



DEMOGRAPHIC RESEARCH

A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

VOLUME 40, ARTICLE 25 PAGES 693–724

PUBLISHED 20 MARCH 2019

<http://www.demographic-research.org/Volumes/Vol40/25/>

DOI: 10.4054/DemRes.2019.40.25

Research Article

Happy parents' tweets: An exploration of Italian Twitter data using sentiment analysis

Letizia Mencarini

Viviana Patti

Delia Irazú Hernández-Farías

Emilio Sulis

Mirko Lai

Daniele Vignoli

This publication is part of the Special Collection on “Social Media and Demographic Research,” organized by Guest Editor Emilio Zagheni.

© 2019 Letizia Mencarini et al.

This open-access work is published under the terms of the Creative Commons Attribution 3.0 Germany (CC BY 3.0 DE), which permits use, reproduction, and distribution in any medium, provided the original author(s) and source are given credit.

See <https://creativecommons.org/licenses/by/3.0/de/legalcode>

Contents

1	Introduction	694
2	Related work	696
3	Developing a data set (corpus) for exploring attitudes towards fertility and parenthood	697
3.1	The collection and filtering of relevant data	698
3.2	Manual annotation criteria for exploring sentiment and irony in parenthood-related topics	699
3.3	Annotation process with CrowdFlower	705
3.4	Analysis of the 'gold standard' Twitter corpus	706
4	Beyond the polarity valence: a lexical analysis based on an emotion lexicon	709
5	Automatic detection of sentiment polarity	710
6	The geographical distribution of positive messages	712
7	Conclusions	713
8	Acknowledgments	715
	References	716

Happy parents' tweets: An exploration of Italian Twitter data using sentiment analysis

Letizia Mencarini¹

Delia Irazú Hernández-Farías²

Mirko Lai³

Viviana Patti³

Emilio Sulis³

Daniele Vignoli⁴

Abstract

BACKGROUND

Demographers are increasingly interested in connecting demographic behaviour and trends with 'soft' measures, i.e., complementary information on attitudes, values, feelings, and intentions.

OBJECTIVE

The aim of this paper is to demonstrate how computational linguistic techniques can be used to explore opinions and semantic orientations related to parenthood.

METHODS

In this article we scrutinize about three million filtered Italian tweets from 2014. First, we implement a methodological framework relying on Natural Language Processing techniques for text analysis, which is used to extract sentiments. We then run a supervised machine-learning experiment on the overall dataset, based on the annotated set of tweets from the previous stage. Consequently, we infer to what extent social media users report negative or positive affect on topics relevant to the fertility domain.

RESULTS

Parents express a generally positive attitude towards being and becoming parents, but they are also fearful, surprised, and sad. They also have quite negative sentiments about their children's future, politics, fertility, and parental behaviour. By exploiting geographical information from tweets we find a significant correlation between the

¹ Dondeña Centre for Research on Social Dynamics and Public Policy, Bocconi University, Milan, Italy.

² INAOE (Instituto Nacional de Astrofísica, Óptica y Electrónica), Puebla, Mexico.

³ Computer Science Department, University of Turin, Italy.

⁴ Department of Statistics, Computer Science, Applications (DiSIA), University of Florence, Italy.

prevalence of positive sentiments about parenthood and macro-regional indicators of both life satisfaction and fertility level.

CONTRIBUTION

We show how tweets can be used to represent soft measures such as attitudes, values, and feelings, and we establish how they relate to demographic features. Linguistic analysis of social media data provides a middle ground between qualitative studies and more standard quantitative approaches.

1. Introduction

Rapid increases in computational power and storage capabilities (Hilbert and López 2011) have radically transformed human communications and societies (Castells 2000). The massive dissemination of information heralds a new era in social studies that brings new research challenges and opportunities (King 2011; Lazer et al. 2009; Aggarwal 2013). This holds true not least for demographic analyses. For example, migrants have been tracked using email data (Zagheni and Weber 2012); migrant stocks have been monitored using Facebook data (Zagheni, Weber, and Gummadi 2017); patterns of short- and long-term migration using Twitter data (Zagheni, Garimella, and Weber 2014); fertility patterns using Google search data (Billari, D'Amuri, and Marcucci 2013); and family change using Twitter data (Billari et al. 2017). But demographers are also interested in connecting demographic behaviour and trends with 'soft' measures, i.e., complementary information on attitudes, values, feelings and intentions. Soft measures play a central role in key theoretical approaches to explaining demographic change, the prime example being the Second Demographic Transition Theory (Van de Kaa 1987; Lesthaeghe 2010). Social media data has great potential in this respect, since it typically contains written text statements. However, as Twitter and Facebook texts are invariably disordered it also raises tremendous challenges; the texts do not provide the same structured measures as, say, survey questionnaires. Still, the advantage of social media data is obvious, since it is continuously produced and is now becoming available for almost all countries, even those where traditional survey data is unavailable. However, demographers who are interested in linking demographic trends and behaviour with soft measures have to pay close attention to defining the meaning of text statements – also known as the annotation process. Often these analyses are quite crude. The number of positively and negatively loaded words are counted and then compared with keywords representing the demographic phenomenon of interest. However, as the concept of interest becomes more complex, semantic analysis becomes more challenging.

The aim of this paper is to demonstrate how computational linguistic techniques can be used to analyse the relationship between a demographic feature and what we refer to as ‘soft measures’. Specifically, we explore opinions and semantic orientations related to fertility and parenthood. This application lends itself to both the burgeoning fertility and parenthood literature and the literature on subjective well-being. There are longstanding academic and non-academic debates about the role children play in parents’ daily lives and parents’ subjective well-being. These range from pure qualitative analysis, such as the book *All Joy and No Fun* (2015) by the award-winning journalist Jennifer Senior, to the more traditional data-driven approaches we have seen in demography (e.g., Kohler, Behrman, Skytthe 2005; Clark et al. 2008; Margolis and Myrskylä 2011; Myrskylä and Margolis 2014). While data-driven studies on the dynamics that link subjective well-being and childbearing provide important information from a quantitative point of view (see Kohler and Mencarini 2016 for a review), they can only provide limited insight into opinions on and emotional attitudes toward fertility choices and parenthood. In addition, most of our knowledge, derived from statistical analysis of survey data, points to a ‘parenthood happiness paradox’. Even in low fertility countries, ‘folk’ beliefs have it that children bring happiness. These folk beliefs are in contrast to recent empirical literature on this topic, which finds that the birth of a child typically has a negative effect on the subjective well-being of parents (Hansen 2012; Cetre, Clark, and Senik 2016; Kohler and Mencarini 2016). In this context, social media data provides a middle ground between the qualitative and the standard quantitative approaches by providing evidence of how people talk spontaneously about parenthood and children.

The approach we present consists of two steps. We first implement a Natural Language Processing (NLP) pipeline, which is a set of modules where the output of one feeds into the next. In this stage, selected tweets are analysed to highlight the relationship between the use of affective language and the subtopics of interest. This step sheds lights on the social media content of messages related to fertility domains. The end product of this phase is known as a ‘gold standard corpus’ about parenthood, which is essentially a body of trustworthy texts used for training and meaningful evaluation in the next stage. The second phase consists of a supervised machine-learning experiment carried out by using a model trained with the annotated tweets resulting from stage one. Employing NLP algorithms, messages concerning children, parenthood, and fertility (‘on-topic’) are distinguished from others (‘off-topic’). For on-topic tweets we also set out to detect related subtopics and the sentiment polarity. In this way we infer the extent to which social media users report negative or positive affect on topics relevant to the fertility domain. The prevalence of positive tweets is then correlated with relevant regional characteristics regarding fertility.

Our data is derived from tweets in Italian. There is currently no up-to-date survey data on individual subjective well-being that can be connected to childbearing and parenthood for Italy: thus the potential value of this material is huge.

2. Related work

Sociodemographic research has already benefited from complex – and large – data sources,⁵ thanks, above all, to the ubiquity and widespread use of new technologies (Reimsbach-Kounatze 2015; Zagheni and Weber 2015; Sulis et al. 2015). For example, mobile phone usage has been employed to estimate demographic indicators (Deville et al. 2014), the distribution of the population and demographic structure of a country (Blumenstock, Gillick, and Eagle 2010), and administrative areas (Sobolevsky et al. 2013). Data on Internet searches is helpful in studying fertility (Billari, D'Amuri, and Marcucci 2013), abortion rates (Reis and Brownstein 2010), and union and marriage formation (Hitsch, Hortaçsu, and Ariely 2010). In addition, online social media like Twitter has been used to study migration patterns (Zagheni, Garimella, and Weber 2014) and post-partum changes (De Choudhury et al. 2013).

Sentiment analysis is defined as “the computational study of opinions, sentiments and emotions expressed in text” (Liu 2010). It has become relevant to Natural Language Processing, especially with respect to the study of new forms of digital and social communication (Meo and Sulis 2017). There are several examples of sentiment analysis in political science and sociology. For instance, sentiment analysis of Twitter has been used to monitor political opinions (Tumasjan et al. 2011), to analyse user stances in social media debates (Stranisci et al. 2016; Lai et al. 2015; Mohammad et al. 2015), and to extract critical information during mass emergencies (Verma et al. 2011; Buscaldi and Hernández-Farías 2015). Examples from the social sciences include estimations of subjective well-being, and such sentiment analysis has helped derive measures of happiness within economics, complementing more traditional measures of well-being such as Gross Domestic Product (Diener 2000). Twitter data has also been used to detect moods and happiness in a given geographical area by extracting sentiments (Mitchell et al. 2013; Allisio et al. 2013). Others have used these methods to look for correlations between mood and traditional economic indicators (Bollen and Mao 2011), or to attempt to measure the well-being of a given population (Quercia et al. 2012).

Sentiment analysis relies on annotated datasets or ‘sentiment lexica’: dictionaries or word lists labelled according to sentiment polarity (Nissim and Patti 2016).

⁵ Big data is a term for data sets that too large or complex for traditional application software to deal with them. Big data sources are (as the name suggests) repositories of large volumes of data.

However, in most cases sentiments are estimated through simple word counting, which is either positively or negatively loaded. As researchers seek to use social media data to answer more complex research questions the demands made on sentiment analysis have become more onerous. Computational linguistic analysis provides a possible way to integrate micro theory into the demographic analysis of social media data (Mencarini 2018).

Key theoretical contributions in demography look to ‘soft’ measures as drivers of family change. One example is the Second Demographic Transition, where new demographic behaviour is argued to be a function of changing values: With the onset of modernization, individuals care more about self-realization and less about traditional family life (Van de Kaa 1987; Lesthaeghe 2010). Another example concerns gender equality and equity, where perceived fairness across genders affects fertility (McDonald 2013).

Measuring such concepts through social media data is clearly a challenge and tweet sets annotated for sentiment analysis and opinion mining become an indispensable resource for secondary analysis. In our case, for instance, machine learning is used to make classifications. Not surprisingly, most applications of this kind are based on English. Italian is used much less frequently (a few examples are Bosco, Patti, and Bolioli 2013; Bosco et al. 2014; Bosco, Patti, and Bolioli 2015; Barbieri et al. 2016), although there has been some evaluation of Italian NLP tools and resources (Attardi et al. 2015; Basile et al. 2016).

3. Developing a data set (corpus) for exploring attitudes towards fertility and parenthood

In this study we are interested in fertility and parenthood and the way these relate to individuals’ emotions. The study is therefore relevant to previous social-media-based studies concerned with subjective well-being, but we are more specific, looking at how social media relates to parenthood. This section describes the data collection and the annotation process. Annotation is a key challenge whenever a new theme is considered and is a crucial step, whatever the topic being analysed.

3.1 The collection and filtering of relevant data

We extracted a set of messages (referred to in linguistics as a 'corpus'⁶) from Twitter for the domain of interest. We used the Twita-2014⁷ dataset, consisting of 259,893,081 Italian-language tweets (of which 4,766,342 had been geotagged). In order to assess its representativeness we computed the correlation between the number of tweets for each Italian province (of which there are 110) and the total resident population as measured by the Italian Office for National Statistics (Istat).⁸ The correlation was estimated to be 0.93, suggesting a geographical distribution of tweets consistent with the actual population size: thus the geographical distribution was quite even by administrative region. It is well known that Twitter users in general are not representative of the overall population, as they tend to come from the younger age-strata of the population (Mitchell et al. 2013), which was also the case in our data. However, there appears to be very little difference in age-structure across provinces. In other words, it is unlikely that the computed correlation between number of tweets and population size is distorted by variation in the proportion of young people in the general population in different provinces.

Next we filtered the data set Twita-2014 to select a subsample of tweets where users talk about the topics of interest. Data filtering exploits hashtags⁹ and keywords in order to select relevant tweets. One common drawback with this method is that the topics of interest will frequently be found in tweets where the main topic of the post is different. Thus the amount of data that is potentially relevant to our specific analysis is wider than can be deciphered through a limited set of hashtags and keywords. In order to overcome this, we followed a two-step approach.

In a first keyword-based filtering step, the inflection (diminutives, singulars, and plurals) of eleven hashtags and keywords¹⁰ were used to select tweets of interest. First,

⁶ In linguistics, a 'corpus' (plural 'corpora') is a large and structured set of texts collected to perform linguistics analysis. We used a corpus in order to apply natural language processing and machine learning experiments.

⁷ Twita-2014 was gathered using the Twitter Streaming Application Programming Interface, as described in Basile and Nissim (2013).

⁸ The 'datetime' of tweets refers to 2014 and Istat data used here refers to the population between 1.1.2014 and 31.12.2014.

⁹ Hashtag is a type of metadata tag used in social network and micro-blogging services. It allows users to apply dynamic, user-generated tagging that makes it possible for others to easily find messages with a specific theme or content. It allows easy informal markups of folk taxonomy without needing any formal taxonomy or markup language. Users create and use hashtags by placing the number sign or pound sign # (colloquially known as the hash character) in front of a string of alphanumeric characters. The hashtag may contain letters, digits, and underscores. Searching for that hashtag will yield each message that has been tagged with it.

¹⁰ Namely, *papà*, *mamma*, *babbo*, *incinta*, *#primofiglio*, *#secondofiglio*, *#futuremamme*, *maternità*, *paternità*, *allattamento*, *gravidanza* (in English: father, mother, dad, pregnant, first child, second child, expectant mums, maternity, paternity, breast-feeding, pregnancy).

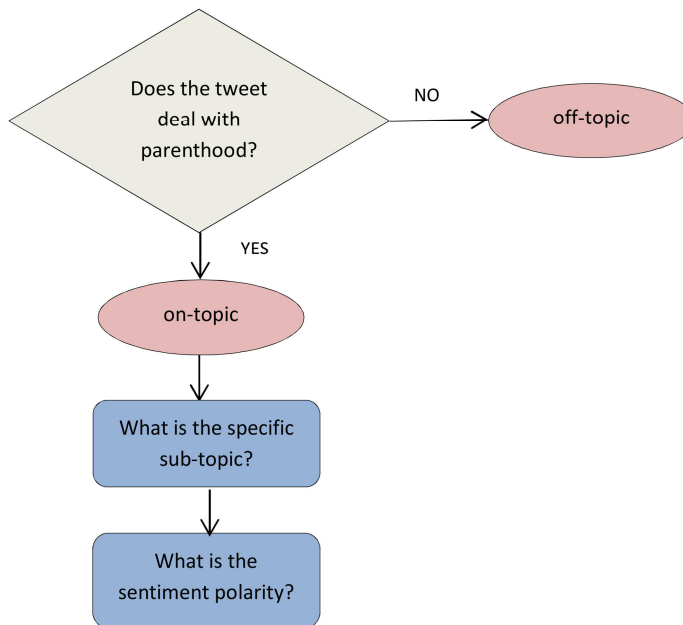
a list of very general Italian keywords were chosen from the *Vocabolario di base della lingua italiana* (VdB) by the linguist Tullio De Mauro,¹¹ all of which were related to the topic of parenthood (e.g., mamma, papà, maternità, figlio, famiglia, incinta). They were selected jointly by a group of linguists and domain experts (demographers). Out of these we randomly chose and manually scrutinized 2,500 tweets. Based on this analysis we selected more keywords (like ‘paternità’), and added them and frequently used hashtags marking Twitter comments relevant to our topic (like *#primofiglio* *#secondofiglio*, *#futuremamme*) to those provided by the VdB. By applying this keyword-based filtering the data set grew to about 3.9 million tweets, all taken from the original Twita-2014 dataset. In the second user-based filtering step we removed ‘noisy’ tweets from the corpus. We defined a ‘noisy’ tweet as a message lacking individual views on fertility and parenthood. This removed all tweets sent from company, institutional, and newspaper accounts. We identified the 500 most prolific Twitter users in Twita-2014, relying on available tweet metadata. By manual inspection of the resulting list of profiles we were able to detect those belonging to online newspapers and news websites, all of which were removed from the corpus. Finally, an automatic duplicate-based filtering step allowed us to delete most advertisements relating to fertility by removing spam tweets, re-tweets, and other duplicated tweets not explicitly marked as re-tweets (having duplicate texts is not interesting for the linguistic analysis of a corpus). After these steps, about 2.8 million tweets remained in the new corpus (henceforth referred to as Twita-2014-parenthood).

3.2 Manual annotation criteria for exploring sentiment and irony in parenthood-related topics

To create a gold corpus with semantic annotation about parenthood, we developed a multi-layered annotation scheme. The scheme is illustrated in Figure 1.

¹¹ It consists of a set of Italian words most commonly used and understood by native speakers, has recently been newly released, and is publicly available here: <https://www.dropbox.com/s/mkcyo53m15ktbnp/nuovovocabolariodibase.pdf>

Figure 1: Multi-layer annotation scheme



This scheme has the benefit of generating tweets that are annotated for both sentiment and subtopics related to parenthood, opening the way for a fine-grained sentiment analysis of the corpus. In particular, it makes it possible to reach beyond generic sentiments by identifying not only different aspects and subtopics in the Twitter debate on parenthood but also sentiments expressed on each specific subtopic.

The first step consisted of manually annotating tweets as being on-topic or off-topic. To continue to filter out off-topic tweets it was necessary to provide annotators with a tag to label any 'noise' still present in the dataset after the automatic filtering steps. We considered tweets as on-topic:

- If the user talked about parenthood, e.g.,

diventare papà è facile. fare il papà un po' di meno [becoming a father is easy, being a father a little bit less so];

- If the user expressed a mood (direct/indirect) with respect to being a parent, e.g.,

grazie di cuore sei una persona splendida e solare come Fiorello forza tanta perché ho 3 bimbi da crescere, buone feste... [Thanks you are a wonderful and sunny person like Fiorello¹². we must be strong because I have 3 kids to raise, happy holidays...].

- If the user posted an advert about being a parent, e.g.,

Confartigianato, aperte le iscrizioni al II anno di Scuola per Genitori [Confartigianato,¹³ enrolment now open for the second year of School for Parents].

On the contrary, we considered tweets off-topic when:

- The user discussed social or economic issues in general terms, e.g.,

#TextYesTo70005ToDonateForRedNoseDay la vita di un bambino costa solo 5 sterline, rendetevi conto, per noi non è niente, per loro tutto. [#TextYesTo70005ToDonateForRedNoseDay the life of a child costs only 5 pounds, for us it's nothing for them everything].

- The user employed a keyword from the keyword-based filtering step in a figurative way, e.g.,

...Ma i sogni son figli del cuore, creati in quanto dolore, spogliati della lor ragione, per questo mandati a morire... [...But dreams are children of the heart, created as pain, stripped of their reason, for this sent to die...].

- The user commented on a VIP's behaviour and actions (which does not tell us anything interesting about users' attitudes to parenthood), e.g.,

ha donato i suoi capelli ai bambini col cancro per dare la possibilità anche a loro di fare il flick [she donated her hair to children with cancer to give them a chance to do the 'flick'].

Furthermore, according to our scheme, tweets could be marked as 'unintelligible', usually because of a lack of context, as in the following example:

¹² *Fiorello* is a famous Italian showman.

¹³ *Confartigianato* is an organization that represents micro and small enterprises in Italy.

@name nè delle sue azioni... nè delle conseguenze nella vita dei figli...
[@name neither his actions... nor the consequences in the lives of the children...].

The second step in the annotation scheme was only applied to on-topic tweets. It is a crucial step since it provides the semantics for analysing the aspects of parenthood discussed on Twitter. For annotation purposes we created seven subtopics, and the annotators' task was to select one tag defining the most relevant subtopic for each post. The tags were:

- Being parents

This tag was introduced to mark when the user generically commented on his/her status as a parent, as in the following example:

Mio figlio mi sta insegnando che nella vita tutto non è mai certo e che ogni giorno può essere un salto temporale in un nuovo progresso... [*My son is teaching me that nothing is certain in life and that every day can be a temporal leap into new kinds of progress*].

- Being sons/daughters

This tag was introduced to mark sons'/daughters' point of view, i.e., a child's comments on the parent-child relationship, as in the following example:

Adolescenti oggi pt84 Sappiamo essere i figli modello. Puliamo, stiriamo, facciamo i carini, il tutto solo perché abbiamo bisogno di qualcosa
[Teenagers today pt84 We know how to be model children. We clean, we iron shirts, we are all very nice, everything because we need something].

- Daily life

This tag marked up tweets on recurring situations in the everyday relationship between parents and children, as in the following example:

@AndrewloveF1 sto aspettando mio figlio all'uscita da scuola.....? Solite cose.... [@AndrewloveF1 I'm waiting for my son after school.....? Usual stuff....].

- Judgment of parents' behaviour

This tag was for comments on children's education, for instance, or comments on behaviour that did not seem appropriate to the parent:

Staccate i bimbi dalla tele a tutto volume, dai tablet, dai centri commerciali, dalla Wii e fateli VIVERE fuori, poveri. [Get children off television at full volume, tablets, shopping centres, Wii and let them LIVE outside, poor children].

- Children's future

This tag was for tweets where parents expressed sentiments, expectations, or fears about the future of children, as in the following example:

Se un giorno i miei figli avranno i valori di questo avrò sbagliato tutto nella vita [If one day my children have the moral values of this person I'll have got everything in my life wrong].

- Becoming parents

This tag was for tweets where users spoke about the fear of becoming parents, as in the following example:

E il mio lui: Amore ci pensi quando torneremo qui saremo genitori. #ansia [And he says: Sweetheart, think that when we come back here we will be parents. #anxious].

- Fertility and politics

This tag was introduced to mark tweets about policies and political initiatives affecting parents. For instance, complaints about welfare policies:

@PMO_W dovrebbe pensare a fare bene la (sig) ministra invece di usare "desiderio" maternità come strumento di propaganda. Che tristezza [She should think about doing her job as minister well instead of using the "desire" for motherhood as a propaganda tool. How sad].

The third level of the annotation scheme was again specific to the on-topic tweets. The purpose of this stage was to provide tags as a means to label the expressed sentiment polarity of the tweets. We relied on a standard set of labels for the annotation of sentiment polarity,¹⁴ which were 'positive', 'negative', 'none', and 'mixed', as provided by Basile et al. (2014).

The presence or absence of irony was marked to examine possible reversal in sentiment polarity in cases where figurative devices were used. Irony may work as an

¹⁴ In linguistics, polarity is a positive or negative mood extracted from the text.

unexpected reverser of polarity: one says something 'good' to mean something 'bad', which risks undermining the accuracy of automatic sentiment classifiers:

Bimbo non è guarito: ha semplicemente impacchettato tutti i germi e me li ha regalati. #balata #SempreNelWeekendMiRaccomando #cosedimamma [kid not better: he simply wrapped up all the germs and gave them to me #flu #alwaysattheweekend #Mummythings].

Trovate le spade di gomma per fare la 'guerra' con mio figlio. ah la favola 'la spada nella roccia' quanti danni fa [Found rubber swords to go to 'war' with my son. the "the sword in the stone" tale. how much damage it does].

Annotating ironic devices is challenging because irony does not always depend on the semantic and syntactic elements in the text but often requires contextual knowledge (Wilson 2006; Reyes and Rosso 2014; Maynard and Greenwood 2014; Ghosh et al. 2015). To mark up irony we introduced two polarized ironic labels: 'negative humour' for negative ironic tweets, and 'positive humour' for positive ironic tweets. The following are examples of each of the six proposed labels:

- Positive

The user expressed a positive opinion or a positive feeling. For example:

Cari genitori della bambina, la state crescendo nel modo giusto [Dear girl's parents, you are raising her in the right way].

- Negative

The user expressed a negative opinion or a negative feeling;

Sono veramente desolata per i bambini di oggi che non avranno tutto questo e non lo rimpiangeranno [I'm really sorry for the children of today who will not have all this and they won't know enough to regret it].

- Mixed

The user expressed both positive and negative opinions or sentiments;

@name: "Cita e rispondi: "Vai d'accordo con i tuoi genitori?" "sì, anche se certe volte facciamo litigate assurde" [*@name: "Question and Answer: "do you get on with your parents?" "Yes, even if sometimes we argue about absurd things"*].

- None

The user did not express positive or negative opinions or sentiments. For example, the user reported a piece of news without expressing an opinion:

@tuttitrogloditi: cita e rispondi sei mai stata sorpresa dai tuoi genitori a fare qualcosa che non dovevi? “No” [@tuttitrogloditi: question and answer have you ever been caught by your parents doing something that you should not have been doing?].

- Positive humour

The tweet included ironic content and conveyed positive polarity. The target of the irony was not important, but there was no intent to insult or to damage the target. Example:

Mi mamma riesce a trovare tutto dal nulla....? “Mammaaaa!!” Ho perso gli One Direction!!!” ?? [My mom manages to find something from nothing...? “Mom!! I lost One Direction!!!” ??].

- Negative humour

The tweet included ironic content and conveyed a negative polarity. The target of the irony was not important, but there was intent to challenge the target. For example:

Vedi figliolo, un giorno tutto questo continuerai a desiderarlo. [Look, kiddo, you will still want all this one day].

3.3 Annotation process with CrowdFlower

Next we drew a random sample of 6,000 tweets from Twita-2014-parenthood (i.e., the Twitter dataset collected and filtered as reported in Section 3.1). This sample was then annotated manually using the scheme defined in Section 3.2. A pre-processing step was applied in order to remove tweets containing only Twitter marks (hashtags, mentions, or urls), which left us with 5,566 tweets. The annotation of the corpus was implemented with the help of CrowdFlower, a crowd-sourcing platform exploited for manual annotation in many similar annotation tasks related to sentiment analysis (Nakov et al. 2016). To ensure high-quality annotations we created 349 test questions in order to evaluate annotator effectiveness. We selected CrowdFlower’s ‘dynamic judgment option’ (ranging between three and five annotators). Annotators were requested to apply the annotation scheme depicted in Figure 1. They started by marking tweets as

being on-topic, off-topic, or unintelligible with respect to the parenthood domain, as defined by precise annotation guidelines.¹⁵ If the CrowdFlower annotator considered the tweet on-topic she/he proceeded to the further steps of annotation, which consisted of determining the subtopic, the sentiment polarity, and the presence of irony.

3.4 Analysis of the 'gold standard' Twitter corpus

By the end of the process, three to five independent annotations had been provided for each tweet. Whether each tweet got a gold label was decided by majority voting: At least 60% of the annotators had to agree on the label. 2,355 tweets were annotated as on-topic (42.3% out of the total of 5,566 submitted to CrowdFlower for human annotation) and 3,136 as off-topic (56.3% of the total), while there was disagreement on the remaining tweets. The proportion of on-topic tweets was high compared to other Twitter-based content and opinion surveys (Ceron, Curini, and Iacus 2014). One thousand five hundred and eight of the 2,355 on-topic tweets got a consistent gold label for all the further annotation layers concerning sentiment polarity, presence of irony, and specific semantic areas (subtopics). This set of 1,508 tweets constituted our 'gold standard corpus'¹⁶ of on-topic tweets with gold labels for sentiment polarity and subtopics. The corpus was named 'Tw-parenthood-gold' and is now publicly available.¹⁷ Table 1 shows the distribution of labels for the sentiment polarity layer in Tw-parenthood-gold, while Table 2 shows the label distribution for subtopics.

Table 1: Distribution of gold standard messages about parenthood, by polarity

Polarity	Num	%
Positive	526	34.9
Positive humour	116	7.7
Mixed	28	1.8
Negative humour	211	14.0
Negative	461	30.6
None	166	11.0
Total	1,508	100

¹⁵ Annotation guidelines (in Italian) are available at: <https://github.com/mirkolai/Happy-Parents/blob/master/guidelines.pdf>

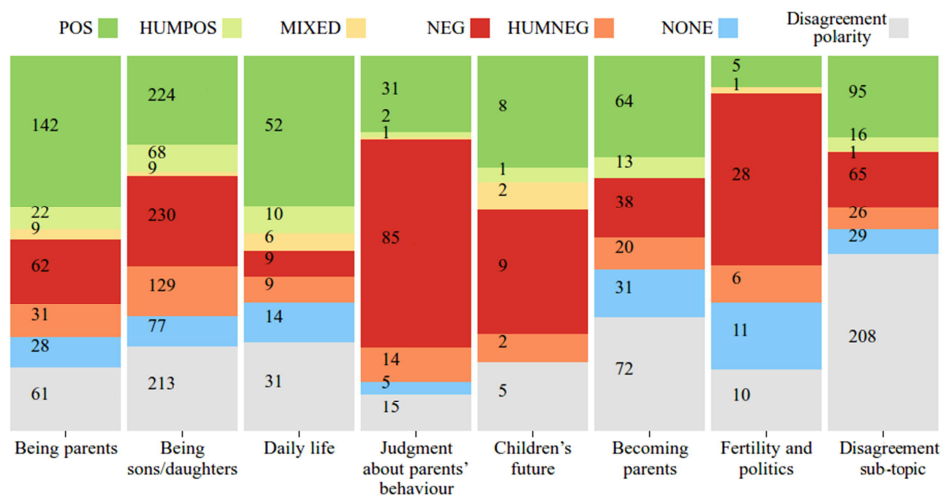
¹⁶ The standard collections called Gold Standard Corpora are trustworthy sets of tweets necessary for training and for the meaningful evaluation of algorithms that use annotations.

¹⁷ The corpus is available in a public repository: https://github.com/mirkolai/Happy-Parents/blob/master/gold_HappyParents.csv The result of the first layer of manual annotation (on-topic vs. off-topic) is also available: <https://github.com/mirkolai/Happy-Parents>.

Table 2: Distribution of gold standard messages about parenthood, by subtopic

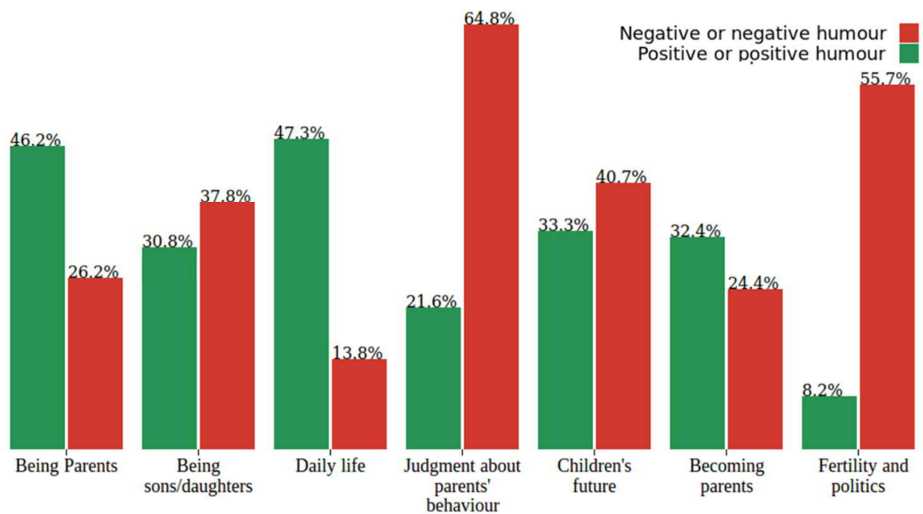
Label	Num	%
Being sons/daughters	737	48.9
Being parents	294	19.5
Becoming parents	166	11.0
Judgment about parents' behaviour	138	9.2
Daily life	100	6.6
Fertility and politics	51	3.4
Children's future	22	1.4
Total	1,508	100

847 remaining tweets (out of 2,355 on-topic tweets) were not included in our gold corpus since annotators could not agree on a common label for either both layers or for one of the two layers. Figure 2 summarizes the annotation process, showing the distribution of disagreements about the various layers and categories. The grey colour inside the bars indicates tweets where there is disagreement in terms of sentiment layer (polarity disagreement). The stacked bar on the right-hand side summarizes the cases where annotators did not agree on the subtopics layer (subtopic disagreement). Interestingly, a high level of polarity disagreement among annotators emerged when they also disagreed on the subtopic.

Figure 2: Label distribution in our gold corpus and disagreement

The overall distribution of the labels in Tw-parenthood-gold provides some clues in support of polarity change, clues that vary according to the subtopic in question. On the one hand, negative polarity prevails in tweets about the subtopics 'Judgment about parents' behaviour' and 'Fertility and politics'. On the other hand, positive polarity prevails in the subtopics 'Beings parents' and 'Daily life'. When we combine tweets classified as negative and those with negative humour and do the same with positive and positive humour tweets (Figure 3), some further aspects emerge of the relationship in the corpus between polarity and subtopic. It is quite clear that positive sentiments emerged and prevailed when people were talking about everyday life with children and the experience of becoming and being parents. On the other hand, negative sentiments were dominant in discourses about children's future, fertility and politics, and parental behaviour. Parents sometimes grumbled about their children's behaviour, but they were mostly happy with and proud of their children.

Figure 3: Prevalence (%) of negative/positive sentiments by parenthood subtopic



4. Beyond the polarity valence: a lexical analysis based on an emotion lexicon

Further analysis was carried out on the corpus at the lexical level, based on an emotion lexicon. This exercise provides more detail than a simple evaluation based on positive or negative polarity. Essentially, it provides cues as to emotions involved in the Twitter discourse on parenthood. This approach is also useful in order to train an automatic classifier with lexical features, as, for instance, in Sulis et al. 2016.

Indeed, a more nuanced result emerged when analysing the Tw-parenthood-gold corpus employing the Word-Emotion Association Lexicon EmoLex¹⁸ (see Table 3). For instance, ‘Being parents’ had a higher incidence of happy words. Messages concerning judgments and comments on the education of children (‘Judgment about parents’ behaviour’) had a high frequency of anger and disgust terms. Anticipation was, as might be expected, more frequent in the ‘Becoming parents’ group of messages. Some other interesting findings concern sadness, which was more relevant to the ‘Politics and fertility’ topic, while ‘Judgments about parents’ behaviour’ included a higher frequency of phrases related to fear. Trust appears to be more closely related to the ‘Daily life’ and ‘Being parents’ topics, consistent with the above-mentioned positive polarity. Finally, phrases expressing surprise were mostly present in the ‘Being parents’ and ‘Becoming parents’ messages.

Table 3: Distribution of emotions in ‘gold standard’ messages, by parenthood subtopic

Polarity	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust
Becoming parents	10%	35%	8%	30%	35%	20%	18%	43%
Being parents	12%	31%	7%	25%	45%	21%	16%	46%
Judgment about parents' behaviour	16%	21%	13%	37%	39%	22%	14%	43%
Children's future	7%	26%	4%	15%	37%	11%	7%	44%
Daily life	9%	36%	8%	19%	40%	18%	13%	47%
Fertility and politics	19%	36%	12%	23%	23%	30%	8%	39%

¹⁸ The Word-Emotion Association Lexicon (aka EmoLex) is a list of English words labelled according to Plutchik’s (Plutchik 2001) eight primary emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, trust) and two sentiments (negative and positive). The annotations were done manually through crowdsourcing. The NRC Emotion Lexicon has affect annotations for English words. Despite some cultural differences, it has been shown that a majority of affective norms are stable across languages. Thus, we exploited the Italian version of the lexicon provided by the NRC research group at <http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>, where the English terms are translated into over twenty languages (by Google Translate).

5. Automatic detection of sentiment polarity

Given the methods described in Sections 3 and 4, the next step was to investigate the polarity of tweets in the complete data set Twita-2014-parenthood. This analysis had two steps. First, we automatically selected tweets of interest (i.e., on-topic tweets). Second, we automatically computed the overall sentiment for each tweet using a machine-learning technique.

The aim of the first phase was to separate on-topic from off-topic tweets. Off-topic tweets were those that did not relate to fertility and parenthood, even though they contained one or more keywords. We trained a binary Support Vector Machine with the labelled tweets of the first annotation layer derived from the manual annotation process described in Section 3.3. The training set consisted of 2,355 on-topic tweets and 3,136 off-topic tweets.

We trained the model with a 'bag-of-words' model, using features like punctuation marks, tweet length (in words and characters), and the frequency of hashtags, mentions, emojis, and interjections. We did not include abbreviations, slang, or swear words, as they were infrequently used in our tweets. The trained model automatically distinguished on-topic and off-topic tweets for the entire data set of around 2.8 million tweets, thus obtaining 1,083,741 on-topic tweets (39.2%). The performance of our classifier was evaluated by the 'F-measure', which provides information on accuracy and is based on the ratio between precision and recall. A five-fold cross validation¹⁹ was applied, obtaining an F-measure value of 0.7496.

Moreover, since in our next analysis (presented in Section 6) we wanted to focus only on messages relating to parental attitudes, filtering out those related to the 'Being sons/daughters' subtopic, we performed a second binary classification experiment, aimed at distinguishing tweets labelled 'Being sons/daughters' from parental tweets (the latter group, i.e., tweets on parenthood, constituted over 39% of the total, i.e., 426,036 of the total 1,083,741). In particular, we performed a binary classification experiment relying on the same feature model as the previous experiment, using as a training set the subtopic layer of the dataset Tw-parenthood-gold described in Section 3.4.²⁰ We obtained an F-measure value of 0.75.

¹⁹ Cross-validation is a technique used to test the general accuracy of the model (Han, Pei, and Kamber 2011). In our case, the whole dataset was split into five equal parts, with one part as a test set and the other four-fifths as a training set.

²⁰ For this binary classification experiment we considered as on-topic only the tweets labelled 'Being parents,' 'Becoming parents,' 'Judgment about parents' behaviour,' 'Daily life,' 'Fertility and politics,' and 'Children's future,' taking only tweets labelled 'Being sons/daughters' as samples of the off-topic class.

In the second step we assigned polarity to the on-topic tweets using the sentiment analysis system IRADABE²¹ (Hernández-Farías, Buscaldi, and Priego-Sanchez 2014). IRADABE relies on a Support Vector Machine with surface (e.g., n-grams, emoticons, exclamation marks, and uppercase-lowercase ratio) and lexicon-based features.²² These are useful in detecting meaning, especially for sentiment and opinion posts, which are interrelated. The model is able to tag each tweet for polarity using the following labels: positive, negative, none (neutral), and mixed (both positive and negative sentiments present in a single tweet). For the experiments presented in this paper, IRADABE was trained with a corpus composed of two data sets: a previous complete data set from the benchmark Italian Twitter corpus released for the Sentipolc 2014 shared task (Basile et al. 2014), composed of 6,448 tweets in Italian on various random topics from politics to football; and the Tw-parenthood-gold corpus described in Section 3.4, considering the sentiment polarity layer. We carried out an experiment using five cross-validations on the training set. The F measure detecting negative polarity obtained by IRADABE was about 70%, with positive polarity above 77%. The performance appeared fully compatible with state-of-the-art system performance for Italian (Basile et al. 2014; Barbieri et al. 2016).

The sentiment analysis results shown in Table 4 show a prevalence of negative tweets (almost 50%), only 10% positive tweets, and a high percentage (36.1%) of mixed tweets,²³ i.e., tweets where both negative and positive attitudes were expressed. Only 4% of tweets were classified with the sentiment label, 'none'.²⁴

²¹ This system obtained one of the best results for subjectivity tasks (3rd with a 0.6706 F-measure), for polarity classification tasks (2nd with 0.6347), and for irony detection tasks (2nd with 0.5415) in an evaluation exercise for Italian (see Basile et al. 2014).

²² Such features relied on an Italian version of the following sentiment lexicons: SentiWordNet (Baccianella, Esuli, and Sebastiani 2010); Hu&Liu (Hu and Liu 2004); AFINN (Nielsen 2011); and the Dictionary of Affect in Language (Whissell 2009).

²³ We manually inspected a sample of mixed tweets and often found the presence of multiple targets and a different polarity. This is interesting, since a finer-grained sentiment analysis might help us understand the targets of the positive and negative components. It may also be possible to investigate the use of automatic stance detection systems in the corpus: the task would be to understand sentiment polarity and its target (Mohammad et al. 2016).

²⁴ Note that there is a certain margin of error in the automatic classification. In particular, consider the following performance analysis of the IRADABE classifier used here on the Sentipolc 2014 benchmark Italian Twitter dataset (Basile et al. 2014, Appendix A): when considering the results per class (positive and negative polarity) in terms of precision and recall, IRADABE's precision was better for the positive class than for the negative class, but the system score was low in recall for the positive class. This partially explains the results described in Table 4.

Table 4: Distribution of sentiment labels annotated by IRADABE

Class	Tweets	Percentage (%)
Positive	109,272	10.1
Negative	538,127	49.7
Mixed	391,522	36.1
None	44,820	4.1
TOT	1,083,741	100.0

6. The geographical distribution of positive messages

As a last step, we extracted geographical information from the messages about parenthood. As most Twitter users do not provide geographical information we could only investigate 120,307 geotagged messages (about one in four of the 426,036 messages on parenthood-related topics).

The aim here was to assess possible correlation between sentiment polarity and population characteristics. A particular measure of interest was the average number of children per woman (Total Fertility Rate). In other words, were positive sentiments related to the fertility rates in different regions? In order to do this we focused on positive messages identified by our automatic classifier, geo-referenced and aggregated by the twenty Italian regions (the administrative level above province). For these regions we relativized the distribution of positive messages over the total number of tweets in the same region, as well as over the sum of positive and negative tweets. These two measures were then compared with the region's total fertility rates. The aggregation is crude, as within these regions there is substantial variation in the fertility rate. Nevertheless, this kind of analysis sheds light on whether social media content relates to demographic variables. We obtained a positive correlation (see Table 5), suggesting an association between higher fertility and the prevalence of individuals with more positive sentiments toward parenthood. This association was reinforced by the fact that there was no correlation between the regional Crude Birth Rate (CBR, the frequency of births in one year out of the total population) and the share of positive parenthood tweets. This suggests that the correlation does not depend on the relative number of newly born children present in the population (which is relatively higher where the birth rate is higher), but rather on the level of fertility per se, measured by the yearly average number of children per woman, i.e., the TFR. Clearly, the direction of the relationship is unknown. On the one hand the higher prevalence of positive sentiment in tweets concerning parenthood might be a result of selection: Fertility might be higher in those areas where childbearing and childrearing is easier and supported by local authority policies. On the other hand, a higher prevalence of positive tweets might reflect how individuals in these regions have a stronger preference for

children – and therefore end up having more children. Independent of the direction of the relationship, there is little doubt that the positive sentiments represent a proxy for being happy with parenthood.

To corroborate this finding we verified the association between the share of tweets positive toward parenthood and the average regional level of life satisfaction. Life satisfaction regional estimates come from the harmonized data sets of the Italian National Statistical Office Multipurpose Household Surveys called Aspects of Daily Life. These cross-sectional, nationally representative surveys were repeated each year through interviews of around 20,000 households, with around 50,000 individuals. The regional values of life satisfaction are population-level estimates obtained using weights provided by the Italian National Statistical Office.²⁵ We found a positive correlation between the share of positive tweets about parenthood and the average regional level of life satisfaction (see Table 5).

Table 5: Correlation of parents' sentiment scores with regional indicators

	% Positive tweets over total tweets	% Positive tweets over sum of positive and negative tweets
Average life satisfaction	0.351	0.278
Total fertility rate	0.283	0.196
Crude birth rate	-0.099	-0.080

Source of macro regional indicators: National Institute of Statistics data for 2014. TFR and CBR are derived from vital statistics; life satisfaction is estimated from the Household Multipurpose Survey "Aspects of Daily Life".

7. Conclusions

In this paper we propose a model for collecting and semantically annotating Twitter data for demographic research on parenthood and fertility. The aim is to demonstrate the necessary steps needed in cases where the concept of interest is multifaceted and not always directly measurable. Whenever the concept is complex, considerably more effort is needed in the annotation procedure to derive meaningful classification results, which is also the case for demographic analysis and family research.

The first step, and a necessary precondition for any further analysis of this kind of content, is the development of a Twitter corpus, annotated with a novel semantic

²⁵ The data was collected using a two-stage sampling design with a stratification of the primary units. The municipalities are the primary units and the households are the secondary units. The municipalities were sampled with probabilities proportional to their population size and without replacement, whereas the households were drawn with equal probabilities and without replacement. All members of the sampled households were interviewed face-to-face. The overall response rate for these surveys was greater than 80%, and there was no major difference in response rates across surveys.

scheme for marking up information. This approach produced data that had been semantically enriched with information about sentiment and specific sentiment targets in Twitter communications between users talking about parenthood. Importantly, the annotation process yielded not only sentiment polarity but also specific semantic areas and subtopics that were sentiment targets in the relationship between parenthood and happiness.

When we consider the sentiment layers the polarity expressed was mainly positive in tweets in which parents talked about their children or their experience of being parents. If we also take into account the different semantic categories that represent the sentiment target (only in the gold standard corpus) the picture becomes more complex, and more interesting for an entangled domain like the one we are focusing on. Our data shows that towards some targets the polarity of the sentiment could also be negative. Interestingly, it emerged that parents expressed positive sentiments when they talked about daily life with children and becoming and being parents, while at times also being fearful, surprised, and sad. In tweets about children's future, fertility, politics, and parental behaviour, negative sentiments prevailed. By scrutinizing opinions on Twitter, which are posted spontaneously, often as a reaction to emotionally driven observations, we thus gain insight into the 'parenthood happiness paradox': Positive and negative feelings toward parenthood co-exist in the Italians tweets.

By using the geocodes associated with (a sub-sample of) tweets, sentiments can, as others have shown before us, be linked to the resident population in a given area (in this case the Italian regions), and then be usefully compared with the socioeconomic characteristics of that area. Here we show how this can be done in relation to fertility. Aggregated measures of positive sentiments appear to be correlated with regional fertility levels. The more positive the sentiments, the higher the fertility. Though the aggregation is crude, this finding is a first for Italy.

Clearly, further information on user characteristics is fundamental to making sense of social media data for demographic purposes. It would have been particularly interesting to know the user's sex, age, and number of children. A caveat of our study and classification is the lack of Twitter users' sociodemographic traits. Twitter does not provide explicit metadata about the age and gender of users. Nevertheless, there are now studies that propose methods to extract this information from social media data, thus opening the way to more ambitious future studies. Some authors have suggested getting information on the sociodemographic traits of Twitter users by manually inspecting data that has been published elsewhere, e.g., on LinkedIn profiles. When age is not given it could be estimated by taking into account any information included, say, in the education section, such as the starting date of a degree. Gender could be inferred from profile photos and names by following a methodology similar to that in Rangel et al. (2014). In particular, the idea of extracting information about the age and gender of

users by automatically analysing their pictures, relying on advanced face-recognition techniques, might allow a novel methodological framework for a demographic-oriented analysis of social media and an assessment of present theoretical ideas. In our case it was possible to extract semantic information on textual content and demographic characteristics from the data set, but we feared that the margin of error would be too large.

In all, we examined a data set that is by its very nature non-representative of the Italian population as a whole. Twitter users tend to be young (see, for instance, the results of a 2012 ISPO poll ²⁶), and tend to use Twitter more for getting timely news than for discussing family-related issues. However, non-representativeness is an issue for any qualitative study. We show how tweets can be used to explore attitudes, values, and feelings related to family life. Social-media-derived linguistic analysis data thus provides a middle ground between qualitative studies and the more standard quantitative approaches.

8. Acknowledgments

The authors gratefully acknowledge financial support from the European Research Council under the FP7 European ERC Grant Agreement StG- 313617 (SWELL-FER: Subjective Well-being and Fertility. P.I. Letizia Mencarini). We thank Michele Mozzachiodi for help with data preparation in the early stages of this research and Arjuna Tuzzi, and Arnstein Aassve for comments on an earlier version of the manuscript.

²⁶ www.ispo.it and specifically <https://www.slideshare.net/MilanIN/twitter-in-italia-ricerca-ispo-click-2012>.

References

- Aggarwal, C.C. and Abdelzaher, T.F. (2013). Social sensing. In: Aggarwal, C.C. (ed.). *Managing and mining sensor data*. New York: Springer: 237–297. doi:[10.1007/978-1-4614-6309-2_9](https://doi.org/10.1007/978-1-4614-6309-2_9).
- Allisio, L., Mussa, V., Bosco, C., Patti, V., and Ruffo, G. (2013). Felicità: Visualizing and estimating happiness in Italian cities from geotagged tweets. In: Battaglini, C., Bosco, C., Cambria, E., Damiano, R., Patti, V., and Rosso, P. (eds.). *Proceedings of the 1st International Workshop on Emotion and Sentiment in Social and Expressive Media (ESSEM 2013)*. Turin: CEUR Workshop Proceedings: 95–106.
- Attardi, G., Basile, V., Bosco, C., Caselli, T., Dell'Orletta, F., Montemagni, S., Patti, V., Simi, M., and Sprugnoli, R. (2015). State of the art language technologies for Italian: The EVALITA 2014 perspective. *Intelligenza Artificiale* 9(1): 43–61. doi:[10.3233/IA-150076](https://doi.org/10.3233/IA-150076).
- Baccianella, S., Esuli, A., and Sebastiani, F. (2010). SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In: Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M., and Tapias, D. (eds.). *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010)*. Paris: ELRA.
- Barbieri, F., Basile, V., Croce, D., Nissim, M., Novielli, N., and Patti, V. (2016). Overview of the EVALITA 2016 SENTiment POLarity classification task. In: Basile, P., Cutugno, F., Nissim, M., Patti, V., and Sprugnoli, R. (eds.). *Proceedings of the 5th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2016)*. Turin: Accademia University Press.
- Basile, V. and Nissim, M. (2013). Sentiment analysis on Italian tweets. In: Balahur, A., van der Goot, E., and Montoyo, A. (eds.). *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. Atlanta: ACL: 100–107.
- Basile, V., Bolioli, A., Nissim, M., Patti, V., and Rosso, P. (2014). Overview of the Evalita 2014 SENTiment POLarity classification task. In: Bosco, C., Cosi, P., Dell'Orletta, F., Falcone, M., Montemagni, S., and Simi, M. (eds.). *Proceedings of the 4th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2014)*. Pisa: Pisa University Press: 50–57.

- Basile, V., Cutugno, F., Nissim, M., Patti, V., and Sprugnoli, R. (2016). Overview of the 5th evaluation campaign of Natural Language Processing and Speech Tools for Italian. In: Basile, P., Cutugno, F., Nissim, M., Patti, V., and Sprugnoli, R. (eds.). *Proceedings of the 5th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2016)*. Turin: Accademia University Press.
- Billari, F.C., Cavalli, N., Qian, E., and Weber, I. (2017). *Footprints of family change: A study based on Twitter*. Paper presented at the Annual Meeting of the Population Association of America, Chicago, USA, April 27–29, 2017.
- Billari, F.C., D’Amuri, F., and Marcucci, J. (2013). *Forecasting births using Google*. Paper presented at the Annual Meeting of the Population Association of America, New Orleans, USA, April 11–13, 2013.
- Blumenstock, J.E., Gillick, D., and Eagle, N. (2010). Who’s calling? Demographics of mobile phone use in Rwanda. In: Eagle, N. and Horvitz, E. (eds.). *AAAI Spring Symposium: Artificial Intelligence for Development 2010*. Menlo Park: AAAI Press: 116–117.
- Bollen, J. and Mao, M. (2011). Twitter mood as a stock market predictor. *Computer* 44(10): 91–94. [doi:10.1109/MC.2011.323](https://doi.org/10.1109/MC.2011.323).
- Bosco, C., Allisio, L., Mussa, V., Patti, V., Ruffo, G., Sanguinetti, M., and Sulis, E. (2014). Detecting happiness in Italian tweets: Towards an evaluation dataset for sentiment analysis in Felicità. In: Schuller, B., Buitelaar, P., Devillers, L., Pelachaud, C., Declerck, T., Batliner, A., Rooso, P., and Gaines, S. (eds.). *Proceedings of the 5th International Workshop on Emotion, Social Signals, Sentiment and Linked Open Data*. Paris: ELRA: 56–63.
- Bosco, C., Patti, V., and Bolioli, A. (2013). Developing corpora for sentiment analysis: The case of irony and senti-TUT. *IEEE Intelligent Systems* 28(2): 55–63. [doi:10.1109/MIS.2013.28](https://doi.org/10.1109/MIS.2013.28).
- Bosco, C., Patti, V., and Bolioli, A. (2015). Developing corpora for sentiment analysis: The case of irony and senti-TUT. In: Yang, Q. and Wooldridge, M. (eds.). *Proceedings of the 24th International Conference on Artificial Intelligence*. Menlo Park: AAAI Press: 4158–4162.

- Buscaldi, D. and Hernández-Farías, D.I. (2015). Sentiment analysis on microblogs for natural disasters management: A study on the 2014 Genoa floodings. In: Gangemi, A., Leonardi, S., and Panconesi, A. (eds.). *Proceedings of the 24th International Conference on World Wide Web Companion (WWW 2015)*. New York: ACM: 1185–1188. doi:[10.1145/2740908.2741727](https://doi.org/10.1145/2740908.2741727).
- Castells, M. (2000). *The rise of the network society*. Cambridge: Blackwell.
- Ceron, A., Curini, L., and Iacus, S.M. (2014). *Social media e sentiment analysis: L'evoluzione dei fenomeni sociali attraverso la Rete*. Milan: Springer. doi:[10.1007/978-88-470-5532-2](https://doi.org/10.1007/978-88-470-5532-2).
- Cetre, S., Clark, A.E., and Senik, C. (2016). Happy people have children: Choice and self-selection into parenthood. *European Journal of Population* 32(3): 445–473. doi:[10.1007/s10680-016-9389-x](https://doi.org/10.1007/s10680-016-9389-x).
- Clark, R., Ogawa, N., Lee, S.-H., and Matsukura, R. (2008). Older workers and national productivity in Japan. *Population and Development Review* 34(Supplement): 257–274.
- De Choudhury, M., Counts, S., and Horvitz, E. (2013). Predicting postpartum changes in emotion and behavior via social media. In: Mackay, W.E., Brewster, S., and Bødker, S. (eds.). *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2013)*. New York: ACM: 3267–3276. doi:[10.1145/2470654.2466447](https://doi.org/10.1145/2470654.2466447).
- Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F.R., Gaughan, A.E., Blondel, V.D., and Tatem, A.J. (2014). Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences* 111(45): 15888–15893. doi:[10.1073/pnas.1408439111](https://doi.org/10.1073/pnas.1408439111).
- Diener, E. (2000). Subjective well-being: The science of happiness and a proposal for a national index. *American Psychologist* 55(1): 34–43. doi:[10.1037/0003-066X.55.1.34](https://doi.org/10.1037/0003-066X.55.1.34).
- Ghosh, A., Li, G., Veale, T., Rosso, P., Shutova, E., Barnden, J., and Reyes, A. (2015). Semeval-2015 task 11: Sentiment analysis of figurative language in Twitter. In: Nakov, P., Zesch, T., Cer, D., and Jurgens, D. (eds.). *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*. Denver: ACL: 470–478.
- Han, J., Pei, J., and Kamber, M. (2011). *Data mining: Concepts and techniques*. Woltham: Elsevier.

- Hansen, T. (2012). Parenthood and happiness: A review of folk theories versus empirical evidence. *Social Indicators Research* 108(1): 29–64. doi:10.1007/s11205-011-9865-y.
- Hernández-Farías, D.I., Buscaldi, D., and Priego-Sánchez, B. (2014). IRADABE: Adapting English lexicons to the Italian sentiment polarity classification task. In: Basili, R., Lenci, A., and Magnini, B. (eds.). *Proceedings of the 1st Italian Conference on Computational Linguistics (CLiC-IT 2017)*. Pisa: Pisa University Press: 75–81.
- Hilbert, M. and López, P. (2011). The world’s technological capacity to store, communicate, and compute information. *Science* 332(6025): 60–65. doi:10.1126/science.1200970.
- Hitsch, G.J., Hortaçsu, A., and Ariely, D. (2010). Matching and sorting in online dating. *American Economic Review* 100(1): 130–163. doi:10.1257/aer.100.1.130.
- Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In: Kim, W., Kohavi, R., Gehrke, J., and DuMouchel, W. (eds.). *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2004)*. New York: ACM: 168–177. doi:10.1145/1014052.1014073.
- King, G. (2011). Ensuring the data-rich future of the social sciences. *Science* 331(6018): 719–721. doi:10.1126/science.1197872.
- Kohler, H.P. and Mencarini, L. (2016). The parenthood happiness puzzle: An introduction to special issue. *European Journal of Population* 32(3): 327–338. doi:10.1007/s10680-016-9392-2.
- Kohler, H.P., Behrman, J.R., and Skyttke, A. (2005). Partner + children = Happiness? The effects of partnerships and fertility on well-being. *Population and Development Review* 31(3): 407–445. doi:10.1111/j.1728-4457.2005.00078.x.
- Lai, M., Virone, D., Bosco, C., and Patti, V. (2015). Debate on political reforms in Twitter: A hashtag-driven analysis of political polarization. In: *Proceedings of 2015 IEEE International Conference on Data Science and Advanced Analytics, Special Track on Emotion and Sentiment in Intelligent Systems and Big Social Data Analysis*. Paris: IEEE: 1–9. doi:10.1109/DSAA.2015.7344884.
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., and Alstyne, M.V. (2009). Social science: Computational social science. *Science* 323(5915): 721–723. doi:10.1126/science.1167742.

- Lesthaeghe, R. (2010). The unfolding story of the second demographic transition. *Population and Development Review* 36(2): 211–251. doi:10.1111/j.1728-4457.2010.00328.x.
- Liu, B. (2010). *Sentiment analysis and subjectivity*. Boca Raton: Taylor and Francis.
- Margolis, R. and Myrskylä, M. (2011). A global perspective on happiness and fertility. *Population and Development Review* 37(1): 29–56. doi:10.1111/j.1728-4457.2011.00389.x.
- Maynard, D. and Greenwood, M. (2014). Who cares about sarcastic tweets? Investigating the impact of sarcasm on sentiment analysis. In: Calzolari, N., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., and Piperidis, S. (eds.). *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*. Paris: ELRA.
- McDonald, P. (2013). Societal foundations for explaining fertility: Gender equity. *Demographic Research* 28(34): 981–994. doi:10.4054/DemRes.2013.28.34.
- Mencarini, L. (2018). The potential of the computational linguistic analysis of social media for population studies. In: Nissim, M., Patti, V., Plank, B., and Wagner, C. (eds.). *Proceedings of the 2nd Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*. New Orleans: ACL: 62–68. doi:10.18653/v1/W18-1109.
- Meo, R. and Sulis, E. (2017). Processing affect in social media: A comparison of methods to distinguish emotions in tweets. *ACM Transactions on Internet Technology* 17(1): 7. doi:10.1145/2996187.
- Mitchell, L., Frank, M.R., Harris, K.D., Dodds, P.S., and Danforth, C.M. (2013). The geography of happiness: Connecting Twitter sentiment and expression, demographics, and objective characteristics of place. *PLoS ONE* 8(5): e64417. doi:10.1371/journal.pone.0064417.
- Mohammad, S.M., Kiritchenko, S., Parinaz, S., Xiaodan, Z., and Cherry, C. (2016). Semeval-2016 task 6: Detecting stance in tweets. In: Bethard, S., Carpuat, M., Cer, D., Jurgens, D., Nakov, P., and Zesch, T. (eds.). *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval 2016)*. San Diego: ACL: 31–41.
- Mohammad, S.M., Zhu, X., Kiritchenko, S., and Martin, J. (2015). Sentiment, emotion, purpose, and style in electoral tweets. *Information Processing and Management* 51(4): 480–499. doi:10.1016/j.ipm.2014.09.003.

- Myrskylä, M. and Margolis, R. (2014). Happiness: Before and after the kids. *Demography* 51(5): 1843–1866. doi:[10.1007/s13524-014-0321-x](https://doi.org/10.1007/s13524-014-0321-x).
- Nakov, P., Ritter, A., Rosenthal, S., Sebastiani, F., and Stoyanov, V. (2016). SemEval-2016 task 4: Sentiment analysis in Twitter. In: Bethard, S., Carpuat, M., Cer, D., Jurgens, D., Nakov, P., and Zesch, T. (eds.). *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval 2016)*. San Diego: ACL: 1–18.
- Nielsen, F.A. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. In: Rowe, M., Stankovic, M., Dadzie, A.-S., and Hardey, M. (eds.). *Proceedings of the ESWC 2011 Workshop on ‘Making Sense of Microposts’: Big things come in small packages (MSM 2011)*. Heraklion: CEUR Workshop Proceedings: 93–98.
- Nissim, M. and Patti, V. (2016). Semantic aspects in sentiment analysis. In: Pozzi, F.A., Fersini, E., Messina, E., and Liu, B. (eds.). *Sentiment analysis in social networks*. Cambridge: Elsevier: 31–48.
- Plutchik, R. (2001). The nature of emotions. *American Scientist* 89(4): 344–350. doi:[10.1511/2001.4.344](https://doi.org/10.1511/2001.4.344).
- Quercia, D., Crowcroft, J., Ellis, J., and Capra, L. (2012). Tracking ‘gross community happiness’ from tweets. In: Gergle, D., Ringel Morris, M., Bjørn, P., and Konstan, J. (eds.). *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW 2012)*. New York: ACM: 965–968. doi:[10.1145/2145204.2145347](https://doi.org/10.1145/2145204.2145347).
- Rangel, F., Rosso, P., Chugur, I., Potthast, M., Trenkmann, M., Stein, B., Verhoeven, B., and Daelemans, W. (2014). Overview of the 2nd author profiling task at PAN 2014. In: Cappellato, L., Ferro, N., Halvey, M., and Kraaij, W. (eds.). *Working notes for the CLEF 2014 Conference*. Sheffield: CEUR Workshop Proceedings: 898–927.
- Reimsbach-Kounatze, C. (2015). The proliferation of ‘big data’ and implications for official statistics and statistical agencies: A preliminary analysis. Paris: OECD (OECD Digital Economy Papers 245).
- Reis, B.Y. and Brownstein, J.S. (2010). Measuring the impact of health policies using internet search patterns: The case of abortion. *BMC Public Health* 10(514): 1–5. doi:[10.1186/1471-2458-10-514](https://doi.org/10.1186/1471-2458-10-514).

- Reyes, A. and Rosso, P. (2014). On the difficulty of automatically detecting irony: Beyond a simple case of negation. *Knowledge and Information Systems* 40(3): 595–614. doi:10.1007/s10115-013-0652-8.
- Senior, J. (2015). *All joy and no fun: The paradox of modern parenthood*. New York: Little Brown Book Group.
- Sobolevsky, S., Szell, M., Campari, R., Couronné, T., Smoreda, Z., and Ratti, C. (2013). Delineating geographical regions with networks of human interactions in an extensive set of countries. *PLoS ONE* 8(12): e81707. doi:10.1371/journal.pone.0081707.
- Stranisci, M., Bosco, C., Patti, V., and Hernández-Farías, D.I. (2016). Annotating sentiment and irony in the online Italian political debate on #labuonascuola. In: Calzolari, N., Choukri, K., Declerck, T., Goggi, S., Grobelnik, M., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., and Piperidis, S. (eds.). *Proceedings of the 10th edition of the Language Resources and Evaluation Conference (LREC 2016)*. Paris: ELRA.
- Sulis, E., Hernández-Farías, D.I., Rosso, P., Patti, V., and Ruffo, G. (2016). Figurative messages and affect in Twitter: Differences between #irony, #sarcasm and #not. *Knowledge-Based Systems* 108: 132–143. doi:10.1016/j.knosys.2016.05.035.
- Sulis, E., Lai, M., Vinai, M., and Sanguinetti, M. (2015). Exploring sentiment in social media and official statistics: A general framework. In: Bosco, C., Cambria, E., Damiano, R., Patti, V., and Rosso, P. (eds.). *Proceedings of the 2nd International Conference on Emotion and Sentiment in Social and Expressive Media (ESSEM 2015)*. New York: ACM: 96–105.
- Tumasjan, A., Sprenger, T.O., Sandner, P.G., and Welp, I.M. (2011). Predicting elections with Twitter: What 140 characters reveal about political sentiment. In: Nicolov, N., Shanahan, J.G., Adamic, L., Baeza-Yates, R., and Counts, S. (eds.). *Proceedings of the 5th International Conference on Weblogs and Social Media*. Menlo Park: AAAI Press: 178–185.
- Van de Kaa, D.J. (1987). Europe's second demographic transition. *Population Bulletin* 42(1): 1–59.
- Verma, S., Vieweg, S., Corvey, W., Palen, L., Martin, J.H., Palmer, M., Schram, A., and Anderson, K.M. (2011). Natural language processing to the rescue? Extracting 'situational awareness' tweets during mass emergency. In: Nicolov, N., Shanahan, J.G., Adamic, L., Baeza-Yates, R., and Counts, S. (eds.).

Proceedings of the 5th International Conference on Weblogs and Social Media.
Menlo Park: AAAI Press: 385–392.

- Whissell, C. (2009). Using the revised dictionary of affect in language to quantify the emotional undertones of samples of natural languages. *Psychological Reports* 2(105): 509–521. doi:[10.2466/PRO.105.2.509-521](https://doi.org/10.2466/PRO.105.2.509-521).
- Wilson, D. (2006). The pragmatics of verbal irony: Echo or pretence? *Lingua* 116(10): 1722–1743. doi:[10.1016/j.lingua.2006.05.001](https://doi.org/10.1016/j.lingua.2006.05.001).
- Zagheni, E. and Weber, I. (2012). You are where you E-mail: Using E-mail data to estimate international migration rates. In: Contractor, N., Uzzi, B., Macy, M., and Nejd, W. (eds.). *Proceedings of the 4th Annual ACM Web Science Conference*. New York: ACM: 348–351. doi:[10.1145/2380718.2380764](https://doi.org/10.1145/2380718.2380764).
- Zagheni, E. and Weber, I. (2015). Demographic research with non-representative internet data. *International Journal of Manpower* 36(1): 13–25. doi:[10.1108/IJM-12-2014-0261](https://doi.org/10.1108/IJM-12-2014-0261).
- Zagheni, E., Garimella, V.R.K., and Weber, I. (2014). Inferring international and internal migration patterns from twitter data. In: Chung, C.-W., Broder, A., Shim, K., and Suel, T. (eds.). *Proceedings of the 23rd International Conference on World Wide Web*. New York: ACM: 439–444. doi:[10.1145/2567948.2576930](https://doi.org/10.1145/2567948.2576930).
- Zagheni, E., Weber, I., and Gummadi, K. (2017). Estimate stock of migrants using Facebook’s advertising platform. *Population and Development Review*: online first.

