Determining Whether and When People Participate in the Events They Tweet About

Krishna C. Sanagavarapu, **Alakananda Vempala** and **Eduardo Blanco**
Human Intelligence and Language Technologies Lab
University of North Texas
Denton, TX, 76203
{KrishnaSanagavarapu, AlakanandaVempala}@my.unt.edu, eduardo.blanco@unt.edu

Abstract

This paper describes an approach to determine whether people participate in the events they tweet about. Specifically, we determine whether people are participants in events with respect to the tweet timestamp. We target all events expressed by verbs in tweets, including past, present and events that may occur in the future. We present new annotations using 1,096 event mentions, and experimental results showing that the task is challenging.

1 Introduction

Twitter has quickly become one of the most popular social media sites: it has 313 million monthly active users, and 500 million tweets are published daily. People tweet about breaking news, world and local events (e.g., eclipses, road closures), and personal events ranging from important life events (e.g., graduating from college) to mundane events such as commuting and attending a party.

People tweet not only about events in which they participate, but also events in which they do not participate but are somehow relevant (e.g., John Doe may tweet about his nephew graduating from college). More specifically, people can participate in the events they tweet about (underlining indicates events of interest below) prior to tweeting (e.g., *When I come back to London, I realise how much I miss living here*), while tweeting (e.g., *Nope. Not yet. Still in my car, enjoying traffic*), or after tweeting (e.g., *Can't wait to fly home this summer*). In the third example, it is not guaranteed that *fly* will occur, so one can only say that the author will probably participate in *fly*.

In this paper, we determine whether people participate in the events they tweet about. More specifically, we determine whether they are par-

ticipants before tweeting, while tweeting and after tweeting, and define event participants as people directly involved in an event, regardless of whether they are the agent, recipient or play another role. The main contributions of this paper are: (a) annotations using 1,096 events from 826 tweets; (b) analysis showing that authors of tweets are often not participants in the events they tweet about before or after tweeting; and (c) experimental results showing that the task can be automated.

2 Previous Work

Most previous efforts on detecting events from Twitter focus on events of general importance (e.g., death of a celebrity, natural disasters) or major life events of individuals (e.g. John Doe getting married, having a baby, being promoted).

Extracting events of general importance often includes extracting the entities involved, date and location, and classifying events into classes such as *trial*, *product launch* or *death* (Ritter et al., 2012). Exploiting redundancy in tweets to extract events is common (Zhou et al., 2014), as well as spatio-temporal information (Cheng and Wicks, 2014), i.e., when and where tweets originate from.

Extracting major life events consists on pinpointing significant events from mundane events (e.g., having lunch, exercising) (Di Eugenio et al., 2013; Li et al., 2014; Dickinson et al., 2015), and determining whether significant events are relevant to Twitter users (e.g., *Why doesn't John marry Mary already?* [not relevant to the author]).

Unlike these previous efforts, the work proposed here determines whether people participate in the events they tweet about, and specifies when with respect to tweet timestamps. As a result, we target past events, ongoing events, and events likely to occur in the future. Additionally, we target all events regardless of importance.
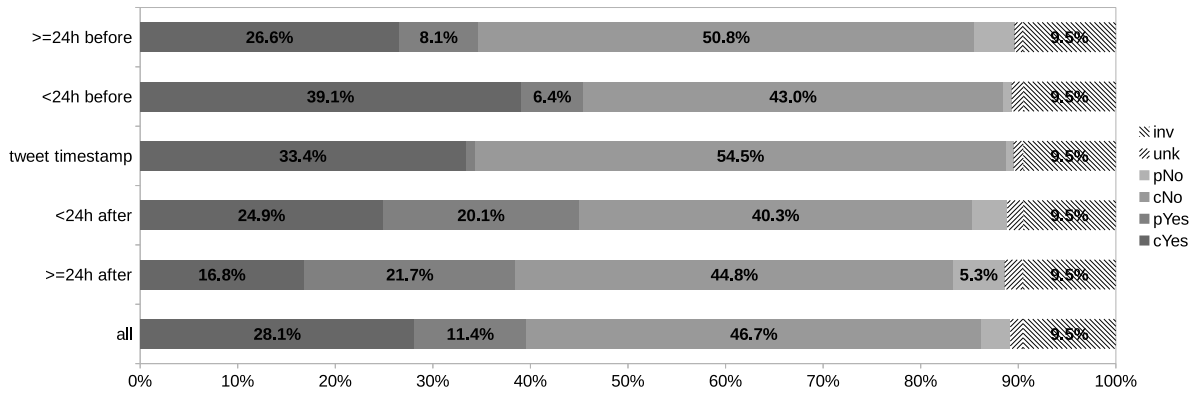
Figure 1: Label distribution per temporal span. Percentages are shown if they are greater or equal than 2%, e.g., percentages for `unk` label are not shown because they range between 0.91% and 1.91%.

3 Corpus Creation

We created a corpus of tweets, events and annotations indicating whether and when the authors of tweets participate in the events as follows.

Selecting Tweets and Events. First, we collected 5,017 tweets from corpora released by previous projects (Owoputi et al., 2013; Kong et al., 2014; Ritter et al., 2011). Second, we run an event detector to tag tokens that are events (Ritter et al., 2012). Third, we selected as events all tokens tagged as events that are verbs. Fourth, we filtered out tweets that did not contain pronouns *I*, *me* or *we* in order to target tweets that are likely to discuss events that involve the author. Finally, we run a dependency parser for tweets (Kong et al., 2014). We decided to work with an automated event detector and parser to experiment with a system that could be deployed in the real world, and discarded events that are nouns because manual inspection revealed that the event detector makes many more mistakes with nouns than with verbs.

The steps above resulted in 1,096 events from 826 tweets. The part-of-speech tags of events are as follows: VBP: 553 (verb, non-3rd person singular present), VBG: 345 (verb, gerund or present participle), VBN: 198 (verb, past participle).

Annotation Process and Quality. For each event in each tweet, we asked annotators "Is the author of the tweet a participant in the *event*?" During pilot annotations with two graduate students, it became clear that a major source of disagreements was due to annotators answering the question for different times with respect to the tweet timestamp. After discussing errors, we decided to ask for five answers: over 24 hours and within 24 hours before tweeting, when tweeting (tweet timestamp), and within 24 hours and over 24 hours after tweeting. Additionally, we allow for six answers partially inspired by previous work on factuality (Sauri and Pustejovsky, 2009):

- `cYes`, `cNo`: I am certain that the author is (or is not) a participant in the event.
- `pYes`, `pNo`: It is probably the case that the author is (or is not) a participant in the event.
- `unk`: the question is intelligible, but none of the four labels above would be correct.
- `inv`: the event at hand is not an event.

The temporal spans and labels were tuned until the two annotators obtained 0.70 Kappa agreement with 10% of selected events. Kappa agreement between 0.60 and 0.80 is considered *substantial* agreement (Landis and Koch, 1977). After tuning, the remaining events were annotated once.[1]

4 Corpus Analysis

In this section, we present a corpus analysis consisting of label distribution per temporal span, label distribution for the top 5 most frequent events, and label distribution per part-of-speech tag.

Figure 1 plots percentages of each label per temporal span. Overall (bottom bar), 28.1% of answers are `cYes`, and 46.7% are `cNo`, i.e., annotators are certain that the author of the tweet is or is not a participant in the event. Percentages for `pYes` and `pNo` are much lower: 11.4% and 2.9%. Regarding time spans, people do not usually tweet about events in which they participate while tweeting (`cYes`: 33.4% vs. `cNo`: 54.5%). People are more likely to tweet about

---

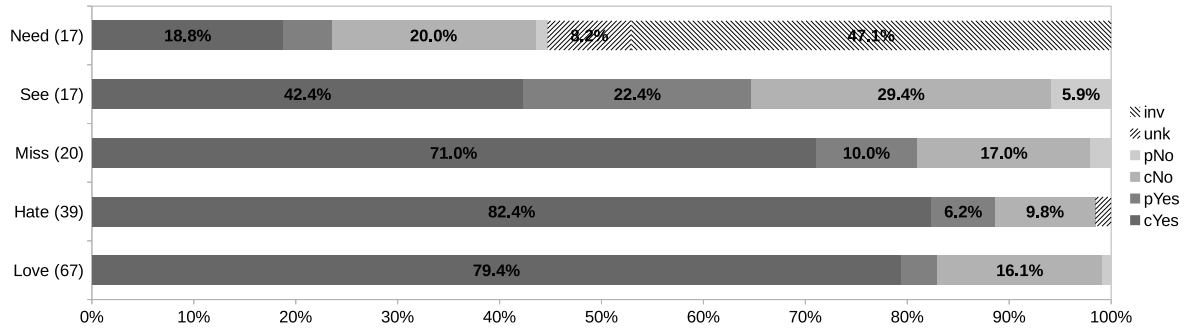[1] Available at `http://www.cse.unt.edu/~blanco/`

Figure 2: Label distribution for the top 5 most frequent events (frequency is shown between parentheses beside the event). Percentages are shown if they are greater or equal than 5%.
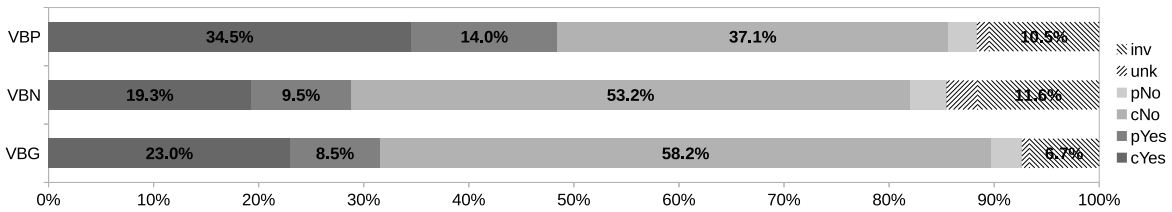

Figure 3: Label distribution per part-of-speech tags of the event. Percentages are shown if they are greater or equal than 5%.

events in which they participated within the last 24 hours (39.1%) than longer before (26.6%), and after tweeting (24.9% and 16.8%). Finally, the percentages of pYes and pNo are below 2% for *tweet timestamp*. Intuitively, it is easier to determine whether somebody participates in the event he tweets about when tweeting rather than before or after. Percentages for labels that do not indicate event participation (unk and inv) are low: unk ranges from 0.91% to 1.91%, and inv is 9.5% for all spans. inv is often used for automatically detected events that are actually states, e.g., *been*.

In Figure 2, we portray the label distribution for the top 5 frequent events and all temporal spans. The top 3 most frequent events (*love*, *hate* and *miss*) have the highest percentages of cYes and pYes labels (> 80% combined), and the fourth most frequent verb (*see*) a lower percentage (cYes: 42.4%, pYes: 22.4%). *Need* has a high percentage of inv (47.1%). This is due to the fact the *need* is often a state (e.g., *Look, I need less friends more bread, less talk, more head*), and asking whether the author is a participant in the event before, during or after the tweet timestamp is nonsensical.

Figure 3 presents the label distribution per part-of-speech tag of the event. VBP (non-3rd person singular present) has the highest percentage of cYes + pYes (48.5% vs. 28.8% and 31.5%), indicating that the author is likely to participate in such events. VBG (gerund or present participle) has highest percentage of cNo + pNo (61.1%), indicating that the author is not likely to participate in those events.

Annotation Examples. Table 1 presents real annotation examples. In Tweet (1), annotators understood that the author has certainly been *addicted to Twitter* for 24 hours before and after tweeting, and probably longer. The author of Tweet (2) was clearly *talking about YOU* before but not after tweeting; annotators indicated that talking most likely occurred within 24 hours before tweeting. The annotations in Tweet (3) indicate that the author was certainly a participant of *selling* when he tweeted and within 24 hours after tweeting, and probably also within 24 hours before and over 24 hours after. Event *seeing* in Tweet (4) may occur *next week*, and annotations capture this information (all cNo except 24h after, which is pYes). Finally, the author of Tweet (5) was never a participant in *scrapbooking* despite she witnessed it.

Note that the annotations also provide information regarding event durations, e.g., *addicted* in Tweet (1) is likely ongoing a day after, but *talking* in Tweet (2) has ended and was a short event.

643

	Tweet	before		tweet	after	
		≥24h	<24h	ts.	<24h	≥24h
1	I'm so <u>addicted</u> to Twitter now that I can tweet all the time. Not good.	pYes	cYes	cYes	cYes	pYes
2	I wonder if you realize we were <u>talking</u> about YOU.	pNo	cYes	cNo	cNo	cNo
3	I'm <u>selling</u> my snorkle, text me for details.	cNo	pYes	cYes	cYes	pYes
4	@Genuine Will I be <u>seeing</u> you at #typeamom next week?	cNo	cNo	cNo	cNo	pYes
5	Today the mall was full of moms who love <u>scrapbooking</u>. kill me.	cNo	cNo	cNo	cNo	cNo

Table 1: Annotation examples. We show annotations for the underlined events only. Recall that events are detected automatically (Section 3). *Tweet ts.* stands for *twitter timestamp*.

Type	No.	Description
Event	1–2	*Event* word form and POS tag
	3	Token number of *event* within the tweet
Situation Entities	7–8	*Event* tense and flag indicating whether *event* tense is perfect tense
	9	Modal type of *event*, if any
	10	Flag indicating whether *event* is a reporting verb
	11	WordNet lexical file name of *event*
Context	12	Position of any pronouns with respect to *event*: left, right, both or none
	13	Position of pronouns *I*, *me*, *we* with respect to *event*: left, right, both or none
	14	Number of outgoing dependencies from *event*
	15	Flag indicating whether there is a dependency between *event* and pronouns *I*, *me*, or *we*

Table 2: Features to determine whether the author of a tweet is a participant in the events he tweets about.

5 Experiments and Results

We follow a standard supervised machine learning approach. We divided the 826 tweets into training and test splits (80% / 20%), and created an instance for each event and temporal span ($1,096 \times 5 = 5,480$ instances). We trained a Support Vector Machine with RBF kernel per temporal span using scikit-learn (Pedregosa et al., 2011) and tuned SVM parameters (C and γ) using 5-fold cross-validation with the training set. We report results when evaluating with the test set.

5.1 Feature Selection

The feature set is presented in Table 2. Most features are fairly simple and well-known, but we borrow some features from previous work on identifying situation entity types (Friedrich et al., 2016) and include features especially designed to capture context around the event we work with.

Event features include the actual event (word form and part-of-speech tag) and number of tokens to left of the event. *Situation Entities* features further characterize the event at hand. To extract them, we first retrieve the clause containing the event by collecting all tokens to the left of the event until we reach a token that is not a verb (including auxiliaries), a modal, or adverb. Friedrich et al. (2016) propose many more features for identifying situation entities, but we selected those that are useful for our task: fine event tense, flag indicating whether the event is in per-

fect tense (past perfect, present perfect, etc.), the type of modals present in the verb clause (if any), whether the event is in a list of 88 reporting verbs, and the WordNet lexical file containing the event (Miller, 1995). Finally, *Context* features mostly indicate the presence of pronouns in the tweet. Specifically, we include the position of any pronoun with respect to the event, and a specialization of this feature for pronouns *I*, *me*, and *we*. We also extract the number of outgoing dependencies from the event, and whether one of those dependencies is between the event and a pronoun *I*, *me* or *we*. Note that the dependency parser for tweets extracts untyped dependencies (Kong et al., 2014), thus we have available information about syntactic dependents but not about specific syntactic relations (nsubj, dobj, auxpass, etc.).

5.2 Experimental Results

Results obtained in the test split with several combinations of features are depicted in Table 3. The baseline always predicts the majority label (cNo for all temporal spans), and is outperformed by all feature combinations.

Feature ablation experiments show (a) little improvement with respect to only training with *Event* features, i.e., the simplest features, and (b) that the optimal combination of features depends on the temporal span for which author participation is being predicted. More specifically, including *Situation Entities* features yields the same results for

	Label	≥24h Before			<24h Before			tweet timestamp			<24h After			≥24h After		
		P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
Baseline	cNo	0.45	1.00	0.62	0.38	1.00	0.55	0.50	1.00	0.67	0.33	1.00	0.49	0.39	1.00	0.56
	others	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Avg.	0.20	0.45	0.28	0.15	0.38	0.21	0.25	0.50	0.34	0.11	0.33	0.16	0.15	0.39	0.22
Event	cYes	0.60	0.51	0.55	0.71	0.71	0.71	0.94	0.69	0.79	0.84	0.59	0.69	0.85	0.69	0.76
	pYes	0.00	0.00	0.00	0.33	0.06	0.10	0.00	0.00	0.00	0.51	0.33	0.40	0.37	0.25	0.30
	cNo	0.56	0.85	0.67	0.54	0.74	0.63	0.69	0.93	0.79	0.45	0.79	0.58	0.48	0.74	0.59
	pNo	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.11	0.12
	unk	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Avg.	0.48	0.56	0.50	0.57	0.61	**0.57**	0.74	0.75	0.73	0.55	0.53	**0.51**	0.50	0.52	0.49
Event + Situation Entities	cYes	0.61	0.52	0.56	0.71	0.69	0.70	0.91	0.73	0.81	0.75	0.60	0.67	0.85	0.71	0.78
	pYes	0.00	0.00	0.00	0.50	0.06	0.11	0.00	0.00	0.00	0.50	0.30	0.37	0.42	0.31	0.36
	cNo	0.57	0.81	0.67	0.56	0.73	0.64	0.72	0.87	0.79	0.46	0.74	0.57	0.49	0.69	0.58
	pNo	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.25	0.11	0.15
	unk	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Avg.	0.47	0.55	0.50	0.58	0.60	**0.57**	0.74	0.74	0.73	0.52	0.51	0.49	0.51	0.53	**0.51**
Event + Situation Entities + Context	cYes	0.71	0.55	0.62	0.69	0.69	0.69	0.91	0.73	0.81	0.73	0.56	0.63	0.85	0.69	0.76
	pYes	0.00	0.00	0.00	0.50	0.06	0.11	0.00	0.00	0.00	0.55	0.40	0.46	0.35	0.31	0.33
	cNo	0.57	0.84	0.68	0.55	0.69	0.61	0.73	0.89	0.80	0.42	0.61	0.49	0.50	0.65	0.56
	pNo	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.11	0.14
	unk	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Avg.	0.50	0.58	**0.52**	0.57	0.59	0.56	0.75	0.76	**0.74**	0.50	0.48	0.48	0.49	0.50	0.49

Table 3: Results obtained with several feature combinations in the test split. *Average* is the weighted average; we do not show results for *invalid* but they are included in the averages.

≥*24h before*, <*24h before* and *tweet timestamp*, slightly worse results for *< 24h after* (F-measures: 0.51 vs. 0.49) and slightly better for ≥*24h after* (F-measures: 0.49 vs. 0.51). Training with all features (Event + Situation Entities + Context), obtains the best results for ≥*24h before* (F-measure: 0.52) and *tweet timestamp* (F-measure: 0.74), but results are slightly lower for the other time spans.

Note that while the results with *Event* features are only outperformed for some temporal spans when training with all features, information beyond the event at hand is needed to solve this task. For example, the correct labels for *I love living in NYC* and *I miss living in NYC* are different (*living* is an ongoing and past event respectively).

6 Conclusions

We have presented a corpus and machine learning models to predict whether people participate in the events they tweet about. More specifically, we determine whether people participate in events when they tweet about them, and also before and after.

Unlike most previous work, we target any event in a tweet regardless of its importance. While major life events (e.g., graduating from college) are arguably more important, we believe that mundane events (e.g., studying for an exam) provide key information to retrieve major life events and predict future events, e.g., *studying for an exam* is (most probably) more likely to lead to *graduating from college* than *going to parties on a regular basis*.

Experimental results show that the task can be automated but is challenging. We believe that features derived from subsequent tweets by the same author and tweet replies would yield better results, and plan to incorporate them in future work.

References

Tao Cheng and Thomas Wicks. 2014. Event detection using twitter: A spatio-temporal approach. *PLOS ONE* 9(6):1–10.

Barbara Di Eugenio, Nick Green, and Rajen Subba. 2013. Detecting life events in feeds from twitter. In *Semantic Computing (ICSC), 2013 IEEE Seventh International Conference on*. Ieee, pages 274–277.

Thomas Dickinson, Miriam Fernandez, Lisa A. Thomas, Paul Mulholland, Pam Briggs, and Harith Alani. 2015. Identifying prominent life events on twitter. In *Proceedings of the 8th International Conference on Knowledge Capture*. ACM, New York, NY, USA, K-CAP 2015, pages 4:1–4:8.

Annemarie Friedrich, Alexis Palmer, and Manfred Pinkal. 2016. Situation entity types: automatic classification of clause-level aspect. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Berlin, Germany, pages 1757–1768. http://www.aclweb.org/anthology/P16-1166.

Lingpeng Kong, Nathan Schneider, Swabha Swayamdipta, Archna Bhatia, Chris Dyer, and

Noah A. Smith. 2014. A dependency parser for tweets. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Doha, Qatar, pages 1001–1012. http://www.aclweb.org/anthology/D14-1108.

J. Richard Landis and Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33(1).

Jiwei Li, Alan Ritter, Claire Cardie, and Eduard Hovy. 2014. Major life event extraction from twitter based on congratulations/condolences speech acts. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Doha, Qatar, pages 1997–2007. http://www.aclweb.org/anthology/D14-1214.

George A. Miller. 1995. Wordnet: A lexical database for english. *Commun. ACM* 38(11):39–41. https://doi.org/10.1145/219717.219748.

Olutobi Owoputi, Brendan O'Connor, Chris Dyer, Kevin Gimpel, Nathan Schneider, and Noah A. Smith. 2013. Improved part-of-speech tagging for online conversational text with word clusters. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Atlanta, Georgia, pages 380–390. http://www.aclweb.org/anthology/N13-1039.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12:2825–2830.

Alan Ritter, Sam Clark, Mausam, and Oren Etzioni. 2011. Named entity recognition in tweets: An experimental study. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Edinburgh, Scotland, UK., pages 1524–1534. http://www.aclweb.org/anthology/D11-1141.

Alan Ritter, Mausam, Oren Etzioni, and Sam Clark. 2012. Open domain event extraction from twitter. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, USA, KDD '12, pages 1104–1112.

Roser Sauri and James Pustejovsky. 2009. Factbank: a corpus annotated with event factuality. *Language Resources and Evaluation* 43(3):227–268.

Deyu Zhou, Liangyu Chen, and Yulan He. 2014. A simple bayesian modelling approach to event extraction from twitter. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Association for Computational Linguistics, Baltimore, Maryland, pages 700–705. http://www.aclweb.org/anthology/P14-2114.