

# Pragmatic Alignment on Social Support Type in Health Forum Conversations

Yafei Wang, John Yen, David Reitter

College of Information Sciences and Technology

The Pennsylvania State University

University Park, PA 16801, USA

yxw184@psu.edu, jyen@ist.psu.edu, reitter@psu.edu

## Abstract

Linguistic alignment, such as lexical and syntactic alignment, is a universal phenomenon influencing dialogue participants in online conversations. While adaptation can occur at lexical, syntactic and pragmatic levels, relationships between alignments at multiple levels are neither theoretically nor empirically well understood. In this study, we find that community members show pragmatic alignment on social support type, distinguishing emotional and informational support, both of which provide benefits to members. We also find that lexical alignment is correlated with emotional support. This finding can contribute to our understanding of the linguistic signature of different types of support as well as the theory of Interactive Alignment in dialogue.

## 1 Introduction

Linguistic alignment is a psycholinguistic phenomenon that causes dialogue participants to adjust their language patterns to those of their conversation partners. These linguistic patterns include words (Gries, 2005), syntax (Bock, 1986; Branigan et al., 2000; Jaeger and Snider, 2007), gestures (Bergmann and Kopp, 2012) and more. This phenomenon has been well examined and explored under experimental settings (Fusaroli et al., 2012; Reitter and Moore, 2007), in naturalistic discourse (Gries, 2005; Reitter et al., 2006), as well as in online conversations (Huffaker et al., 2006; Scissors et al., 2008; Backstrom et al., 2013), social media, and fictional dialogue in

film scripts (Danescu-Niculescu-Mizil and Lee, 2011). Moreover, according to the Interactive Alignment Model (IAM) (Pickering and Garrod, 2004), linguistic alignment has been suspected to be a driver of mutual understanding, building up over different levels (lexicon, syntax, situation, pragmatic, agreement).

Alignment is a universal phenomenon that reaches beyond the linguistic decisions we make once we have decided to communicate an idea. *Pragmatics* is commonly taken to refer to the way we express and understand communications in context, encoding higher-level intent. How people understand words and phrases in a given situation is indeed subject to alignment (Garrod and Anderson, 1987). Generally, games can elucidate pragmatic reasoning and mutual adaptation thereof (Frank and Goodman, 2012). To the best of our knowledge, there is no prior report of pragmatic alignment in naturalistic situations. For the purposes of this study, we define the pragmatics of an utterance as the intended effect on the reader or listener, regardless of the way it is semantically expressed. Unlike in pragmatics in linguistics, however, our focus is not on the differences between explicitly stated and implied meaning.

The first question this paper will focus on is analyzing pragmatic alignment in naturalistic dialogue, specifically in internet forum conversation. To understand what we mean by higher-level semantics or pragmatics in these data, we need to understand the motivation and dynamics of these communities.

An increasing number of people with serious

disease seek and give social support in group discussions in online social networks such as Facebook and online health support communities. Basically, there are four types of social support, emotional support, informational support, tangible/instrumental support and appraisal support (Langford et al., 1997; Malecki and Demaray, 2003). In online communities, the social support exchange is primarily of an informational or emotional nature (Wang et al., 2012; Rimer et al., 2005). Understandably, people with a life-threatening illness are in need of both information, such as side-effects of a specific drug, and emotional care, such as empathy. Previous research on behavior analysis, such as stress-buffering theory (Cohen and McKay, 1984), also suggested that exchanging useful social support protects people from stressful and pathological events. Analyzing the social support and the kind of support conveyed in the messages will be of benefit to support-oriented community building. Furthermore, previous studies (Zhao et al., 2014) suggested that earlier responses to a new support seeking request help predict leaders in self-supported communities. Although, the proportion of emotional or informational support in a message can, of course, be influenced by many factors, such as previous messages in the conversation, word choices and personality. Nevertheless, we use this measure for further analysis. From the alignment perspective, we will focus on whether people tend to align in the type of support in online health communities. In other words, we first analyze the pragmatic alignment phenomenon, which is defined as alignment of the type of support provided by one community member to another. We validate it in one of the largest online health communities, Cancer Survivor Network. To the best of our knowledge, pragmatic alignment in online communities has not been explored yet.

The second question is whether we could find evidence for or against the Interactive Alignment Model (Pickering and Garrod, 2004) in this dataset. As IAM suggested, alignment at different levels is linked, building up from lower-level adaptation. At a functional level, linguistic alignment indicates and may help build social relationships (Danescu-Niculescu-Mizil and Lee, 2011), reveal

social status (Danescu-Niculescu-Mizil et al., 2011; Jones et al., 2014) and strengthen situational awareness in dialogic tasks (Fusaroli et al., 2012; Reitter and Moore, 2007, 2014).

Thus, an important question in this context is whether adaptation also applies to higher-level pragmatic goals, such as providing support that is more informational or more emotional. Convergence at lower levels would theoretically be expected to correlate to higher-level convergence, and conversations that show convergence would be expected to be more effective. Do priming effects at levels of lexicon and syntax influence the proportion of the type of support in a message within the conversation? We predict that social support adaptation exists in thread based discussions. Theoretically, we would also expect that low-level priming facilitates any social support adaptation we find.

To sum up, there are two concrete questions we will address in this paper:

- (1) Does the type of support (i.e, emotional vs. informational) provided by early responders (i.e, first responder) on a thread influence the type of support provided by later responders in self-support communities?
- (2) Does lexical and syntactic alignment (henceforth “linguistic alignment”) between early responders and later responders correlate to the type of support matching?

The alignment we are concerned with would clearly happen at the level of communicative intent. We consider this *pragmatics*. The pragmatics we refer to is not the same as it’s used in linguistics concerning contextual and indirect interpretation of sentence semantics, but rather the sense of intent, in a psychological sense. In psychology, pragmatic communication comprises social and conventional messages that take the recipient’s needs into account. Social support adaptation specifically considers the unspoken rule that we perceive an interlocutor’s emotional and informational needs and react accordingly.

Some studies in behavior analysis (Backstrom et al., 2013; Cheng et al., 2014) showed that word use in the conversations may influence members’

behavior in the communities. Althoff et al. (2014) stated that request presentation influences members’ feedbacks in a variety of ways, such as sentiment, politeness and length of reply posts. Cheng et al. (2014) mentioned that members’ feedbacks also shapes users’ behavior in the communities. Furthermore, automated content (Qiu et al., 2011) and discourse analysis using machine learning methods provided important insights about the benefits and causal relationships (Bui et al., 2015) with support behaviors in online health communities. Thus, modeling members’ feedbacks at the pragmatic level could help us build better communities.

A recent study from Vlahovic et al. (2014) was similar to our study. They used profit regression to predict members’ satisfaction after receiving emotional and informational support in a breast cancer online support community. For one thread, a trained profit regression model predicted the thread initiators’ satisfaction scale from 1 to 7. In this study, both receiving emotional and informational support increased thread initiators’ satisfaction in general. However, if a thread initiator received support that did not match the type requested, this user’s satisfaction decreased. In this work, we focus on whether previous messages will influence other responders’ behavior in the ensuing conversation.

## 2 Measures

In this paper, we use *adaptation* measures at two levels, *linguistic* alignment and *pragmatic* similarity. Linguistic alignment quantifies by how much conversation participants adapt their language patterns to those of their interlocutors. Studies differ in the kinds of patterns examined: Some approaches measure linguistic adaptation using Linguistic Inquiry and Word Count (LIWC) (Tausczik and Pennebaker, 2010; Danescu-Niculescu-Mizil et al., 2011), and some focus on functional words (Jones et al., 2014). Other approaches measure repetition of words or syntactic rules (Church, 2000; Dubey et al., 2005; Fusaroli et al., 2012; Gries, 2005; Reitter et al., 2006). We use Indiscriminate Local Linguistic Alignment (Fusaroli et al., 2012) to measure linguistic alignment in this paper. Pragmatic similarity, for the purposes of the present

study, evaluates the degree of matching social support types in conversation messages. In the following, we will introduce these measures.

### 2.1 Linguistic Alignment Measures

In this paper, we implement *Indiscriminate Local Linguistic Alignment* (Fusaroli et al., 2012) at lexical and syntactic levels to evaluate linguistic alignment. Generally, it measures the repetition of linguistic patterns among messages in the same conversation.

To be specific, *Lexical Indiscriminate Local Linguistic Alignment (LILLA)* measures word repetition between between pairs of messages (Wang et al., 2014; Fusaroli et al., 2012). The messages, ordered by occurrence in a thread of messages, will be called *prime post* and *target post*, respectively. In this study, they will sampled from the Cancer Support Network corpus. Formally, LILLA is calculated as

$$LILLA(target, prime) = \frac{\sum_{word_i \in target} \delta(word_i)}{length(prime) * length(target)} \quad (1)$$

$$\delta(word_i) = \begin{cases} 1 & \text{if } word_i \in \text{prime} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where  $length(X)$  is the number of words in post  $X$ .

We also measure syntactic alignment. Every sentence in each post is annotated with phrase structure trees using the Stanford CoreNLP parser (Klein and Manning, 2003). Each syntax tree is translated to a series of syntactic rules to encode the sequence of syntactic decisions. *Syntactic Indiscriminate Local Linguistic Alignment (SILLA)* is analogous to LILLA and measures repetition of syntactic rules between prime and target post pair, where  $length(X)$  in SILLA is the number of rules in post  $X$ . (Fusaroli et al., 2012; Wang et al., 2014).

### 2.2 Support Measures

As discussed above, emotional support and informational support are two most major support types in self support health communities (Wang et al., 2012; Rimer et al., 2005). Emotional support gives individual a feeling that s/he is cared for, or the facility of “*understanding/empathy, encouragement, affirmation/validation, sympathy,*

and caring/concern” (Bambina, 2007). Emotional support does not include information. Here is an example of emotional support in CSN: “*I pray for you every night XXX.....and send you hugs and encouragement....you have the very BEST attitude and you must have a totally wonderful family. Love, XX*”

However, different from emotional support, informational support provides facts, advices and referrals (Bambina, 2007). Also, in our case, informational support only provides experience and information, without any emotional support. Another example of informational support in our data is: “*I am having similar problem with sacrum and hip, however not ready for biopsy in those areas. If you can tolerate pain waiting for new drugs to come will be beneficial. a new drug palbociclib (PD-0332991) expected to receive FDA approval in April of 2015*”.

In order to quantify the amount of one type of support in a reply post, we quantify the amount of one type of support (i.e. informational support, emotional support) in a comment post as *support index* Biyani et al. (2014), as follows:  $Index_{type} = num_{type}/num_{classified}$ , which is the proportion of sentences of a specific type in a post.

We will predict the emotional support index,  $Index_{emo} = 1 - Index_{info}$  (for presentational reasons). The measure is produced automatically using the previously published classifier (Biyani et al., 2014).

### 3 Data Description

The data we use in this paper is from Cancer Survivor’s Network (CSN) ([csn.cancer.org](http://csn.cancer.org)), which is the largest active online community for cancer survivors. The CSN contains more than 166,000 users and 41 sub-communities (Portier et al., 2013). Users in one sub-community have experienced the same primary disease, similar health issues, surgeries. Furthermore, many users express depression. Most of the discussions in CSN are goal-directed and support-oriented conversations, which attracted our attention. Users would like to exchange their experiences and emotions in facing these tough situations.

We used threads from two largest sub-forums

in the CSN: Breast cancer and Colorectal cancer. These sub-forums contain posts from the period of June 2000 to October 2010. The majority of posts in the breast cancer sub-forum are from female members, while most posts in the colorectal cancer sub-forum were authored by male patients. Thus, the two corpora are from relatively distinct, but representative user groups.

Mirroring the structure of other online communities, we refer to an initial post followed by a sequence of reply post as a *thread*. We treat the structure of these threads as a sequence of plain texts in temporal order, as members often use a general “reply” button to initiate replies, even when such messages are direct replies to a post. Thus, more detailed post relationships within each thread are sparse and not very reliable. A discussion thread is represented as a sequence of posts,  $\langle P_0, P_1, \dots, P_i, \dots, P_n \rangle$ , where  $P_0$  is called *initial post*,  $P_1$  is called the *first reply of one thread* (simply called *first reply*) and the author of initial post is called the *thread initiator*. In most cases, the rest of replies provide help and emotional support to the thread initiator. The variable  $i$  is called the *absolute position* of post  $P_i$ . Posts in which the thread initiator replies to his or her message are excluded (as thread initiator may not provide support to themselves). The number of replies in a thread (without the initial post) is called the *length* of that thread. Both sub-communities have similar distributions of thread length. 90% of threads in Breast Cancer forum and Colorectal Cancer forum are shorter than 23 and 19, respectively.

We used a binary sentence classifier described by Biyani et al. (2014), which classifies sentences as providing either emotional or informational support. The classifier was trained in that work on more than 1,000 hand-annotated sentences; annotators reached 89% initial agreement. The classifier uses a variety of features, including subjective and cancer-related words, part-of-speech, phrases indicating support types. It yields an F-measure of 0.840.

### 4 Method

We treat the first reply of a thread as the prime, and each following reply as a target; thus, we include several prime-target data points per thread.

Table 1: Data Distribution in Breast Cancer and Colorectal Cancer sub-forums in CSN

	# of Users	# of Threads	# of Comments	# of Comment Pairs	% of Users who have posted in both forums
Breast Cancer	4,290	14,061	153,160	139,444	2.66%
Colorectal Cancer	2,348	13,282	125,527	112,602	4.85%

To address our first and second research questions, we fit a generalized linear mixed effects regression model with binomial kernel to predict the support index of a reply post given lexical and syntactical alignment measures, the support index in the first reply, post distance, number of sentences of that post and interactions terms among predictors in the thread. All the predictors in the model have been rescaled (but not centered).

#### 4.1 Covariates

As a reminder, the model predicts the support index of a reply post (response variable) as a function of the following variables and interaction terms based on the previous work Wang et al. (2014). While the predicted informational support level is the sum of all predictors, pragmatic alignment is indicated as a positive correlation between the first reply’s informational support index and the response variable (see Table 3).

*First Reply Support Index:* This variable measures the proportion of informational support sentences in the post. A positive estimate ( $\beta$ ) would indicate positive correlation of support type between the posts.

*Post Distance:* The distance between the data point (current post) and first reply in the thread can be seen as a proxy for how much information has been discussed so far in this thread, or for how much time has elapsed between the posts. A large post distance indicates that a post is far away from the initial post. Distance is measured in number of posts, as this is the most informative number: dates and times are not indicative of when a member has actually read the posts. Distance is interesting in our context, as the priming effect decays rapidly, as shown previously for the case of this corpus (Wang et al., 2014).

*Linguistic Alignment:* As discussed in the previous section, we use two linguistic alignment measures, Linguistic Alignment (LILLA) and Syntactic Alignment (SILLA) to link the linguistic to support index. This main effect helps us address the second research question.

*Number of Sentences:* This variable approximates the complexity and the amount of information in a given post.

*Interaction terms between first reply Support Index and Post Distance:* The distance effect on pragmatic alignment would indicate a decay effect which is similar to decay previous observations for linguistic repetition (Reitter et al., 2006).

*Interaction terms between first reply Support Index and Linguistic Alignment Measurement:* To address our second research question, we measure the correlation between linguistic alignment and support matching.

*Interaction terms between Linguistic Alignment and Post Distance:* These two interaction terms evaluate a correlation between linguistic decay and support index, which follows the IAM’s cascade of alignment effects at different representational levels.

#### 4.2 Experimental Settings

We treat predicting online social support as a generalized mixed effects linear regression with binomial kernel (e.g., Jaeger, 2008). Compared to other black-box machine learning algorithms, such as SVM, this model is directly interpretable. It predicts the probability of a message being emotional/informational support message using logit-link kernel. In this regression model, we also consider the effect of different threads. This is of concern because social support types in different posts influenced by various topics and authors of initial posts. Therefore, we employ a logistic

Table 2: Model performance of full model and drop one feature off model. Numbers in bold shows the best performing feature set.

Dataset	Breast Cancer Sub Forum			Colorectal Cancer Sub Forum		
	pseudo R-sq	AIC	BIC	pseudo R-sq	AIC	BIC
- first reply Info Support Index	11.19%	61838	61912	14.53%	49514	49586
- PostDistance	18.29%	58658	58732	27.93%	46987	47060
- Lexical Alignment	10.84%	59060	59142	10.52%	47519	47601
- Syntactic Alignment	20.46%	58633	58717	28.78%	46943	47026
- # of Sentence in the current post	19.63%	58898	59000	<b>29.63%</b>	47111	47211
- Lexical Alignment $\times$ first reply Info Support Index	15.95%	58836	58938	28.97%	46889	46997
- Syntactic Alignment $\times$ first reply Info Support Index	19.55%	58695	58796	28.27%	46844	46944
- first reply Info Support Index $\times$ Post Distance	21.34%	58750	58851	29.40%	46963	47063
- Lexical Alignment $\times$ Post-Distance	18.96%	58513	58614	29.39%	46883	46983
- Syntactic Alignment $\times$ Post-Distance	17.99%	58633	58735	28.93%	<b>46664</b>	<b>46763</b>
Full Model	<b>21.82%</b>	<b>58511</b>	<b>58622</b>	28.58%	46774	46882

regression model with random effects, grouped by *ThreadID*.

We use the lme4 R package (Bates et al., 2014). To evaluate the performance of drop one feature off models, we give conditional pseudo R-squared (pseudo R-sq), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) (Burnham and Anderson, 2002). The conditional R squared shows the proportion of variance explained (Barto, 2014, using R package MuMIN). AIC and BIC are measures of the quality of logistic regression model of current dataset.

All results reported in Table 2 are produced from 10 random sub-sampling validation with 70% training and 30% testing splits of threads.

### 4.3 Experiment Results

Table 3 reports effect sizes and directions, and Table 2 gives the performance of the informational support index prediction using generalized mixed effects linear regression model with binomial kernel. In Table 2, it shows the pseudo R-squared, AIC and BIC of models for different set of features and the

full model.<sup>1</sup>

Overall, the full model which considers all the features we have listed out-performs other models. From the proportion of explained variable perspective (R-squared), it shows that the linguistic alignment and informational support index of the first reply are the two most important predictors, and increase a relatively larger proportion of variance explained in the model. Also, interaction terms considerably improve the model performance.

We rebuilt the model using the whole dataset in order to show and interpret effects of different predictors. The estimates and associated p-values given in Table 3 pertain to the two best models, predicting informational support index separately for the two sub-forums.

## 5 Discussion

Initially, we focus on the effect of the first reply support index addressing research question 1, whether online support provided by early responders

<sup>1</sup>We use the AIC scores to select the best generalized mixed effects linear regression model with binomial kernel, since AIC score evaluates the model based on both goodness of fit and model complexity (Burnham and Anderson, 2002).

Table 3: **Predicting Information Support Index** with generalized mixed effects linear regression models with binomial kernel, fitted to data from Breast Cancer and Colorectal Cancer Sub-Communities.

Dataset	Breast Cancer Sub Forum			Colorectal Cancer Sub Forum		
	beta	SE	p	beta	SE	p
Intercept	-1.552	0.022	0.000	-1.545	0.023	0.000
first reply Info Support Index	1.295	0.045	0.000	1.204	0.048	0.000
PostDistance	-0.351	0.051	0.000	-0.253	0.057	0.000
Lexical Alignment	-29.369	2.180	0.000	-47.786	2.588	0.000
Syntactic Alignment	0.989	0.266	0.000	0.861	0.276	0.001
# of Sentence in the current post	7.464	0.243	0.000	5.426	0.169	0.000
Lexical Alignment $\times$ first reply Info Support Index	58.150	3.592	0.000	73.163	4.410	0.000
Syntactic Alignment $\times$ first reply Info Support Index	0.433	0.655	0.509	-0.084	0.806	0.917
Post Distance $\times$ first reply Info Support Index	-0.114	0.113	0.310	-0.244	0.128	0.057
Lexical Alignment $\times$ Post-Distance	-23.346	6.080	0.000	-20.691	7.108	0.004
Syntactic Alignment $\times$ Post-Distance	0.256	0.702	0.715	0.181	0.807	0.823

influences the support index of replies from later responders. According to the regression models in Table 3, the support index of the first reply is positively correlated with the support indices of the later replies in both datasets. In short, people align at the pragmatic level when it comes to overall communicative intent. The intent of the first reply is matched by the intent shown in future replies.

Similar to linguistic alignment effects, we also consider the post distance effect. Previous studies (Reitter et al., 2006) showed that strong syntactic adaptation diminishes in seconds in spoken dialogue corpora. This phenomenon also has been found for individual syntactic constructions in written and spoken language (see Pickering and Ferreira, 2008, for a review) and also in dialogues in online communities (Wang et al., 2014). In order to test and measure the effect of early messages on later messages, we examine whether support index has the same characteristic. There are two components to an answer. First, the regression model (Table 3) suggests that informational support index generally decreases by post distance. In other words, less informational support is given as discussions proceed. It is worthwhile to note that

conversations shift towards emotional support in this support-oriented community. Does alignment decay by distance? This answer is given by the interaction between distance and support index of the first reply. Evidence for such decay is weak: we have no support for decay in the Breast Cancer case, and some decay ( $\beta = -0.244, p = 0.057$ ) in the Colorectal Cancer forum.

Another notable result is how linguistic and pragmatic alignment interact. The LILLA measure quantifies lexical adaptation between messages. Lexical alignment is reliably indicative of emotional support (negative information support) in both forums. The reasons for this correlation may be found in properties of informational support in both datasets. Informational support provided at a later time is likely to include new information, introducing new words. Emotional support, on the other hand, implies more consistent word choice. Syntactic adaptation (SILLA) shows no effect of syntactic alignment on support index.

To address our second initial question, we also evaluate the relationship of linguistic and pragmatic alignment using interaction terms between Linguistic Alignment and first reply

Support Index. From a theoretical perspective and previous empirical results, we expect to see that adaptation is consistent across different linguistic choices: it may be due to a cascade of priming effects and joint situational understanding (Garrod and Pickering, 2009), joint languages (Fusaroli et al., 2012), and/or a cognitive (memory) process that is common to the different choices (Reitter et al., 2011). We find a strong positive interaction between lexical alignment and Informational Support Index in the first reply. This means that when first-reply (prime) and other-reply (target) align in terms of their kinds of social support, then they also tend to show much more lexical alignment. The same cannot be said for the syntactic level.

Linguistic adaptation is correlated with high-level alignment. In order to validate this theoretical effect on our corpora, we observe interaction effects between lexical alignment and the support type alignment. We caution the reader, however, that this interaction effect is expected given that our measure of support type is a function not least of word choices. Thus, these predictors are by no means independent. However, as stated before, lexical alignment also correlates with stronger emotional support. The interaction effect of lexical alignment and post distance, present in both datasets, suggests that in later portions of each thread, lexical alignment is no longer predictive of such emotional support.

To summarize, the observations of main effects suggest that the type of support provided by early responders on the thread positively influences the type of support provided by later responders in our data. That is, pragmatic adaptation based on support index exists in our data. Also, the observations provide clues that informational support messages are more likely to be provided at the beginning of the thread discussions.

Moreover, with regard to our research question 2, there is a correlation between some linguistic alignment measurements and support index. Naturally, *these results are observational*: taken by themselves, they suggest no causality. We make our argument solely because the hypotheses tested were motivated by theoretical predictions. Our results are compatible with a theoretical perspective that explains mutual understanding and successful

communication as being aided by a cascade of priming or language adaptation effects (Pickering and Garrod, 2004).

## 6 Conclusion

Motivated by the large proportion of online social support in peer-to-peer support online communities, we quantify and predict online support in the thread-based conversations. In a regression model, we have considered multiple factors, such as previous messages, linguistic alignment, and complexity. The results point to alignment phenomena at a pragmatic level. Such alignment tends to coincide with alignment of word choices. Both of these results are, to our knowledge novel. The interpretation of our regression model is congruent with the interactive alignment theory (Pickering and Garrod, 2004).

From an applied perspective the models we fitted to the forum data could facilitate filters to display certain useful posts, or to improve ranking of search results after analyzing a specific users' needs (i.e. providing results with high informational support index for seeking informational support). We believe that it might help health communities to improve user experience.

## 7 Acknowledgement

This research was made possible by a collaboration agreement between Penn State and the American Cancer Society. The authors would like to thank Kenneth Portier and Greta E. Greer for preparing the CSN dataset. Also, the authors would like to thank Prakhar Biyani for providing the source code of sentence classifier as well as Prasenjit Mitra and Yang Xu for their helpful discussions.

## References

- Tim Althoff, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. How to ask for a favor: A case study on the success of altruistic requests. *arXiv preprint arXiv:1405.3282*, 2014.
- Lars Backstrom, Jon Kleinberg, Lillian Lee, and Cristian Danescu-Niculescu-Mizil. Characterizing and curating conversation threads: expansion, focus, volume, re-entry. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 13–22. ACM, 2013.



- Antonina Bambina. *Online social support: the interplay of social networks and computer-mediated communication*. Cambria press, 2007.
- Kamil Barto. *MuMIn: Multi-model inference*, 2014. R package version 1.10.5.
- Douglas Bates, Martin Maechler, Ben Bolker, and Steven Walker. *lme4: Linear mixed-effects models using Eigen and S4*, 2014. R package version 1.1-7.
- Kirsten Bergmann and Stefan Kopp. Gestural alignment in natural dialogue. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society (CogSci 2012)*, 2012.
- Prakhar Biyani, Cornelia Caragea, Prasenjit Mitra, and John Yen. Identifying emotional and informational support in online health communities. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 827–836, Dublin, Ireland, August 2014. Dublin City University and Association for Computational Linguistics.
- J Kathryn Bock. Syntactic persistence in language production. *Cognitive psychology*, 18(3):355–387, 1986.
- Holly P Branigan, Martin J Pickering, and Alexandra A Cleland. Syntactic co-ordination in dialogue. *Cognition*, 75(2):B13–B25, 2000.
- Ngot Bui, John Yen, and Vasant Honavar. Temporal causality of social support in an online community for cancer survivors. 2015.
- Kenneth P Burnham and David R Anderson. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer, 2002.
- Justin Cheng, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. How community feedback shapes user behavior. In *Eighth International AAAI Conference on Weblogs and Social Media*, 2014.
- Kenneth W Church. Empirical estimates of adaptation: the chance of two noriegas is closer to  $p/2$  than  $p$ . In *Proceedings of the 18th conference on Computational linguistics*, pages 180–186. Association for Computational Linguistics, 2000.
- Sheldon Cohen and Garth McKay. Social support, stress and the buffering hypothesis: A theoretical analysis. *Handbook of psychology and health*, 4:253–267, 1984.
- Cristian Danescu-Niculescu-Mizil and Lillian Lee. Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs. In *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics*, pages 76–87. Association for Computational Linguistics, 2011.
- Cristian Danescu-Niculescu-Mizil, Michael Gamon, and Susan Dumais. Mark my words!: linguistic style accommodation in social media. In *Proceedings of the 20th international conference on World Wide Web*, pages 745–754. ACM, 2011.
- Amit Dubey, Frank Keller, and Patrick Sturt. Parallelism in coordination as an instance of syntactic priming: Evidence from corpus-based modeling. In *Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing*, pages 827–834, Vancouver, Canada, 2005.
- Michael C Frank and Noah D Goodman. Predicting pragmatic reasoning in language games. *Science*, 336(6084):998–998, 2012.
- Riccardo Fusaroli, Bahador Bahrami, Karsten Olsen, Andreas Roepstorff, Geraint Rees, Chris Frith, and Kristian Tylén. Coming to terms quantifying the benefits of linguistic coordination. *Psychological Science*, 23(8):931–939, 2012.
- Simon Garrod and Anthony Anderson. Saying what you mean in dialogue: A study in conceptual and semantic coordination. *Cognition*, 27:181–218, 1987.
- Simon Garrod and Martin J Pickering. Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, 1(2):292–304, 2009.
- Stefan Th. Gries. Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research*, 34(4):365–399, 2005.
- David Huffaker, Joseph Jorgensen, Francisco Iacobelli, Paul Tepper, and Justine Cassell. Computational measures for language similarity across time in online communities. In *Proceedings of the HLT-NAACL 2006 Workshop on Analyzing Conversations in Text and Speech*, pages 15–22. Association for Computational Linguistics, 2006.
- T. Florian Jaeger. Categorical data analysis: Away from anovas (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4): 434–446, 2008.
- T. Florian Jaeger and Neal Snider. Implicit learning and syntactic persistence: Surprisal and cumulativity. *University of Rochester Working Papers in the Language Sciences*, 3(1):26–44, 2007.
- Simon Jones, Rachel Cotterill, Nigel Dewdney, Kate Muir, and Adam Joinson. Finding zelig in text: A measure for normalising linguistic accommodation. In *25th International Conference on Computational Linguistics*. University of Bath, 2014.
- Dan Klein and Christopher D Manning. Fast exact inference with a factored model for natural language

- parsing. *Advances in Neural Information Processing Systems*, pages 3–10, 2003.
- Catherine Penny Hinson Langford, Juanita Bowsher, Joseph P Maloney, and Patricia P Lillis. Social support: a conceptual analysis. *Journal of advanced nursing*, 25(1):95–100, 1997.
- Christine Kerres Malecki and Michelle Kilpatrick Demaray. What type of support do they need? investigating student adjustment as related to emotional, informational, appraisal, and instrumental support. *School Psychology Quarterly*, 18(3):231, 2003.
- Martin J Pickering and Victor S Ferreira. Structural priming: a critical review. *Psychological Bulletin*, 134(3):427, 2008.
- Martin J. Pickering and Simon Garrod. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27:169–225, 2004.
- Kenneth Portier, Greta E Greer, Lior Rokach, Nir Ofek, Yafei Wang, Prakhar Biyani, Mo Yu, Siddhartha Banerjee, Kang Zhao, Prasenjit Mitra, and John Yen. Understanding topics and sentiment in an online cancer survivor community. *JNCI Monographs*, 2013(47):195–198, 2013.
- Baojun Qiu, Kang Zhao, Prasenjit Mitra, Dinghao Wu, Cornelia Caragea, John Yen, Greta E Greer, and Kenneth Portier. Get online support, feel better—sentiment analysis and dynamics in an online cancer survivor community. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom)*, 2011 IEEE Third International Conference on, pages 274–281. IEEE, 2011.
- David Reitter and Johanna D Moore. Predicting success in dialogue. In *Annual Meeting of the Association for Computational Linguistics*, volume 45, page 808, 2007.
- David Reitter and Johanna D. Moore. Alignment and task success in spoken dialogue. *Journal of Memory and Language*, 76:29–46, 2014.
- David Reitter, Johanna D. Moore, and Frank Keller. Priming of syntactic rules in task-oriented dialogue and spontaneous conversation. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society (CogSci)*, pages 685–690, Vancouver, Canada, 2006.
- David Reitter, Frank Keller, and Johanna D. Moore. A computational cognitive model of syntactic priming. *Cognitive Science*, 35(4):587–637, 2011.
- Barbara K Rimer, Elizabeth J Lyons, Kurt M Ribisl, J Michael Bowling, Carol E Golin, Michael J Forlenza, and Andrea Meier. How new subscribers use cancer-related online mailing lists. *Journal of medical internet research*, 7(3), 2005.
- Lauren E Scissors, Alastair J Gill, and Darren Gergle. Linguistic mimicry and trust in text-based cmc. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, pages 277–280. ACM, 2008.
- Yla R Tausczik and James W Pennebaker. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54, 2010.
- Tatiana A Vlahovic, Yi-Chia Wang, Robert E Kraut, and John M Levine. Support matching and satisfaction in an online breast cancer support community. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 1625–1634. ACM, 2014.
- Yafei Wang, David Reitter, and John Yen. Linguistic adaptation in online conversation threads: Analyzing alignment in online health communities. In *Proceedings of the Fifth Workshop on Cognitive Modeling and Computational Linguistics*, pages 55–62, Baltimore, Maryland, USA, June 2014. Association for Computational Linguistics.
- Yi-Chia Wang, Robert Kraut, and John M Levine. To stay or leave?: the relationship of emotional and informational support to commitment in online health support groups. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 833–842. ACM, 2012.
- Kang Zhao, John Yen, Greta Greer, Baojun Qiu, Prasenjit Mitra, and Kenneth Portier. Finding influential users of online health communities: a new metric based on sentiment influence. *Journal of the American Medical Informatics Association*, pages amiajnl–2013, 2014.