

#mygoal: Finding Motivations on Twitter

Marc T. Tomlinson, David B. Bracewell, Wayne Krug, David Hinote, Mary Draper

Language Computer Corporation

Richardson, TX 75080

{marc,david,wayne,dhinote}@languagecomputer.com

Abstract

Our everyday language reflects our psychological and cognitive state and effects the states of other individuals. In this contribution we look at the intersection between motivational state and language. We create a set of hashtags, which are annotated for the degree to which they are used by individuals to mark-up language that is indicative of a collection of factors that interact with an individual's motivational state. We look for tags that reflect a goal mention, reward, or a perception of control. Finally, we present results for a language-model based classifier which is able to predict the presence of one of these factors in a tweet with between 69% and 80% accuracy on a balanced testing set. Our approach suggests that hashtags can be used to understand, not just the language of topics, but the deeper psychological and social meaning of a tweet.

Keywords: motivation, twitter, psychological state

Language is used to communicate. Its words and phrases can endow a listener with knowledge about events and entities in the world, but language also has a purpose that transcends the topic of a conversation. Language allows an individual to express information about their cognitive state, their desires, opinions, and motivations. A particularly interesting aspect of language is how its content can inform our perception of an individual's motivations. When someone makes a new year's resolution, or your colleague says that they'll finish a paper, how motivated are they? In this paper we present a methodology for creating a large corpus of tweets containing language which informs an individual's level of motivation.

We cast our approach to understanding the expressions of motivational factors within a more general framework for understanding speech acts (Searle, 1969; Bunt, 2011). Speech acts provide a theoretical framework to explore the motivational implicatures of an utterance. In this contribution we use the term motivational act to represent utterances by individuals that either inform their motivation for an action or affect the motivation of another individual. We outline a set of three factors, detectable through analysis of an individual's language, that are derived from work in psychology and could be combined to model an individual's motivational level.

Social media platforms have enabled new forms of discourse and also provided enormous quantities of communications. These communications provide a rich source of evidence about how people express their personal states. For instance, the popular micro-blogging service Twitter provides an exceptionally useful source of user-generated content which has attracted considerable interest from researchers in computational linguistics (Ritter et al., 2009; Gimpel et al., 2011). Most of the language processing on tweets has involved the identifications of sentiment (Davidov et al., 2010), emotion (Roberts et al., 2012), summarization (Sharifi et al., 2010), and conversational models of Dialogue acts (Ritter et al., 2009), or lexical and semantic processing.

Instead of directly annotating the presence of a motivational

act within a given text span we take an approach based on distant supervision (Mintz et al., 2009). Distant supervision involves the use of a small set of annotations that link to a larger knowledge base that contains noisy instances of those annotations. For our small set of annotations, we annotate hashtags for their degree of association with tweets that exhibit a given motivational act. We can then train a model from the language contained in the tweets marked by those hashtags. In contrast to approaches that learn the topical (Ramage et al., 2010) or sentimental (Davidov et al., 2010) value of a hashtag we suggest that these tags can be used to signal complex personal states.

We first test the accuracy of our corpus creation methodology through inter-annotator agreement. We then demonstrate that a language model can be used to effectively learn the correlation between the language contained within a tweet and its likely effect on motivation. Finally, we present a time-series analyses investigating the changes in the likelihood of individuals tweeting about motivational content as a function of the day of week.

1. Related Work

While twitter is difficult for NLP, it also provides important insight into language use because authors generate a plethora of posts about themselves, and more importantly, authors use hashtags to highlight the meaning of their posts. To date, most of the research investigating hashtag usage has focused on utilizing them to inform models of topics (Ramage et al., 2010) or models of sentiment (Davidov et al., 2010). However, hashtags can also signal that the tweet contains language expressing goals (#iwish) and rewards (#iwon). By pairing the twitter tags with the language contained in the tweets we can, for the first time, have a very large, very valuable dataset of self-expressed intentions in language.

One related area of work covers annotations of beliefs, opinions, sentiment, and desires of individuals; *private states* (Wilson and Wiebe, 2005). This approach to annotation, in general, focuses on creating rich frames which espouse the world-view of participants in a discourse

Table 1: Example motivational acts, associated hashtags, and sample tweets.

Motivational Action	Sample Tags	Sample Tweets
Goal	#goalinlife, #mywish	“3 more days of studying”
Control	#dowhatisay, #kissmyfeet	“I defy the law of gravity”
Skill	#ikickass, #madskills	“I can’t be stopped”
Lack of Control	#help, #ihavenocontrol	“the worlds caving in”
Negative Reward Self	#fml, #crap	“I just locked the keys in my car”
Negative Reward Other	#worstdriverever, #awkward	“It does make me cringe”
Positive Reward Self	#whyismile, #victoryismine	“my cats make me smile”
Positive Reward Other	#ff, #thatsbadass	“Solar panels on the white house”

or a narrative (Elson, 2012). These frames attempt to solve the problem of correctly capturing the state of an individual, what they know and who it is known about. In contrast, our current contribution focuses on identifying the elements that belong within the frame, though we do look at recovering the source of positive and negative rewards.

Similarly focused to the idea of identifying motivational acts is work looking at identifying social acts. In the spirit of dialogue acts (Core and Allen, 1997; Stolcke et al., 1998; Bunt, 2011), social acts capture the social implications present within a statement. Researchers have recently begun to construct and annotate social acts. For example, Bender et al. (2011) created an annotated corpus of social acts relating to *authority claims* and *alignment moves*. Authority claims are statements by individuals that demonstrate their authority, while alignment moves are statements which suggest solidarity between two individuals or set an individual against another. Similarly, Bracewell et al. (2012) and Tomlinson et al. (2012) annotate corpora for a broader selection of social acts breaking down alignment moves into a more fine grained category. For example, the researchers identify utterances that show disrespect, gratitude, solidarity, challenges to credibility, and others.

2. Motivational Acts

We classify three main types of motivational acts that an individual can use which inform their intentions and the effort they are willing to expend towards those intentions. These acts are derived from work in psychology understanding how individuals perceive intentionality in others (Malle and Knobe, 1997; Sloman et al., 2012), and what factors change an individual’s motivation.

The first act that we look at are comments that express goals or indicators of a goal orientation in an individual. The second act looks for evidence that the individual has (or thinks they have) skill or control to act within the environment. The act identifies expressions which express a value for an action indicate positive social value for the individual’s work. This act is further refined into separate categories for self-directed rewards and reward statements directed at other individuals. The three acts: Goals, Control, and Rewards, and their subtypes are shown in Table 1.

Goals encode an individual’s desire for an event or the reward associated with an event’s outcome. Examples of linguistic expressions demonstrating this factor are as follows.

(1) I want to quit smoking and be healthy.

This goal expresses an action (quitting smoking) and establishes a goal for a reward (being healthy). In contrast, a statement such as

(2) I want to be the new president.

expresses an action (becoming president) but requires making inferences about the probable rewards for the individual if they are successful. Alternatively, expressions can also establish goals but only express rewards, such as

(3) I want to be famous.

which expresses a clear expectation for a reward resulting from some series of events, but requires inference about the details of the future events that the individual might precipitate to achieve their goal. Does the individual want to be a rock star or a college professor? Expressions of goals are important not just for understanding motivations, but also for understanding the probability of success (Albarracin et al., 2011; Locke, 1968).

The second act that we considered are expressions which indicate skill and control over a situation. Like the previous factor, these expressions can range from abstract to concrete

(4) I can do anything.

(5) I am able to syntactically parse a sentence.

However, this factor could also cover expressions indicating differences in an individual’s perception of their locus of control, such as by saying

(6) He attacked me.

which indicates that the speaker lacked control over the conflict event. Individual’s that perceive themselves as being in control of an event are likely to expend more effort on the event’s outcome (Ajzen, 1991). We further refined this act into three separate subtypes, expressions of control, expressions of skill, and expressions indicating a lack of control.

The third act that we considered were mentions of rewards. Rewards interact with goals and intentions through the work of Kehneman and Tversky (1979) treatise on

prospect theory. Kahneman and Tversky discuss three concepts that affect how an individual values a reward: a) Reference points - rewards are valued in how far they deviate above (positive reward, gain) or below (negative reward, loss) a given reference point; b) Loss Aversion - avoiding a loss is treated as being more important than an equivalent gain. An individual expressing

(7) I am not going to pay that five dollar fine.

is more motivated than an individual expressing

(8) If I get this done I get an extra five dollars.

and 3) Diminishing sensitivity - the value of a change is not linear but decreases as the point gets further from the referent. The individual expressing

(9) With good work, I can add forty dollars to my twenty dollar take-home pay.

is more motivated than one expressing

(10) With good work, I can add forty dollars to my ten-thousand dollar take-home pay.

In addition to monetary rewards, one must also consider social rewards that come from support by community members, such as statements like “good work”. Reward mentions can also influence and reveal the moral values of individuals, and hence their motivations towards an event (Knobe, 2003). However, some care needs to be taken in this as some people actually seek out and are motivated by negative comments (Finkelstein and Fishbach, 2012). We further divided rewards into four subtypes, self-directed positive rewards, other directed positive rewards, self-directed negative rewards, and self-directed other rewards.

The ways in which individuals utilize these three motivational acts reflect the individual’s underlying motivation to perform. Evidence that an individual is in control of a situation that they have expressed a goal for and would receive positive rewards for would indicate a greater motivation for that goal. In contrast, an individual that expresses a lack of control coupled with negative rewards suggests that the individual is not motivated.

3. Data and Annotation

Goals, rewards, and perceptions of control can be expressed in a myriad of different ways in text, sometimes very clearly, e.g. “I want to do better”, and sometimes only implied through the use of future tense. To create a resource for understanding the wide variety of ways in which individuals can express motivational acts we created a large repository of tweets. We looked at twitter because it provides a unique challenge and set of benefits for natural language understanding. Tweets provide a significant challenge because they often contain spelling mistakes, non-traditional grammatical usage, and shorthand. However, tweets have the benefit of conveying many different forms of information, being written by people across a spectrum of socio-economic and cultural background, and being available in massive quantities. Further, we suspect

that computational systems trained to detect motivational acts in twitter posts could be easily applied to the problem of identifying the same factors in other forms of communications. However, training a model on tweets requires a large number of annotations, which can be an expensive and time-consuming operation. Instead we can look at annotating groups of hashtags for how they are used by individuals to mark text which indicate our factors of interest.

Hashtags are words or phrases that are often included in tweets to signify the topic or non-obvious meaning of the tweet. Some hashtags have meanings that can be derived from the words making up the tag (e.g. #mygoal – is used to indicate that the linguistic content of the tweet expresses a goal), while others are related to Internet memes and require cultural knowledge (e.g. #fml, used to express negative things happening in an individual’s own life). To date, most of the research has focused on the topical nature of hashtags. However, hashtags can also signal that the tweet contains language expressing complex psychological factors (#iwish), rewards (#iwon), or sociological phenomena (#ioweyou).

The goal of the annotation task was to generate a set of hashtags which were used by individuals to mark language that exhibited one of the motivational acts. The first task for the annotators was to produce a set of probable hashtags that could be used to mark each motivational act. These tags were generated through trial and error utilizing twitter’s default search website to understand the frequency of the tag and the characteristics of the tweets labeled with the hashtag. Examples of hashtags generated during this stage where #mygoal, #iwon (goal based); #madskills, #imapro (skill); #dowhatisay, #kissmyfeet (control); #fml, #noob (negative rewards); #ff, #proud (positive rewards).

The second task was to look at the hashtags that were used with tweets labeled with one of the known hashtags. For example, an inspection of the tweets marked with #mygoal were sometimes labeled with #icandothis. The annotators would then check the validity of the tag #icandothis. The annotators would bin each discovered hashtag as to its likely category based on the tweets shown through twitter’s online search utility or throw the tag out. After a set of 200 tags had been identified a single annotator, utilizing a scale from 1 - 5, rated each hashtag as to how well it was associated with the category. Because of the initial pre-identification and association step, only six tags were marked with scores less than 3 and fourteen of the tags received a 3. The remaining 180 tags were marked with either a 4 or a 5.

All hashtags receiving a 4 or a 5 were then reviewed by a second annotator, the agreement rate on hashtags marked with a 4 or 5 was 86%. This produced a final set of 157 hashtags which were strongly associated with one of the motivational acts. Example tags are shown in Table 1. Example of the goal related tweets, marked with #mygoal range from “3 more days of studying #iwillsurvive #4.0 #mygoal” to “Looking for a bigger house By December I wane[sic] be out this house in a bigger house #mygoal”.

We then utilized the hashtags to find tweets from twitter’s streaming api that contained one of the hashtags of interest or were related to a hashtag of interest (same author).

Table 2: Training and testing sizes for N-gram classifiers with resultant accuracy and classifier bias for labeling a tweet as a positive instance of the class. All test and train splits are 50/50 between positive and negative instances.

Motivational Action	# Hashtags	# Train	# Test	accuracy
Goal	23	83,838	20,960	79.8%
Control	18	153,136	38,286	70.2%
Skill	7	78,146	18,764	74.1%
Lack of Control	19	127,040	28,296	68.9%
Negative Reward Self	30	100,996	25,250	68.6%
Negative Reward Other	47	157,582	39,396	69.6%
Positive Reward Self	8	158,250	39,564	69.3%
Positive Reward Other	5	103,948	25,988	78.9%

This created an index of approximately 7.5 million tweets. In our collection hashtags exhibiting control contained the largest number with 153136 tweets, while we only collected 78,146 tweets which were marked with hashtags indicating skill. The number of hashtags used to define each category and the number of tweets found with one of those hashtags is shown in Table 2. We discarded any tweets which were labeled with more than one hashtag from one of our categories.

4. Automatically recognizing motivational acts on twitter through distant supervision

While the most successful approaches to speech act classification on text have focused on integrating multi-layer models which examine both the utterance content and the surrounding utterances (?), for tweets, the surrounding context is problematic to define. Many tweets are sent in reference to external context, commenting on happening in the real world instead of in response to other tweets. Thus for our initial experiments we only modeled the self-contained language within each tweet.

For training and testing purposes we removed all URLs, hashtags, and @users from the tweets. We then discarded tweets that were less than two words long. Removing all hashtags from the tweets is conservative, because we removed the classifier’s ability to directly learn co-occurring hashtags, however we wanted to ensure that we would minimize deficient solutions and maximize our ability to transfer the twitter classifier to other domains, such as Flickr.

We utilized a naive-Bayes language model for each of the different motivational acts. The models used n-grams within the tweets as features. We used n-grams of 2-4 words in length. The model was trained using tweets containing one of the hashtags representing that motivational act as positive data and a random sampling of tweets containing hashtags from one of the other categories as negative data. We tested the performance of each language model using an 80/20 split for testing and training. Any tweets which contained hashtags for more than one class were not included in either sample.

The accuracy of the resultant classifiers (shown in table 2) suggest that our annotation procedure was accurately iden-

tifying tweets that had strong similarities in the language that was used in the tweet and that the language expressing each motivational action could be adequately captured by the model.

Inspections of tweets with the labels suggest that one of the main sources of error were sarcastic usage of the hashtags. This problem was particularly poignant with one expression *thank you*. Tweets containing only the words *thank you* were more associated with a negative reward than a positive reward. Inspection of our twitter data showed that when *thank you* was tagged, it was more likely to be with a negative tag indicating that it was being used sarcastically. In contrast, when *thank you* was being used literally, the authors felt no need to tag it or often included more details about the action that precipitated the gesture of thanks. This suggests one area of fruitful research.

In addition to results presented in Table 2, we also evaluated the annotations and n-gram model by inspecting the tweets that were automatically labeled by the model. We looked at the distribution of the hashtags within those tweets (as a reminder, hashtags were not used as input in training). We found that the most frequently occurring hashtags with each of our factors were indeed related to our features. In addition, the set of highly ranked hashtags contained many which we had not considered. For example, tweets marked by the model as being about a goal were very likely to contain the hashtags #day1 and #day2. These two tags are used by individuals to express the first and second day of pursuing a new goal.

This result shows that our approach can also identify new hashtags which are associated with our cognitive factors. This supports the idea that an iterative distant-supervised annotation procedure would be an extremely beneficial. This would allow the system to propose new hashtags which could be evaluated by an annotator and reincorporated into the system, resulting in a more accurate system.

5. Cycles of Motivation

The classifier allows for investigation into the nature of motivational statements on twitter. As a first step we wanted to look at the effect of time on tweeting. A seven day seasonality effect is common in twitter analyses of public sentiment (cf. OConnor et al. 2010) which shows variation in

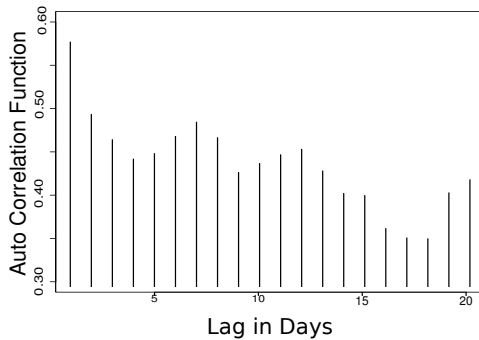


Figure 1: The autocorrelation of the probability of a tweet expressing a goal over time. Of note is the wave form that peaks approximately every 7 days.

an individual’s likelihood of expressing positive sentiment throughout the week. The seasonality of a time-series reflects a correlation between changes in the expression of a factor and seasons (often days, months, or years). In contrast to seasonal time-series, a stationary time-series is one where changes in the factor are not correlated across time. We assess the stationarity of a time-series by looking at the autocorrelation of our factor y , which is the probability of a tweet sent on a given day exhibits a particular motivational act. We find the probability of p_j^c , which compares the correlation of y_t^c with y_{t-j}^c for a given lag j .

$$p_j^c = \frac{\text{cov}(y_t^c, y_{t-j}^c)}{\sqrt{\text{var}(y_t^c) \cdot \text{var}(y_{t-j}^c)}}$$

Figure 1 shows a plot of y for tweets in English that express a goal. The lag 0 autocorrelation is not shown, since it is always 1. From the plot we can see that there is a sinusoidal pattern to the autocorrelation. It peaks and troughs approximately every seven days. This suggests that our data is not stationary, but instead seasonal, with a 7 day window. This suggests that the number of goal mentions on a given day change depending on the time of week.

The next graph has been augmented with days of the week. Figure 2 shows how goals and statements of control differ in their pattern throughout the week. Goals are more likely to be mentioned on Monday, associated with the start of the work week. In contrast, statements of control are more likely to be issued later in the week. The other acts exhibit similar cyclical patterns though with smaller magnitudes. For example, mentions of positive rewards peak on the weekend and then dip on Monday. The graphs for negative rewards show that they are infrequent at the beginning of the week, but slowly build until peaking on Thursday, then declining for the weekend, though this shows the weakest seasonality effect.

Presumably, further analysis will reveal differences in the types of goals that are mentioned throughout the work week as well. The non-stationarity certainly suggests that goals on Mondays should be focused on work related goals, though it could also be simply a strong association between the start of the week and establishment of general goals.

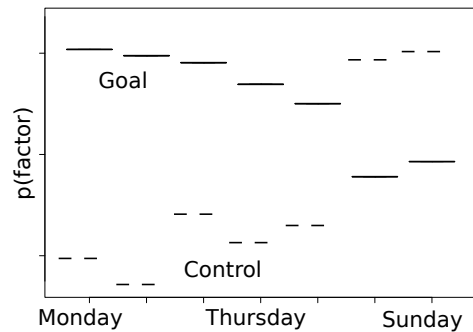


Figure 2: The probability of a tweet exhibiting a goal act (solid line) or control act (dashed line) by day of the week on twitter.

6. Conclusion

We created a novel corpus designed for identifying the motivational implicatures of language used by individuals on twitter. These annotations were then used to build a language model of three different types of motivational acts; goals, statements of control, and expressions of rewards. The resulting motivational acts could be utilized to understand the amount of effort an individual would be willing to expend to accomplish a goal or task. Critically, this work suggests that hashtags are used for more than just marking the topic of a tweet but can function to mark complex expressions which indicate an individual’s social or cognitive state.

Acknowledgment

This research was funded by the Intelligence Advanced Research Projects Activity (IARPA) through the Department of Defense US Army Research Laboratory (DoD / ARL). The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoD/ARL, or the U.S. Government.

7. References

- Ajzen, I. (1991). The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes*, 50:179–211.
- Albarracin, D., Hepler, J., and Tannenbaum, M. (2011). General Action and Inaction Goals: Their Behavioral, Cognitive, and Affective Origins and Influences. *Current Directions in Psychological Science*, 20(2):119–123, April.
- Bender, E., Morgan, J., Oxley, M., Zachry, M., Hutchinson, B., Marin, A., Zhang, B., and Ostendorf, M. (2011). Annotating social acts: Authority claims and alignment moves in wikipedia talk pages. *ACL HLT 2011*, (June):48.
- Bracewell, D. B., Tomlinson, M. T., Brunson, M., Plymale, J., Bracewell, J., and Boerger, D. (2012). Annotation of Adversarial and Collegial Social Actions in Dis-

- course. In *6th Linguistic Annotation Workshop*, number July, pages 184–192.
- Bunt, H. (2011). The semantics of dialogue acts. *International Conference on Computational Semantics*, pages 1–13.
- Core, M. G. and Allen, J. F. (1997). Coding Dialogs with the DAMSL Annotation Scheme. In Traum, D., editor, *AAAI Fall Symposium on Communicative Action in Humans and Machines*, pages 28–35. AAAI, American Association for Artificial Intelligence.
- Davidov, D., Tsur, O., and Rappaport, A. (2010). Enhanced Sentiment Learning Using Twitter Hashtags and Smileys. In *Coling*, number August, pages 241–249.
- Elson, D. K. (2012). DramaBank : Annotating Agency in Narrative Discourse. In *LREC*, pages 2813–2819.
- Finkelstein, S. R. and Fishbach, A. (2012). Tell Me What I Did Wrong: Experts Seek and Respond to Negative Feedback. *Journal of Consumer Research*, 39(1):22–38, June.
- Gimpel, K., Schneider, N., Connor, B. O., Das, D., Mills, D., Eisenstein, J., Heilman, M., Yogatama, D., Flanigan, J., and Smith, N. A. (2011). Part-of-Speech Tagging for Twitter : Annotation , Features , and Experiments. In *Proceedings of the Association for Computational Linguistics*, number 2.
- Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, 63(3):190–194, July.
- Locke, E. A. (1968). Toward a theory of task motivation and incentives. *Organizational behavior and human performance*, 3(2).
- Malle, B. and Knobe, J. (1997). The Folk Concept of Intentionality. *Journal of Experimental Psychology*, 33(2):101–121.
- Mintz, M., Bills, S., Snow, R., and Jurafsky, D. (2009). Distant supervision for relation extraction without labeled data. In *Proceedings of ACL-IJCNLP*, number 2005.
- Petukhova, V. and Bunt, H. (2011). Incremental dialogue act understanding. *IWCS '11 Proceedings of the Ninth International Conference on Computational Semantics*, pages 235–244.
- Ramage, D., Dumais, S., and Liebling, D. (2010). Characterizing Microblogs with Topic Models. In *Proc of ICWSM 2010*.
- Ritter, A., Cherry, C., and Dolan, B. (2009). Unsupervised Modeling of Twitter Conversations. In *HTL-NAACL*.
- Roberts, K., Roach, M., Johnson, J., Gurthrie, J., and Harabagiu, S. M. (2012). Empatweet: Annotating and detecting emotions on Twitter. *Proceedings of LREC2012*, pages 3806–3813.
- Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*, volume 0. Cambridge University Press.
- Sharifi, B., Hutton, M.-a., and Kalita, J. (2010). Summarizing Microblogs Automatically. In *ACL-HLT*, number June, pages 685–688.
- Slooman, S. a., Fernbach, P. M., and Ewing, S. (2012). A Causal Model of Intentionality Judgment. *Mind & Language*, 27(2):154–180, April.
- Stolcke, A., Shriberg, E., Bates, R., Coccaro, N., Jurafsky, D., Martin, R., Meteer, M., Ries, K., Taylor, P., and Ess-Dykema, C. V. (1998). Dialog Act Modeling for Conversational Speech. In *Applying Machine Learning to Discourse Processing*, pages 98–105. AAAI Press.
- Tomlinson, M., Bracewell, D., Draper, M., Almissour, Z., Shi, Y., and Bensley, J. (2012). Pursing power in arabic on-line discussion forums. *Proceedings of LREC2012*, pages 1359–1364.
- Wilson, T. and Wiebe, J. (2005). Annotating Attributions and Private States. In *Workshop on Frontiers in Corpus Annotation II: Pie in the Sky*, number June, pages 53–60.