior) and contextual (e.g., community quality, size) aspects.
Third, we contribute to research on communication by considering text not only as a data source but also as the societal context in which the text is nested (Fairclough 1992).
We find that synchronicity in communicative behavior—or cohesiveness across community members' linguistic styles—adds substantively to the explanation of individual members' participation behavior.

the current study offers additional insights on user community participation. Specifically, whereas previous research has stressed the criticality of frequent participation and the generation of good quality content (Ransbotham and Kane 2011), drawing on research on argumentation quality, we highlight the importance of argument development quality in group communication processes.

we view participation quality as the degree to which members substantiate and develop their statements and arguments, which enables a better analysis of participation quality across communities that vary in their content.

Critiques

This paper shows us how the Linguistic Style Match is coming from and also how the different LSM (trend…) impact the Participation Quality and Quantity in different community.