Subjective Well-Being and Social Media: A Semantically Annotated Twitter Corpus on Fertility and Parenthood

Emilio Sulis, Cristina Bosco, Viviana Patti
Dipartimento di Informatica
University of Turin
Italy
{bosco,patti,sulis@di.unito.it}

Mirko Lai
Delia Irazú Hernández Farías
University of Turin, Italy
Univ. Politècnica de València, Spain
{hernande,lai}@unito.it

Letizia Mencarini
Dondena Centre for Research on
Social Dynamics and Public Policy
Bocconi University, Italy
letizia.mencarini@unibocconi.it

Michele Mozzachiodi
Daniele Vignoli
University of Florence, Italy
{mozzachiodi,vignoli}@disia.unifi.it

Abstract

English. This article describes a Twitter corpus of social media contents in the Subjective Well-Being domain. A multi-layered manual annotation for exploring attitudes on fertility and parenthood has been applied. The corpus was further analysed by using sentiment and emotion lexicons in order to highlight relationships between the use of affective language and specific sub-topics in the domain. This analysis is useful to identify features for the development of an automatic tool for sentiment-related classification tasks in this domain. The gold standard is available to the community.

Italiano. *L'articolo descrive la creazione di un corpus tratto da Twitter sui temi del Subjective Well-Being, fertilità e genitorialità. Un'analisi lessicale ha mostrato il legame tra l'uso di linguaggio affettivo e specifiche categorie di messaggi. Questo esame è utile per se e per l'addestramento di sistemi di classificazione automatica sul dominio. Il gold standard è disponibile su richiesta.*

1 Introduction

The key research questions we address in this paper concern how subjective well-being drives fertility trends (and vice versa). We developed a Twitter Italian corpus annotated with a novel semantic annotation scheme for marking information not only about sentiment polarity, but also about the specific semantic areas/sub-topics which are the target of sentiment in the fertility-SWB domain. The relationship between big data and official statistics is increasingly a subject of attention (Mitchell et al., 2013; Reimsbach-Kounatze, 2015; Sulis et al., 2015; Zagheni and Weber, 2005). In this work we focus on Twitter data for two main reasons. First, Twitter individuals' opinions are posted spontaneously (not responding to a question) and often as a reaction to some emotional driven observation. Moreover, using Twitter we can incorporate additional measures of attitudes towards children and parenthood, with a wider geographical coverage than what is the case for traditional survey. Sentiment analysis in Twitter has been also used to monitor political sentiment (Tumasjan et al., 2010), to extract critical information during times of mass emergency (Verma et al., 2011; Buscaldi and Hernández Farías, 2015), or to analyse user stance in political debates on controversial topics (Stranisci et al., 2016; Bosco et al., 2016; Mohammad et al., 2015). A comprehensive overview of sentiment analysis with annotated corpora is offered in (Nissim and Patti, 2016). Focusing on Italian, among the existing resources we mention the Senti-TUT corpus (Bosco et al., 2013) and the TWITA corpus (Basile and Nissim, 2013) that were recently exploited in the SENTIment POLarity Classification (SENTIPOLC) shared task (Basile et al., 2014). The corpus described in this paper enriches the scenario of datasets available for Italian, enabling also a finer grained analysis of sentiment related phenomena in a novel domain related to parenthood and fertility.

2 Dataset and Methodology

As a reference dataset, we adopted all the tweets posted in Italian language in 2014 (TWITA14

henceforth), which were retrieved through the Twitter Streaming API and applying the Italian filter proposed within the TWITA project (Basile and Nissim, 2013). The TWITA14 dataset included 259,893,081 tweets (4,766,342 geotagged). We applied a multi-step methodology in order to filter and select those relevant tweets concerning fertility and parenthood.

2.1 Filtering steps on the dataset

A number of filtering steps have been applied for selecting from TWITA14 a corpus of tweets where users talk about fertility and parenthood (TW-SWELLFER corpus, henceforth). We could not rely on the exploitation of one or few hashtags or other elements that allow identifying posts on fertility and parenthood. In fact, these topics are somehow spread in the dataset and messages may contain relevant information on such subjects even if the main topic of the post is different. Therefore, we are facing a situation where, on the one hand, the set of the data that are potentially relevant for our specific analysis is wider than usual; on the other hand, it is more difficult to identify the presence of information related to the topics we are interested in. This leaded us to adopt a multi-step thematic filtering approach. In a first step (*Keyword-based filtering step*), eleven hashtags[1] and other 19 keywords have been chosen for selecting tweets of interest, including 8 roots to consider diminutives, singulars and plurals. This list is the result of a combination of a manual content analysis on 2,500 tweets sampled at completely random (taken as a starting point) and a linguistic analysis on synonyms. We obtain a total amount of 3.9 million tweets. A second filtering step consisted in removing noisy tweets from corpus (*User-based filtering step*), as the off-topic ones (messages not concerning individual expression on fertility and parenthood topics). Tweets posted by company/institutions/newspapers accounts have been deleted. Finally, duplicated tweets not marked as RT were deleted (*Duplicate-based filtering step*). The resulting TW-SWELLFER corpus consists of 2,760,416 tweets.

[1] #papà, #mamma, #babbo, #incinta, #primofiglio, #secondofiglio, #futuremamme, #maternità, #paternità, #allattamento, #gravidanza

2.2 Annotation scheme

Given the TW-SWELLFER dataset, we developed and applied an annotation model aimed at studying not only the sentiment expressed in the tweets, but also specific parenthood-related topics discussed in Twitter that are the target of the sentiment. To build our annotation model, we relied on a standard annotation scheme on sentiment polarity (POLARITY), by exploiting the same labels POS, NEG, NONE and MIXED provided the organizers of the shared task for sentiment analysis in Twitter for Italian (Basile et al., 2014). Also the presence/absence of irony has been marked in order to be able to reason on sentiment polarity also in case of use of figurative devices. Annotating the presence of ironic devices is a challenging task because the inferring process of this figure of speech does not always lie on semantic and syntactic elements of texts (Ghosh et al., 2015; Reyes et al., 2013; Hernández Farías et al., 2016), but often requires contextual knowledge (Wilson, 2006). In order to mark irony, we introduced two polarized ironic labels: HUMNEG, for ironic tweets with negative polarity, and HUMPOS for ironic tweets with positive polarity. Finally, a set of labels marks the specific semantic areas (or SUBTOPICS) of the tweets related to the parenthood domain. This part of the annotation scheme is very important since somehow provides us with a semantic grid in order to analyse which are the aspects of parenthood that are discussed on Twitter. For the annotation of sub-topics we considered 7 labels, suggested by a group three experts on the SWELLFER (subjective well-being and fertility) domain, after a manual analysis of a subset of the tweets:

- TOBEPA - *To be parents*. This tag is introduced to mark when the user generically comments about his status of parent.
- TOBESO - *To be sons*. This tag marks the sons point of view, i.e. on when the user is a son that comments on the parent-son relationship.
- DAILYLIFE - *Daily life*. This tag marks messages commenting on recurring situation in everyday life for what concerns the relationship between parents and children.
- JUDGOTHERPA - *Judgment over other parents behaviour*. The tag allows to mark

comments on educations of children, for instance comments of behaviours which does not seems to be appropri - ated for the parent role.

- FUTURE - *Children' future*. This tag is used for tweets where parents do express sentiments about the future of children.
- BECOMPA - *To become parents*. This tag is introduced to mark tweets where users speak about the prospect or fear of being parents.
- POL - *Political side*. This tag is introduced to mark tweets talking about laws having impact on being parents.

Finally, two additional tags (IN-TOPIC/OFF-TOPIC) have been added to allow annotators to mark if the tweet is relevant. The addition of this tag was necessary because of the noise still present in the dataset. Moreover, in this way, the manual annotation will produce also data to be used in order to create a supervised topic classifier from the whole TW-SWELLFER corpus. This opens the way to the exploitation of the corpus for a fine-grained sentiment analysis, by identifying different aspects and topics of the Twitter debate on parenthood and the sentiment expressed towards each aspect/topic.

2.3 Manual annotation

A random sample of 5,566 tweets from TW-SWELLFER has been collected. On this sample we applied crowdsourcing for manual annotation via the Crowdflower platform already used in literature (Nakov et al., 2016). We relied on CrowdFlower controls to exclude unreliable annotators and spammers based on hidden tests, which we created by developing a set of gold-standard test questions equipped with gold reasons[2]. The annotator's task was, first, to mark if the post is IN- or OFF-TOPIC (or unintelligible), and then to mark for IN-TOPIC posts, on the one hand, the polarity and presence of irony, on the other hand, the subtopics. Precise guidelines were provided to the annotators. Overall, for each tweet at least three independent annotations were provided[3]. In order to select the true label we used majority voting.

In-topic vs off-topic: manual annotation on this aspect resulted in 2,355 in-topic tweets (42.3%) and 3,136 off-topic (56.3%); the remaining 75 tweets were unknown or null (cases of disagreement). Thanks to the preliminary filtering steps, the proportion of in-topic tweets is pretty high compared to common results from different Twitter based content and opinion analysis (Ceron et al., 2014).

Polarity, irony, sub-topics (in-topic tweets): at the end of the manual annotation process we collected 1,545 labeled with the same tags for all the layers.

Notice that in the analysis in the next section will report results also on tweets labeled as IN-TOPIC after the manual annotation (2,355), but where annotators did not agree on the polarity, irony and subtopics labels. We refer to those tweets as NULL messages.

Summarizing, the TWSWELLFER-GOLD corpus includes 1,545 IN-TOPIC tweets labeled with the same tags for all the layers (POLARITY, IRONY and SUBTOPICS).

3 Analysis of the gold standard

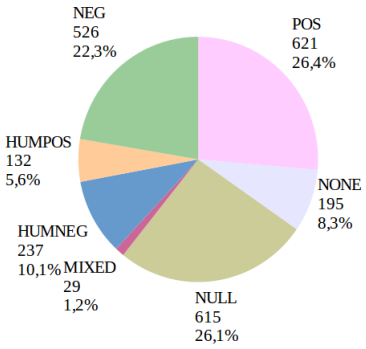

Figure 1: TW-SWELLFER: Distribution of polarity tags in IN-TOPIC messages

Regarding IN-TOPIC tweets, the 26.4% has been labeled as positive and 22.3% as negative (See Fig.1), giving us a guidance on what might be the general feeling in Twitter about the research topics on happiness and parenthood. The irony issue is limited to a 15.7% of all the messages and negative irony prevails (10.1% of negative ironic tweets and 5.6% of positive ironic tweets), while neutral tweets are just the 8.3%.

[2]Test questions resulted from the agreement of three expert annotators.

[3]We selected the CrowdFlower's *dynamic judgment* option: having the goal of collecting at least 3 reliable annotations for each tweet, the system was collecting up to a maximum of 5 annotations (to deal with cases when row's confidence score is low). In our jobs we set 0.7 as minimum accuracy threshold.

The amount of mixed tweets is limited to 1.2% (remaining 26% are labelled as NULL because of annotators disagreement). Regarding these results, it appears that positive and negative feelings towards family, parenthood and fertility appear more or less equally spread through Twitter Italy. Even if the positive posts are a little bit more than the negative ones, ironic tweets must be considered: most of them are negative ironic posts (i.e., insulting/damaging the target) balancing the slight difference between pure positive and negative tweets. Furthermore, this particular topic, combined with the Twitter nature which provides short direct message, discourages people to stand in the grey (neutral) area, as could happens in other cases: about the 90% of the tweets shows an explicit polarity, meaning people take a side and express their opinions.

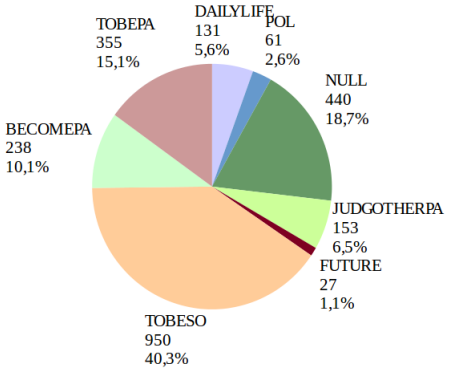

Figure 2: TW-SWELLFER: Distribution of subtopic tags in IN-TOPIC messages

Which are these opinions and about what? Going further with the analysis and looking also at the contents, so taking into consideration the "topic specification attribute and its values (Fig. 2), the largest category refers to sons tweets (TOBESO) (40.3%), in which children are discussing and posting about being children and/or about relating themselves with parents. Parents tag (TOBEPA) settles on 15% and becoming tag (BECOMEPA) on 10%. Remaining categories have minor impact, all being in between 1% and 6%.

3.1 Sentiment and emotion analysis

The exam of the corpus includes a lexical analysis on different aspects of affect: sentiment and emotions. The distribution of terms in each group of messages reveals interesting patterns. Adopting sentiment lexical resources[4] the whole polarity of messages is computed summing positive and negative terms. A normalization is finally performed, i.e. dividing the polarity value by the number of terms in each group. In particular, the four lexica considered count more positive terms in positive messages. Similarly, negative terms are more frequent in negative messages. Ironic messages reveal a similar pattern, even if smoothed. Table 1 presents some of these results.

In addition, the emotion lexicon indicates a larger frequency of terms related to anger, sadness, fear and disgust in negative messages than in positive ones (See Fig. 3). On the contrary, positive

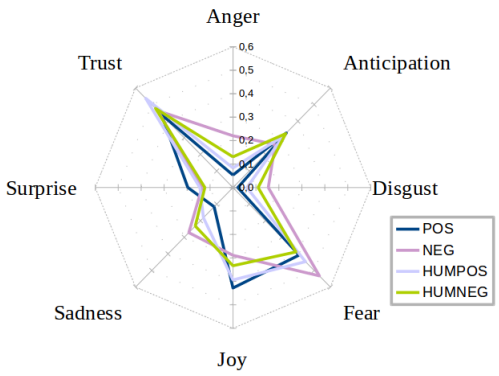

Figure 3: Distribution of emotions by polarity tags

messages contain more terms related to joy, anticipation and surprise. Some suggestions can be derived in the comparison of polarity categories and the corresponding ironic ones. For instance, terms related to joy are more frequent in ironic negative messages than in negative ones. It is an insight of the polarity reversal phenomena, where a shift is produced by the adoption of a seemingly positive statement, to reflect a negative one (Sulis et al., 2016).

The analysis of topic specification messages reveals a positive polarity for messages concerning TOBEPA (to be parents), while BECOMEPA (to become parents) has a more negative polarity (See Table 1). Focusing on emotion lexicon, TOBEPA has an higher incidence of Joy words (Fig. 4).

Messages concerning educations of children (JUDGOTHERPA) contain a high frequency of anger and disgust term. The category TOBESO (to

[4]EmoLex (Mohammad and Turney, 2013) as well as an own-house Italian version of LIWC (Pennebaker et al., 2001), Hu&Liu (Hu and Liu, 2004), AFINN (Nielsen, 2011). Lexicons were translated from English in (Buscaldi and Hernández Farías, 2015).

tag	pLIWC	pHuLiu	pEmolex	Afinn	pAVG
POS	1.06	0.22	0.62	3.51	1.35
NEG	-1.61	0.04	0.12	0.39	-0.27
HUMPOS	0.19	0.12	0.23	2.29	-0.71
HUMNEG	-0.34	0.08	0.64	0.61	0.25
TOBESO	1.97	0.88	0.02	1.56	1.11
TOBECOMEPA	1.5	0.73	0.18	-1.64	0.19
TOBEPA	1.94	1.38	0.18	5.04	2.13
DAILYLIFE	1.13	1.56	0.32	6.04	2.26

Table 1: Polarity values according to different lexicons in tweets tagged with the following labels: POS, NEG, HUMPOS, HUMNEG (polarity tags) and TOBESO, TOBECOMEPA,TOBEPA, DAILYLIFE (sub-topic tags).

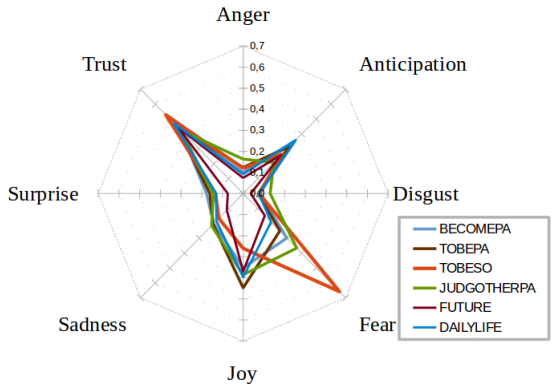

Figure 4: Distribution of emotions by sub-topic tags

be sons) is more controversial, having the higher frequency of negative terms as fear, but also trust, as well as having the lower frequency of Joy terms. Coherently, anticipation is more frequent in the BECOMEPA group of messages. Summarizing, it seems that children are more critics toward parents. On the contrary, parents seem express an attitude more positive towards children.

4 Conclusions and Future Work

The contribution of this paper is the exploration of opinions and semantic orientation about fertility and parenthood by scrutinizing about 3 million Italian tweets. This analysis is useful to identify features for the development of an automatic system to address automatic classification tasks in this domain. The corpus is available to the community. Its development constitutes a first step and a precondition to a further analysis that can be applied on such contents in order to extract, from semantically enriched data, measures of SWB constructed in an indirect way. This will hopefully improve our understanding of attitudes on fertility and parenthood.

We are currently extending the corpus by exploring the very interesting debate around the "Fertility Day's initiative" from the Italy's Minister of Health Beatrice Lorenzin, which had a remarkable echo on social media such as Twitter, with a substantial number of (also sarcastic) messages with hashtag #fertilityday posted.

Acknowledgments

The authors gratefully acknowledge financial support from the European Research Council under the European ERC Grant Agreement n. StG-313617 (SWELL-FER: Subjective Well-being and Fertility, P.I. Letizia Mencarini).

References

Valerio Basile and Malvina Nissim. 2013. Sentiment analysis on italian tweets. In *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 100–107, Atlanta, Georgia. Association for Computational Linguistics.

Valerio Basile, Andrea Bolioli, Malvina Nissim, Viviana Patti, and Paolo Rosso. 2014. Overview of the Evalita 2014 SENTIment POLarity Classification Task. In *Proc. of EVALITA 2014*, pages 50–57, Pisa, Italy. Pisa University Press.

Cristina Bosco, Viviana Patti, and Andrea Bolioli. 2013. Developing corpora for sentiment analysis: The case of irony and senti-tut. *IEEE Intelligent Systems*, 28(2):55–63, March.

Cristina Bosco, Mirko Lai, Viviana Patti, and Daniela Virone. 2016. Tweeting and being ironic in the debate about a political reform: the french annotated corpus twitter-mariagepourtous. In *Proceedings of*

the Tenth International Conference on Language Resources and Evaluation (LREC 2016), pages 1619–1626, Portoroz, Slovenia. ELRA.

Davide Buscaldi and Delia Irazú Hernández Farías. 2015. Sentiment analysis on microblogs for natural disasters management: A study on the 2014 genoa floodings. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, pages 1185–1188. ACM.

A. Ceron, L. Curini, and S.M. Iacus. 2014. *Social Media e Sentiment Analysis: L'evoluzione dei fenomeni sociali attraverso la Rete*. SxI - Springer for Innovation / SxI - Springer per l'Innovazione. Springer Milan.

Aniruddha Ghosh, Guofu Li, Tony Veale, Paolo Rosso, Ekaterina Shutova, Antonio Reyes, and Jhon Barnden. 2015. Semeval-2015 task 11: Sentiment analysis of figurative language in Twitter. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 470–475, Denver, Colorado, USA. Association for Computational Linguistics.

Delia Irazú Hernández Farías, Viviana Patti, and Paolo Rosso. 2016. Irony detection in Twitter: The role of affective content. *ACM Transaction of Internet Technology*, 16(3):19:1–19:24.

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, pages 168–177, Seattle, WA, USA. ACM.

L. Mitchell, M. R. Frank, K. D. Harris, P. S. Dodds, and C. M. Danforth. 2013. The geography of happiness: Connecting Twitter sentiment and expression, demographics, and objective characteristics of place. *PLoS ONE*, 8(5), 05.

Saif M. Mohammad and Peter D. Turney. 2013. Crowdsourcing a Word–Emotion Association Lexicon. *Computational Intelligence*, 29(3):436–465.

Saif M. Mohammad, Xiaodan Zhu, Svetlana Kiritchenko, and Joel Martin. 2015. Sentiment, emotion, purpose, and style in electoral tweets. *Information Processing and Management*, 51:480–499.

Preslav Nakov, Alan Ritter, Sara Rosenthal, Fabrizio Sebastiani, and Veselin Stoyanov. 2016. Semeval-2016 task 4: Sentiment analysis in twitter. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 1–18, San Diego, California, June. Association for Computational Linguistics.

Finn Årup Nielsen. 2011. A new ANEW: evaluation of a word list for sentiment analysis in microblogs. In *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages*, volume 718 of *CEUR Workshop Proceedings*, pages 93–98, Heraklion, Crete, Greece. CEUR-WS.org.

Malvina Nissim and Viviana Patti. 2016. Semantic aspects in sentiment analysis. In Fersini Elisabetta, Bing Liu, Enza Messina, and Federico Pozzi, editors, *Sentiment Analysis in Social Networks*, chapter 3, pages 31–48. Elsevier.

James W. Pennebaker, Martha E. Francis, and Roger J. Booth. 2001. Linguistic Inquiry and Word Count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, 71.

Christian Reimsbach-Kounatze. 2015. The proliferation of big data and implications for official statistics and statistical agencies.

Antonio Reyes, Paolo Rosso, and Tony Veale. 2013. A multidimensional approach for detecting irony in twitter. *Lang. Resour. Eval.*, 47(1):239–268, March.

Marco Stranisci, Cristina Bosco, Delia Irazú Hernández Farías, and Viviana Patti. 2016. Annotating sentiment and irony in the online italian political debate on #labuonascuola. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France, may. European Language Resources Association (ELRA).

Emilio Sulis, Mirko Lai, Manuela Vinai, and Manuela Sanguinetti. 2015. Exploring sentiment in social media and official statistics: a general framework. In *Proceedings of the 2nd International Workshop on Emotion and Sentiment in Social and Expressive Media, co-located with AAMAS 2015, Istanbul, Turkey, May 5, 2015.*, volume 1351 of *CEUR Workshop Proceedings*, pages 96–105. CEUR-WS.org.

Emilio Sulis, Irazú Hernández Farías, Paolo Rosso, Viviana Patti, and Giancarlo Ruffo. 2016. Figurative messages and affect in twitter: Differences between #irony, #sarcasm and #not. *Knowledge-Based Systems*, 108:132 – 143. New Avenues in Knowledge Bases for Natural Language Processing.

Andranik Tumasjan, Timm Sprenger, Philipp Sandner, and Isabell Welpe. 2010. Predicting elections with Twitter: What 140 characters reveal about political sentiment. In *International AAAI Conference on Web and Social Media*.

Sudha Verma, Sarah Vieweg, William Corvey, Leysia Palen, James Martin, Martha Palmer, Aaron Schram, and Kenneth Anderson. 2011. Natural language processing to the rescue? extracting "situational awareness" tweets during mass emergency. In *International AAAI Conference on Web and Social Media*.

Deirdre Wilson. 2006. The pragmatics of verbal irony: Echo or pretence? *Lingua*, 116(10):1722 – 1743.

Emilio Zagheni and Ingmar Weber. 2005. Demographic research with non-representative internet data. *International Journal of Manpower*, 36(1):13–25.