Dondena Working Papers

Carlo F. Dondena Centre for Research on Social Dynamics and Public Policy

Population Dynamics and Health

Italian happy parents in Twitter

Letizia Mencarini
Delia Irazú Hernández-Farías
Mirko Lai
Viviana Patti
Emilio Sulis
Daniele Vignoli

Working Paper No. 117
April 2018

Università Bocconi • The Dondena Centre
Via Guglielmo Röntgen 1, 20136 Milan, Italy
http://www.dondena.unibocconi.it

ISSN-2035-2034

Italian happy parents in Twitter*

Letizia Mencarini, Dondena Centre for Research on Social Dynamics and Public Policy, Bocconi University;
Delia Irazú Hernández-Farías, Universitat Politècnica de València;
Mirko Lai, University of Turin;
Viviana Patti, University of Turin;
Emilio Sulis, University of Turin;
Daniele Vignoli, University of Florence.

Abstract This article explores opinions and semantic orientation around fertility and parenthood by scrutinizing filtered Italian Twitter data. We propose a novel methodological framework relying on Natural Language Processing techniques for text analysis, which is aimed at extracting sentiments from texts. A manual annotation for exploring sentiment and attitudes to fertility and parenthood was applied to Twitter data. The resulting set of tweets (*corpus*) was analysed through sentiment and emotion lexicons in order to highlight how affective language is used in this domain. It emerges that parents express a generally positive attitude towards their children and being and become parents, but quite negative sentiments on children's future, politics and fertility and also parental behaviour.
Exploiting geographical information from tweets, we find a significant correlation between the prevalence of positive sentiments about parenthood and macro-regional indicators for both life satisfaction and fertility levels.

Key words: sentiment analysis, social media, fertility, parenthood, subjective well-being, linguistic corpora

* 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 on data preparation at early stages of this research and Arjuna Tuzzi for comments on an earlier version of the manuscript.

1. Introduction

Rapid increases in computational power and storage capabilities (Hilbert and Lopez 2011) have radically transformed human communications and societies (Castells 2000). The massive dissemination of information heralds a new era in social studies, one that brings with it new research challenges and opportunities (King 2011; Lazer et al. 2009; Aggarwal 2013): this holds true, not least, for demographic analyses. For example, email data has been used to track migrants (Zagheni and Weber 2012); Facebook data to monitor stocks of migrants (Zagheni et al. 2017); patterns of short- and long-term migration using Twitter data (Zagheni et al. 2014); fertility patterns using Google search data (Billari et al. 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: namely, complementary information on attitudes, values, feelings or intentions. Such concepts play a central role in key theoretical approaches for explaining demographic change, the prime example being the Second Demographic Transition Theory (Van de Kaa 1986; Lesthaeghe 2010). Social media data have great potential in this respect since they typically contain written text statements. However, they also raise tremendous challenges since Twitter or Facebook texts are invariably disordered: we do not have the structured measures one gets from, say, survey questionnaires. Still, the advantage of social-media data is obvious since they are continuously produced and are now becoming available for almost all countries, even those where traditional surveys data are not available. For demographers, however, who are interested in linking demographic trends and behaviour with soft measures, much attention has to be given over to defining the meaning of text statement – also known as the annotation process. Often these analyses are done in quite a crude way. The number of positive and negatively loaded words are counted, and, then, held up against keywords representing the demographic phenomenon of interest. But as the concept of interest becomes more complex, semantic analysis becomes more challenging.

The aim of this paper is to demonstrate how computational linguistics techniques can be used to analyse the relationship between a demographic feature and what we earlier referred to as a “soft measure”. In particular, we explore opinions and semantic orientations related to fertility and parenthood. This application lends itself to a burgeoning literature on fertility and parenthood, on the one hand, and subjective well-being, on the other. There is a longstanding academic and non-academic debate about the role played by children in parents’ daily life and for parents’ subjective well-being. These range from pure qualitative analysis, such as the book by the award-winning journalist Jennifer Senior (*All Joy and No Fun* 2015), to the more traditional data-driven approaches we have seen in demography (e.g., Kohler et al. 2005; Clark et al. 2008; Margolis and Myrskylä 2011; Myrskylä and Margolis 2014). While data-driven studies provide important information on the dynamics that link subjective well-being and childbearing from a quantitative point of view (see Kohler and Mencarini 2016 for a review), they can only provide limited insights into opinions and emotional attitudes toward fertility choices and parenthood. In addition, most of our knowledge, derived from statistical analysis on survey data, point to a “parenthood happiness paradox”. Here “folk” beliefs stand against recent empirical literature on this topic. Even in low fertility countries “folk” beliefs have it that children bring happiness. Empirical studies, meanwhile, find that the birth of a child has an average 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. They give evidence on how people talk spontaneously about parenthood and children.

The approach we present consists of two steps. We implement first a Natural Language Processing (NLP) pipeline, which is a set of modules where the output of one feeds into the next. Included in this stage is an analysis of the selected tweets to highlight relationships between the use of affective language and sub-topics of interest. This step sheds lights on the social-media content of messages related to fertility domains. The end product of this phase is what is known as a *gold standard corpus*, which is essentially a body of trustworthy texts used for training and for meaningful evaluation in the next stage. The second phase consists of a supervised machine-learning experiment carried out on the overall dataset and based on the annotated tweets from stage one. Employing well-known algorithms from NLP, we distinguish, in an automated way, messages concerning children, parenthood or fertility (*in-topic*) from others (*off-topic*). We also set out to detect sentiment *polarity*. In this way, we try to infer to what extent 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. The value of this material is, thus, potentially great.

2. Why sentiment analysis matters in demographic studies

Socio-demographic research has already benefited from complex – and big – data sources[1], 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 et al.2010); and administrative areas (Sobolevsky et al.2013). Data on Internet searches are helpful in researching fertility (Billari et al.2013); abortion rates (Reis and Brownstein 2010); and union and marriage formation (Hitsch et al. 2010). In addition, online social media, like Twitter, has been used to study migration patterns (Zagheni et al. 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 for 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 in Twitter was 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 Hernandez-Farias 2015). Examples from the social sciences include estimations of subjective well-being and, as such, sentiment analysis has helped derive measures of happiness within economics, complementing more traditional measures of well-being such as Gross Domestic Product

[1] Big data is a term for data sets that are so large or complex that traditional application software is inadequate in dealing with them. Big data sources are, as the name suggests, repositories of large volumes of data.

(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 Hao 2011), or attempted to measure the well-being of a given population (Quercia et al. 2012).

Sentiment analysis naturally relies on annotated datasets or sentiment *lexica*, resources such as dictionaries or word lists labelled with sentiment polarity (Nissim and Patti 2016). However, in most cases, sentiments are estimated through simple word-counting, which are either positively or negatively loaded. As researchers seek to use social media data to answer more complex research questions, the demands made on the sentiment analysis also become more onerous.

In demography, key theoretical contributions 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 values change. 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 annotated tweet sets for sentiment analysis and opinion mining become an indispensable resource for secondary analysis. Here, for instance, machine-learning statistical models are used for making classifications. Not surprisingly most applications of this kind are based on English. Italian is much less frequently used (though see Bosco et al. 2013, 2014, 2015; Barbieri et al. 2016), though 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 toward fertility and parenthood

In this study we are interested in fertility and parenthood and the way these relate to individuals' emotional opinions. The study, then, is relevant to previous social-media-based studies considering subjective well-being, but we are more specific, looking at how social media relates to parenthood. This section describes data collection and the annotation process. Annotation is a key challenge whenever a new theme is considered and it is a crucial step no matter what topic is being analysed.

3.1 The collection and filtering of relevant data

We extracted a set of messages (*corpus* in linguistics[2]) from Twitter for the domain of interest. We used dataset "Twita-2014"[3] consisting of 259,893,081 Italian-language tweets (of which 4,766,342 had been geotagged). The geographical distribution is quite even by administrative region and is directly and proportionally related to the resident population of the country (Pearson Correlation Coefficient of 0.93). This assures coverage of all Italian territory.

Next we filter the data set "Twita-2014" to select a sub-sample of tweets where users talk about the topics of interest. Data filtering exploits *hashtags*[4] *or keywords* in order to

---

[2] In linguistics. a *corpus* (plural *corpora*) is a large and structured set of texts collected to perform linguistics analysis. We used it in order to apply natural language processing and machine learning experiments.

[3] Twita-2014 has been gathered using the Twitter Streaming Application Programming Interface as described in Basile and Nissim (2013).

[4] Hashtag is a type of metadata tag used on 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

select relevant tweets through two steps. In the first *keyword-based filtering* step, the inflection (diminutives, singulars and plurals) of eleven *hashtags and keywords*[5] were used to select tweets of interest. We manually constructed this list through a content analysis of 2,500 randomly sampled tweets. This was followed by a linguistic analysis of synonyms. The subset after this keyword-based filtering step included about 3.9 million tweets.

In the second *user-based filtering* step, we attempted to remove 'noisy' tweets from the *corpus*. We defined a *noisy* tweet as a message without individual views on fertility and parenthood. To this end, all tweets sent from company, institutional and newspaper accounts were deleted. We also manually scanned the 500 most prolific users in 2014 and automatically searched, in the top 3,000, for online newspapers and news websites. This *duplicate-based filtering* step allowed us to delete most advertisements relating to fertility: re-tweets and other duplicated tweets not explicitly marked as re-tweets were also detected and deleted. After having removed this "noise" about 2.8 million tweets remained in the new *corpus* (henceforth "Twita-2014-parenthood").

3.2. Criteria for manual annotation for exploring sentiment and irony in parenthood-related topics

For our definitive dataset "Twita-2014-parenthood" we developed an annotation model to study the sentiments expressed in tweets and also specific parenthood-related topics (sub-topic) discussed in Twitter. This opens the way to a fine-grained sentiment analysis of the *corpus*. With this type of analysis, it is possible to reach beyond generic sentiments, by

---

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.

[5] 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, and pregnancy).

identifying different aspects and topics in the Twitter debate on parenthood and the sentiments expressed toward each aspect or topic. To build the annotation model, we relied on a standard scheme for the annotation of sentiment *polarity*[6]. We did so by exploiting the same labels, namely "positive", "negative", "none" and "mixed", as already 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 are used. Irony may work as an unexpected reverser of polarity: one says something "good" to mean something "bad". This 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. Often it 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. Finally, we created an annotation scheme to mark the specific semantic areas (i.e. all the sub-topics) of all tweets related to parenthood. This part of the annotation scheme has been crucial since it provided us with the

---

[6] In linguistics, polarity is a positive or negative mood extracted from the text.

semantics for analysing the aspects of parenthood discussed on Twitter. For annotation purposes we created seven sub-topics:

- *Being parents*

This tag is introduced to mark when the user generically comments 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 is introduced to mark the 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 is to mark 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 about parents' behaviour*

This tag is for comments on children's education, for instance, or comments on behaviour that does not seems appropriate for the parent:

"L'asilo di mia figlia mi ha fatto capire che oggi non si mandano più i figli x imparare qualcosa, ma per depositarli mezza giornata"

("My daughter's kindergarten has made me realize that today parents no longer send children there to learn something, but to park them there for half a day")

- *Children's future*

This tag is for tweets where parents express 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 is for tweets where users speak 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 is introduced to mark tweets about laws affecting parents. For instance, the so-called "baby bonus", an income transfer to new parents with low resources introduced by the government:

"@alinomilan io non mi rassegno. La Meloni aveva fatto una buona proposta a riguardo. 200 euro al mese x ogni figlio x < 5 anni. #sesivuolesipuò"

("@alinomilan I cannot just accept this. Meloni[7] had made a good proposal in this regard. 200 EUR per month per child < 5-year-old. #ifyouwantyoucan")

The task for the annotators was to select one tag for each post to define the most relevant sub-topic. Two additional tags related to the "*in/off topic*" were, subsequently, added to allow

[7] Meloni is the leader of a right–wing Italian party.

annotators to mark whether the tweet was relevant or not to the domain. The addition of an "*off topic*" tag was necessary because of the "noise" still present in the dataset.

3.3. Annotation process with *"CrowdFlower".*

We selected a random sample of 6,000 tweets from our *corpus* to be annotated manually. Before proceeding we manually inspected this sample by removing other "noisy" messages, missed by the automatic filtering steps, and had, after "noise" removal, 5,566 tweets. The annotation of the *corpus* was performed with "CrowdFlower", a crowdsourcing platform for manual annotation (Nakov et al. 2016). To encourage high-quality annotations and to discard poor ones, we created 349 test questions in order to evaluate annotator affordability. We selected "CrowdFlower's dynamic judgment option" (from three to five annotators). The annotation task had two steps. In the first step annotators had to mark the tweets as being *in-topic, off topic* or unintelligible with respect to the domain, as defined by precise guidelines. We considered tweets *in-topic*:

If the user talks 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 expresses 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 Fiorello[8]*.
we must be strong because I have 3 kids to raise, happy holidays…")*

If the user posts an advert about being a parent. e.g.:

---

[8] *Fiorello* is a famous Italian showman.

"Confartigianato, aperte le iscrizioni al II anno di Scuola per Genitori"

("Confartigianato[9], enrollment now open for the second year of Parents School ")

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

The user discusses 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 uses 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 comments on a VIP's behaviour and actions (we think that this 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'")

Some tweets are marked as unintelligible, usually because of a lack of context, as in the following example:

"@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…")

---

[9] *Confartigianato* is an organization that represents micro and small enterprises in Italy.

If the CrowdFlower annotator considers the tweet *in-topic*, she/he proceeds to the second annotation step. This step consists in determining the *polarity*, the *sub-topics* and the presence of *irony*, according to the annotation scheme. For sentiment and irony, annotators were given instructions and examples for the following tags:

Positive: the user expresses 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 expresses 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 expresses both positive and negative opinion or sentiment;

"@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 does not express positive or negative opinion or sentiments. For example, the user reports 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 includes ironic content and conveys positive polarity. The target of the irony is not important, but there is 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 includes ironic content and conveys a negative polarity. The target of the irony is not important, but there is intent to insult or damage the target. For example:

"Vedi figliolo. un giorno tutto questo continuerai a desiderarlo."

("Look, kiddo, you will still want all this one day.")

3.4. Analysis of the *gold standard* Twitter *corpus*

Three to five independent annotations were provided for each tweet. We selected the gold label for each tweet by majority voting when at least 60% of annotators agreed. "Tw-parenthood-topic" data set consists in 2,355 *in-topic* tweets (42.3% out of the total of 5,566 submitted to CrowdFlower), 3,136 *off-topic* tweets (56.3% of the total), while for the remaining tweets there was disagreement. The proportion of *in-topic* tweets is high compared to other Twitter-based, content and opinion surveys (Ceron, Curini, Iacus. 2014). 1,545 tweets obtain a consistent gold label for sentiment polarity, presence of irony, and specific semantic areas. These tweets constitute our *gold standard corpus*[10], named "Tw-parenthood-polarity&subtopic". Where annotators disagree on polarity, irony or subtopics, we labelled the tweets as *null*.

If we take into consideration the tweets about children or parenthood (see table 1), we can see that among those *in-topic*, 28.3% were labelled positive and 21.1% negative. This gives us a general sense about Twitter feelings about happiness and parenthood. Only 12.3% of all the messages are ironic and negative irony prevails, while neutral (*none)* tweets stand at

[10] The standard collections called Gold Standard Corpora are trustworthy sets of tweets necessary for training and for the meaningful evaluation of algorithms which use annotations. In our case it is available on request.

just 8.4%. The number of mixed tweets is limited to 1.4%: the remaining 28.6% are *null*, as annotators disagreed. Even if there are slightly more purely positive posts than negative ones, ironic tweets must also be considered: most of them are negative ironic posts, balancing out the slight disparity between purely positive and purely negative tweets. Furthermore, the nature of Twitter demands short direct messages, something that makes nuances difficult. People cannot easily stand in a grey (neutral) area: about 90% of tweets shows a clear polarity, where people come down on one side or the other and express their opinions.

Our *gold standard data set* gives us some clues in support of polarity change, clues that vary depending on the sub-topic under discussion. First, it is important to note a user's status relative to a sub-topic. In figure 1 we put together tweets classified as negative together with those with negative humour and we did the same with positive and positive humour tweets. It is quite clear that positive sentiments emerged and prevail when people are talking about everyday life with children and the experience of becoming and being parents. The negative sentiments are dominant, instead, in discourses about children's future, fertility and politics and parental behaviour. Parents sometimes grumble about their children's behaviour, but they are mostly happy and proud of them.

A similar if more nuanced result emerges employing lexical sentiment resources when the polarity of messages is computed, summing up positive and negative terms, following the approach used in (Kramer 2010)[11]. The analysis of topic specification messages reveals a

[11] Linguistic normalization is performed dividing the polarity value by the number of terms in each group. As LIWC, HuLiu and EmoLex *lexica* contain a different number of positive and negative terms, the occurrences count is normalized by these values (considering the percentage). Similarly, as each category contains a different number of tweets and words, we take the number of terms in each group into account (multiplied by 100 to improve readability). In addition, as Afinn includes a single value for each term in a range from -5 to +5, the count of Afinn term occurrences is normalized by the different number of terms in each group of messages. The presence of a negation term usually reverses the polarity of the following sentiment terms in the sentence. Every time we found a negation – the list of Italian negations includes three terms: "non", "né", "nemmeno" – followed by a term belonging to our three polarity *lexica*, in a three–word range, we reverse the polarity of that term.

positive polarity for *being parents* messages, while *becoming parents* messages have a more negative polarity. Focusing on the emotion lexicon[12] (see table 2), *being parents* has a higher incidence of happy words. Messages concerning judgments and comments on the education of children (*judgment about parents' behaviour*) have a high frequency of anger and disgust terms. Anticipation is, as we might have expected, more frequent in the *becoming parents* group of messages.

Table 1 Distribution of *gold standard messages* about parenthood by *polarity*

Polarity	num	%
positive	397	28.3
negative	296	21.1
mixed	20	1.4
positive humour	64	4.6
negative humour	108	7.7
none	118	8.4
null	402	28.6
TOT	*1.405*	*100*

[12] Word-Emotion Association Lexicon (aka EmoLex) is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive). The annotations were manually done by 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 provide versions of the lexicon in over twenty languages by translating the English terms with Google Translate (July 2015). There is a translation for Italian (see http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm).

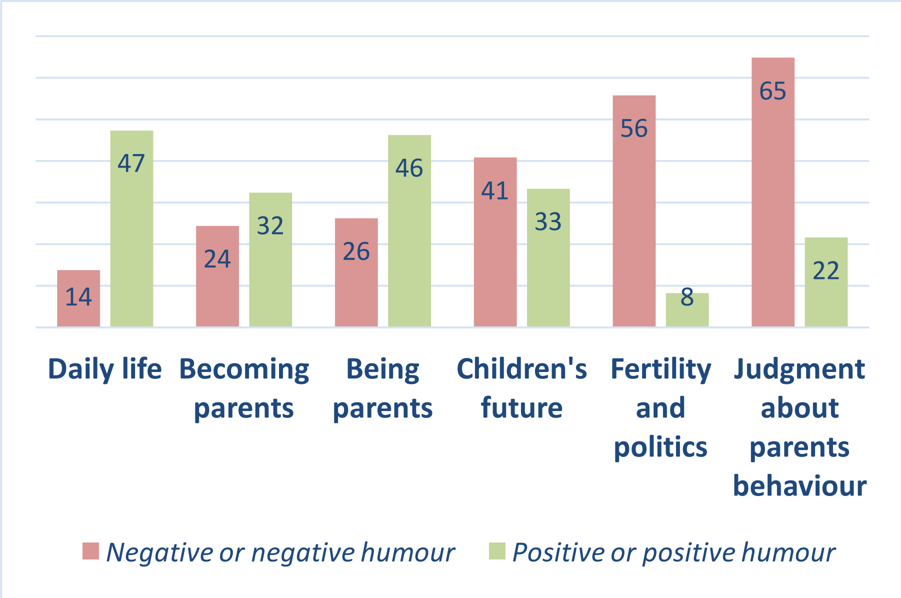

Figure 1 Prevalence (%) of negative/positive sentiments by parenthood sub-topics

Table 2 Distribution of emotions in *gold standard* messages by parenthood sub-topics

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

4. Automatic detection of sentiment polarity

To investigate the polarity of tweets in the data set "Twita-2014-parethood", we performed a two-phase experimental setting. First, we selected tweets of interest (i.e., *in-topic* ones). Second, we computed the overall sentiment for each tweet by using a machine-learning technique.

The aim of the first phase is to separate *in-topic* from *off-topic* tweets. *Off-topic* tweets are those that, even with one (or more) keywords, do not relate to fertility and parenthood. We trained a binary *Support Vector Machine* using the "Tw-parenthood-topic".

The training-set was made up of 2,355 *in-topic* tweets, whereas the 3,136 *off-topic* and the 30 *unknown* tweets were put in the same class.

We trained the model using a *bag-of-words* model and features like punctuation marks, length of the tweet (in words and characters), the frequency of hashtags, mentions, emoji, and interjections. We do not considered abbreviations, slang or swear words as they are not frequently used in our tweets. By exploiting the trained model, we automatically discriminated *in-topic* and *off-topic* tweets for the entire data set of about 2.7 million tweets. We obtained, in this way, 1,083,741 *in-topic* tweets (39.2%)[13]. To further investigate the *in-topic* category, we distinguished messages from children (in the subtopic "Being sons/daughters") from all parental tweets and tweets on parenthood (the latter group, i.e. tweets on parenthood, making up above 39% of the total, i.e., 426,036 out of the total 1,083,741). Our training-set was "Tw-parenthoodpolarity&subtopics".

To assign polarity to tweets we used a sentiment analysis system called *IRADABE*[14] (Hernandez-Farias et al. 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[15]. These are useful in detecting meaning, especially for sentiment and opinion posts, which are interrelated. The algorithm 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 with 6,448

[13] The performance of our classifier was evaluated in F-measure terms with a five-fold-cross validation being applied. We obtained a F-measure value of 0.7496.

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

[15] Such features relied on an Italian version of the following sentiment lexicons: SentiWordNet (Esuli et al. 2010); Hu&Liu (Hu and Liu 2004); AFINN (Nielsen 2011); and the Dictionary of Affect in Language (Whissell 2009).

tweets in Italian on various random topics, from politics to football; and our manually-labelled data set described in the previous section. We carried out an experiment using a five cross-validation[16] on the training set. The accuracy, i.e. the lack of mistakes or errors, obtained by IRADABE in detecting subjectivity is about 76%, in negative polarity 70% and in positive polarity above 77%. Performances seem to be fully compatible with state-of-the-art system performances for Italian (Basile et al. 2014).

The polarity results are shown in Table 4, with a prevalence of *negative* tweets (almost 50%), only 10% of *positive* ones; and a high percentage (36.1%) of *mixed* ones[17], i.e. tweets where both negative and positive attitudes are expressed. Only 4% were classed as *none* sentiment label tweets. All the tweets were classified as being subjective. Tweets about fertility and parenthood are, after all, likely to see the users expressing themselves subjectively.

Table 4. Distribution of the 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

5. The geographical distribution of positive messages

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

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

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

It would be interesting to understand, while considering sentiment polarity, whether there is a correlation with population characteristics. A particular measure of interest in this case would be the average number of children *per* women (Total Fertility Rate). In other words, are positive sentiments related to the fertility rates in different regions? In order to do this, we focus on positive messages identified by our automatic classifier, geo-referenced and aggregated by the twenty Italian regions. For these regions we relativize 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 ones. These two measures are, then, held up against the region's specific fertility rates. Obviously, the aggregation is crude, as there is substantial variation in the fertility rate within these regions. But this kind of analysis, nevertheless, sheds light on whether social-media content relates to demographic variables. We obtain 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 is reinforced by the fact that there is no correlation between the regional Crude Birth Rate (CBR; namely, 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 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 child-rearing is easier and supported by local authority's 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, that they end up having more children. Independent of the direction of the relationship, there is little doubt that the positive sentiments represent a proxy for happiness with parenthood. This state of affairs accords with the positive correlation we find between the share of positive tweets toward parenthood and the average regional level of life satisfaction (see again table 5).

Table 5. Correlations of parents' sentiment scores and 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".

6. Conclusions

In this paper we proposed a model for collecting and semantically annotating Twitter data for demographic research into parenthood and fertility. The aim was 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 in deriving meaningful classification results, and this is also the case for demographic analysis and for family research.

The development of a Twitter *corpus*, annotated with a novel semantic scheme for marking up information is the first step, and a necessary precondition for any further analysis of this kind of content. The approach produces data that have 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 sub-topics, which are sentiment targets in the relationship between parenthood and happiness.

In the tweets in which parents talk about their children or their experience of being parents, the polarity expressed is mainly positive when we consider the sentiment layers. 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 focussing in on. Our data show that the polarity of the sentiment can also be negative toward some targets. Interestingly, it emerges that parents seem to express, generally speaking, positive sentiments when they talk about daily life with children and becoming and being parents. Negative sentiments prevail, meanwhile, in tweets about children's future, fertility, politics and parental behaviour. By scrutinizing opinions on Twitter, which are posted spontaneously and often as a reaction to some emotionally-driven observation, we, thus, gain some insights into the "parenthood happiness paradox". Positive and negative feelings toward parenthood seem, in fact, to co-exist in Italians tweets. There are more positive feelings when people talk about abstract and emotion-based concepts; and more negative feelings in connection with real-world, context-based considerations.

By using the geocodes associated with (a sub-sample of) tweets, sentiments can, as others have showed 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 socio-economic 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 fertility. Though the aggregation is crude, this finding is a first for Italy.

Clearly any further information about the characteristics of users are fundamental in 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 socio-demographic traits among Twitter users. Twitter does not give explicit metadata about the age and gender of users. Nevertheless, there are now studies that propose methods for extracting this information from social-media data, thus opening the way to more ambitious future studies. Some authors have suggested getting information on the socio-demographic traits of Twitter users by manually inspecting data that they have published elsewhere, e.g. on their LinkedIn profiles. When age is not given, it could be estimated by taking into account, if present, the 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 we could have extracted the semantic information on textual content and the demographic characteristics from the data set, but we fear that the margin of error would just have been too large.

In all, we examined a dataset which 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 ISPO poll in 2012[18]), and tend to use Twitter more for getting timely news than for discussing family-related issues. Nevertheless, non-representativeness is an issue for any qualitative study. Here, 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.

[18] www.ispo.it and specifically https://www.slideshare.net/MilanIN/twitter-in-italia-ricerca-ispo-click-2012.

References

Aggarwal, C.C., Abdelzaher, T.F. (2013). Social Sensing. In C.C. Aggarwal (ed.) *Managing and Mining Sensor Data*. New York: Springer: 237-297.

Allisio, L., Mussa, V., Bosco, C., Patti, V. and Ruffo, G. (2013). Felicittà: Visualizing and estimating happiness in Italian cities from geotagged Tweets. In *Proceedings of the 1st International Workshop on Emotion and Sentiment in Social and Expressive Media, ESSEM@AI*IA*. CEUR-WS.org. 1096: 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. *Journal of Intelligenza Artificiale*, 9(1): 43–61.

Barbieri F., Basile, V., Croce, D., Nissim, M., Novielli, N. And Patti, V. (2016). Overview of the EVALITA 2016 SENTiment fertility and politics. In Proceedings of Third Italian Conference on Computational Linguistics (CLiC-it 2016) and Fifth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2016). (AILC). CEUR-WS.org: 1749.

Basile, V. and Nissim, M. (2013). Sentiment analysis on Italian tweets. In Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Atlanta, Georgia. ACL: 100–107.

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 Proceedings of Third Italian Conference on Computational Linguistics (CLiC-it 2016) and Fifth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2016). AILC. CEUR-WS.org: 1749.

Basile, V., Bolioli, A., Nissim, M., Patti, V. and Rosso, P. (2014). Overview of the Evalita 2014 SENTIment fertility and politicsarity Classification Task. In Proceedings of the 4th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITA'14). Pisa, Italy, 2014. Pisa University Press: 50-57.

Billari, F.C., D'Amuri, F., and Marcucci, J. (2013). Forecasting births using google. Paper presented at Annual Meeting of the Population Association of America, New Orleans, LA, April 11-13 2013.

Billari, F.C., Cavalli, N., Qian, E. and Weber I. (2017). Footprints of Family Change: A Study Based on Twitter, Paper presented Annual Meeting of the Population Association of America, Chicago, IL, April 27-29 2017.

Blumenstock, J. E., Gillick, D., and Eagle, N. (2010). Who's calling? Demographics of mobile phone use in Rwanda. In AAAI Spring Symposium: Artificial Intelligence for Development, 2010.

Bollen, J. and Hao, M. (2011). Twitter mood as a stock market predictor. *Computer*, 44(10): 91-94.

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.

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 Felicittà. In *Proc. of the 5th International Workshop on Emotion, Social*

Signals, Sentiment and Linked Open Data, ESSSLOD 2014, Reykjavik, Iceland, 2014. ELRA: 56-63.

Bosco, C., Patti, V. and Bolioli, A. (2015). Developing corpora for sentiment analysis: The case of irony and senti-tut (extended abstract). In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, July 25-31 2015*, Buenos Aires, Argentina, July 25-31, 2015. AAAI Press / International Joint Conferences on Artificial Intelligence: 4158-4162.

Buscaldi, D., and Hernandez-Farias, D.I. (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 Companion, WWW 2015, Florence, Italy, May 18-22, 2015 - Companion Volume: 1185–1188.

Castells, M., (2000). *The Rise of the Network Society* (2nd ed.). Cambridge, MA: Blackwell Publishers, Inc.

Ceron, A., Curini, L., and Iacus, S. M. (2014). *Social Media e Sentiment Analysis: L'evoluzione dei fenomeni sociali attraverso la Rete.* Milan: Springer-Verlag.

Cetre, S., Clark, A.E. & Senik, C. (2016). Happy People Have Children: Choice and Self-Selection into Parenthood, *European Journal of Population*, 32 (3): 445–473.

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 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM: 3267-3276.

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.

Diener, E. (2000). Subjective well-being: The science of happiness and a proposal for a national index. *American Psychologist*, 55(1): 34–43.

Esuli, A., Baccianella, S. and Sebastiani, F. (2010). SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May 2010, Valletta, Malta. ELRA.

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 Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Denver, Colorado, June 2015. ACL: 470-478.

Han, J., Pei, J., and Kamber, M. (2011). *Data mining: concepts and techniques*. Woltham, MA: Elsevier.

Hansen, T. (2012). Parenthood and Happiness: a Review of Folk Theories Versus Empirical Evidence, *Social Indicators Research*, 108(1): 29-64.

Hernandez-Farias, D. I., Buscaldi, D. and Priego-Sánchez, B. (2014). IRADABE: Adapting English Lexicons to the Italian Sentiment Polarity Classification task. In First Italian Conference on Computational Linguistics (CLiC-it 2014) and the fourth International Workshop EVALITA2014, Proceedings of the First Italian Conference on Computational Linguistics (CLiC-it 2014) and the fourth International Workshop EVALITA2014, Pisa, Italy, 2014: 75-81.

Hilbert, M. and López, P. (2011). The world's technological capacity to store, communicate, and compute information. *Science*, 332(6025): 60–65.

Hitsch, G. J., Hortaçsu, A., and Ariely, D. (2010). Matching and sorting in online dating. *The American Economic Review*, 100(1): 130-163.

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

King, G. (2011). Ensuring the Data-rich Future of the Social Sciences. *Science* 331(6018): 719-721.

Kohler, H. P., Behrman, J. R., and Skytthe, A. (2005). Partner+ Children= Happiness? The Effects of Partnerships and Fertility on Well-Being. *Population and development review* 31(3): 407-445.

Kohler, H.-P. (2012). Do children bring happiness and purpose in life?, In: Wilcox, W.B., Kaufman, E., (Eds). *Whither the child: Causes and consequences of low fertility*. Paradigm Publishers: 47–75.

Kohler, H.P., Mencarini, L. (2016). The Parenthood Happiness Puzzle: An Introduction to Special Issue. *European Journal of Population* 32 (3): 327-338.

Kramer, A. D. (2010). An unobtrusive behavioral model of gross national happiness. In Proceedings of the SIGCHI conference on human factors in computing systems, ACM: 287-290.

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, France, 2015. IEEE: 1-9

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.

Lesthaeghe, R. (2010). The unfolding story of the second demographic transition, *Population and Development Review* 36(2): 211-251.

Liu, B. (2010). *Sentiment analysis and subjectivity*. Boca: Taylor and Francis Group.

Margolis, R. and Myrskylä, M. (2011). A global perspective on happiness and fertility. *Population and Development Review* 37(1): 29-56.

Maynard, D. and Greenwood, M. (2014). Who cares about sarcastic tweets? investigating the impact of sarcasm on sentiment analysis. In Proceedings of the Ninth International Conference on Language Resources and Evaluation, Reykjavik, Iceland, May 2014. ELRA.

McDonald, P. (2013). Societal foundations for explaining fertility: gender equity. *Demographic Research* 28: 981–994.

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 (TOIT)*, 17(1): 7.

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.

Mohammad, S. M., Kiritchenko, S., and Zhu, X. (2013). NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets. In Proceedings of the seventh international workshop on Semantic Evaluation Exercises, June 2013, Atlanta, USA.

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.

Mohammad, S.M., Kiritchenko, S., Parinaz, S., Xiaodan, Z., and Cherry, C. (2016). Semeval-2016 Task 6: Detecting Stance in Tweets. Proceedings of the 10th International Workshop on Semantic Evaluation, San Diego, California, ACL.

Myrskylä, M. and Margolis, R. (2014). Happiness: Before and after the kids. *Demography 51*(5): 1843-1866.

Nakov, P., Ritter, A., Rosenthal, S., Sebastiani, F. and Stoyanov, V. (2016). SemEval-2016 Task 4: Sentiment Analysis in Twitter. Proceedings of the 10th International Workshop on Semantic Evaluation, San Diego, California. ACL.

Nielsen, F.A. (2011). A new ANEW: evaluation of a word list for sentiment analysis in microblogs. Proceedings of the ESWC 2011 Workshop on 'Making Sense of Microposts': Big things come in small packages, CEUR Workshop Proceedings, 718: 93–98.

Nissim, M., and Patti, V. (2016). Semantic aspects in sentiment analysis. In: F. Elisabetta, B. Liu, E. Messina, F. Pozzi (eds.) *Sentiment Analysis in Social Networks*. Cambridge, MA: Elsevier: 31–48.

Quercia, D., Crowcroft, J., Ellis, J. and Capra, L. (2012) Tracking "gross community happiness" from tweets. In Proceedings of the ACM Conference on Computer Supported Cooperative Work: 965–968.

Rangel, F., Rosso, P., Chugur, I., Potthast, M., Trenkmann, M., Stein, B., Verhoeven, B., Daelemans, W. (2014). Overview of the 2nd author profiling task at PAN 2014, in: L. Cappellato, N. Ferro, M. Halvey, W. Kraaij (eds.), CLEF 2014 Labs and Workshops, Notebook Papers, 1180, CEUR-WS.org: 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(1), 1.

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.

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.

Stranisci, M., Bosco, C., Patti, V. and Hernández-Farias, D.I. (2016). Annotating Sentiment and Irony in the Online Italian Political Debate on #labuonascuola. In Proceedings of the 10th edition of the Language Resources and Evaluation Conference, Portoroz (Slovenia), ELRA.

Sulis, E., Lai, M., Vinai, M., and Sanguinetti, M. (2015). Exploring Sentiment in Social Media and Official Statistics: a General Framework. In Proceedings of International Conference on Autonomous Agents and Multiagent Systems, Istanbul, Turkey, May 5, 2015. ceur.org: 96-105.

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.

Tumasjan, A., Sprenger, T.O., Sandner, P.G. and Isabell M. Welpe (2011) Predicting elections with Twitter: What 140 characters reveal about political sentiment. In Proceedings of the The International Conference on Weblogs and Social Media Barcelona, July 17 -21: 178–185.

Van de Kaa, D.J. (1987). Europe's second demographic transition. *Population Bullettin* 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 Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 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.

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

Zagheni, E. and Weber, I. (2012). You are where you E-mail: Using E-mail Data to Estimate International Migration Rates. Proceedings of the 4th Annual ACM Web Science, Evanston, IL: 348-351.

Zagheni, E., Weber, I. (2015). Demographic research with non-representative internet data, *International Journal of Manpower*, 36(1): 13-25.

Zagheni, E., Weber, I., Gummadi, K. (2017). Estimate stock of migrants using Facebook's advertising platform, *Population and Development Review*, on-line first.

Zagheni, E., Garimella, V.R.K., and Weber, I. (2014). Inferring international and internal migration patterns from twitter data. In Proceedings of the 23rd International Conference on World Wide Web. ACM: 439-444.