The Joy of Birth? Sentiment Analysis of Childbirth Narratives from Interwar Poland

Bartosz Ogórek (Institute of History, Polish Academy of Sciences)

Abstract

Historical demography often gauges fertility intentions from reproductive outcomes and their distribution over the individuals' lifespan (Timaeus & Moultrie 2008), but little is known about the purposefulness of such behaviour and individuals' attitudes towards its results in the pre-survey era. This paper uses natural language processing methods (mainly sentiment analysis) to scrutinise over 100 narratives, written by women and men from different generations and social classes, describing childbirth in the 1930s. The emotional appeal of each fragment is measured using the Nencki Affective Word List (NAWL – Riegel et al. 2015) and then modelled in *tobit* regression with robust SE to establish the associations with author characteristics. The study shows that the happiness encapsulated in the description of childbirth is clearly a generational phenomenon, decreasing with each birth order above three. The positive sentiment towards the childbirth expressed in ego-documents varies also by occupation as different social groups place the description of this event in different contexts of everyday life. Surprisingly, the most optimistic and least anxious about childbirth are those who live from hired labour (agricultural workers and labourers). The paper demonstrates a promising method for using medium and large corpora of ego-documents in historical fertility analysis, providing a potential platform for communication between quantitative and qualitative scholars. It also reduces the subjectivity of the researcher interpreting the narratives (etic approach), while retaining a considerable amount of subjectivity inherent in the source itself and its author (emic approach).

Research problem

Measuring fertility intentions has a fairly recent tradition, dating back to early family and fertility surveys in the mid-twentieth century (Bhushan and Hill 1996; Oris 2016). In the pre-survey era, the attitudes of individuals and couples towards successive births are much harder to capture. Demographers typically infer an intention to limit the number of offspring from fertility outcomes, using concepts such as spacing, stopping or postponing as signs of deliberate contraceptive practice (Timaeus & Moultrie 2008). This apparent reductionism does not allow researchers to assess whether the outcomes (number of children, sex ratio and birth order) were seen as favourable or unfavourable to the historical actors. Such attitudes can only be traced through qualitative material, which poses many methodological challenges. Probably the most common problem in the use of narrative sources for demographic research

could be described as qualitative 'cherry picking'. Researchers often look for those narrative sources (ego documents, proverbs, customs) that fit well into the mechanics revealed by the statistical models (Van de Kaa 1996). It is worth noting that while sampling on the outcome variable is an unpardonable error in quantitative analysis, it is a common practice when using qualitative material. Furthermore, it is reasonable to assume that the scientist should be looking for refutation of his findings in order to provide better insight, not a problem, rather than trying to confirm them (Popper 1972). Other sciences, especially cultural anthropology, offer many useful tools for understanding narratives (Handwerker 1986, Greenhalgh 1995, Kertzer and Fricke 1997, Szreter et al. 2004). However, the testimonies of historical actors can remain superficial or conventional, and rarely explicitly report precise decision-making processes, especially when the decisions are largely beyond their awareness (Luker 1975, Carter 1995, Angin and Szreter 2004).
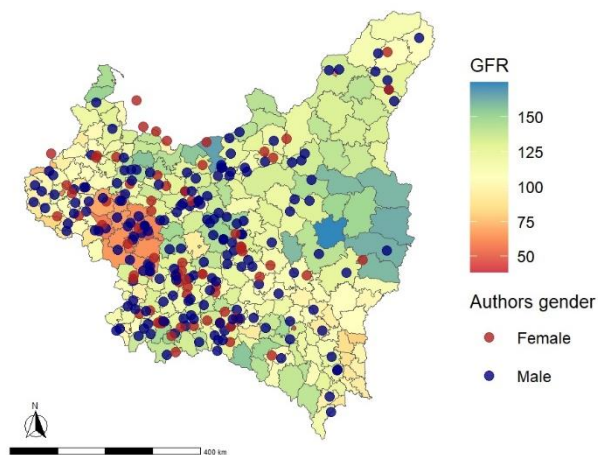
In this context the natural language processing seems well suited for capturing the implicitly expressed emotions around the facts reported by the individuals. It can also reduce the researcher subjectivity while retaining the uniqueness of each actor narrative account. This paper uses sentiment analysis to scrutinise over 100 narratives, written by women and men from different generations and social classes, describing childbirth in the 1930s to capture the emotions accompanying this event. By combining the measured sentiment score with various characteristics of authors and their offspring, the study tries to answer the very simple question – were people happy, sad, angry or fearful when they recalled and described the moment when their babies came into the world. How this emotions correlated with demographic and socio-economic characteristics of a parent as well as birth order or gender of a child? I argue that by answering these questions we can gain unusual insight into the fertility intentions of couples during the historical fertility decline.
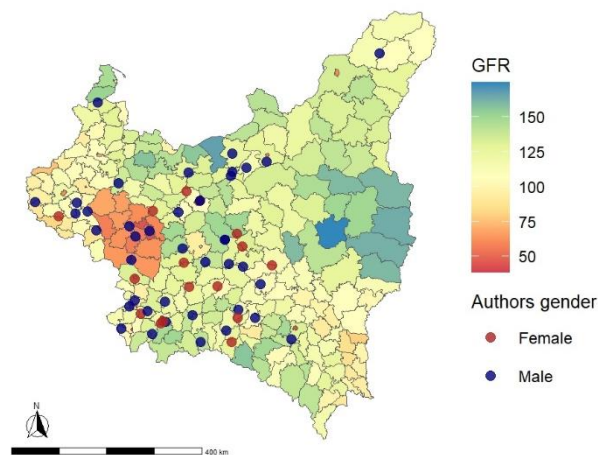
Data and methods

Poland has a long and rich tradition of memoirism popularized also among the lower social strata through memoirs competitions first organized by Florian Znaniecki and continued for the whole 20th century by various institutions (IGS, Polish radio, press). For this study I use 238 published memoirs[1], which contain 108 descriptions of the childbirth. I follow the as-if survey methodology to collect the data on the authors (Humphries 2010).

---

[1] Krzywicki (ed.), *Pamiętniki chłopów*,I-II, (Warszawa,1935); Idem (ed.), *Pamiętniki bezrobotnych*; Idem (ed.), *Pamiętniki emigrantów – Francja*, (Warszawa, 1939); Józef Chałasiński (ed.), *Młode pokolenie chłopów*, I-IV, (Warszawa, 1938); Władysław Grabski (ed.), *Wiejscy działacze społeczni*, I-II, (Warszawa, 1937); Barbara Chlabicz (ed.)., *Pamiętniki kobiet. Biografie Trzydziestolecia*, (Warszawa, 1977); Eugenia Jagiełło-Łysiowa E., *Rola przeorana, dom piękny* (Warszawa, 1975); Eadem, *Czyste wody moich uczuć* (Warszawa, 1975); Eadem., *Być w środku życia*, (Warszawa, 1975); Dyzma Gałaj, *Być matką*, (Warszawa, 1986).

General Fertility Rate in Poland, 1930-32
Memoirs' authors locations and gender

General Fertility Rate in Poland, 1930-32
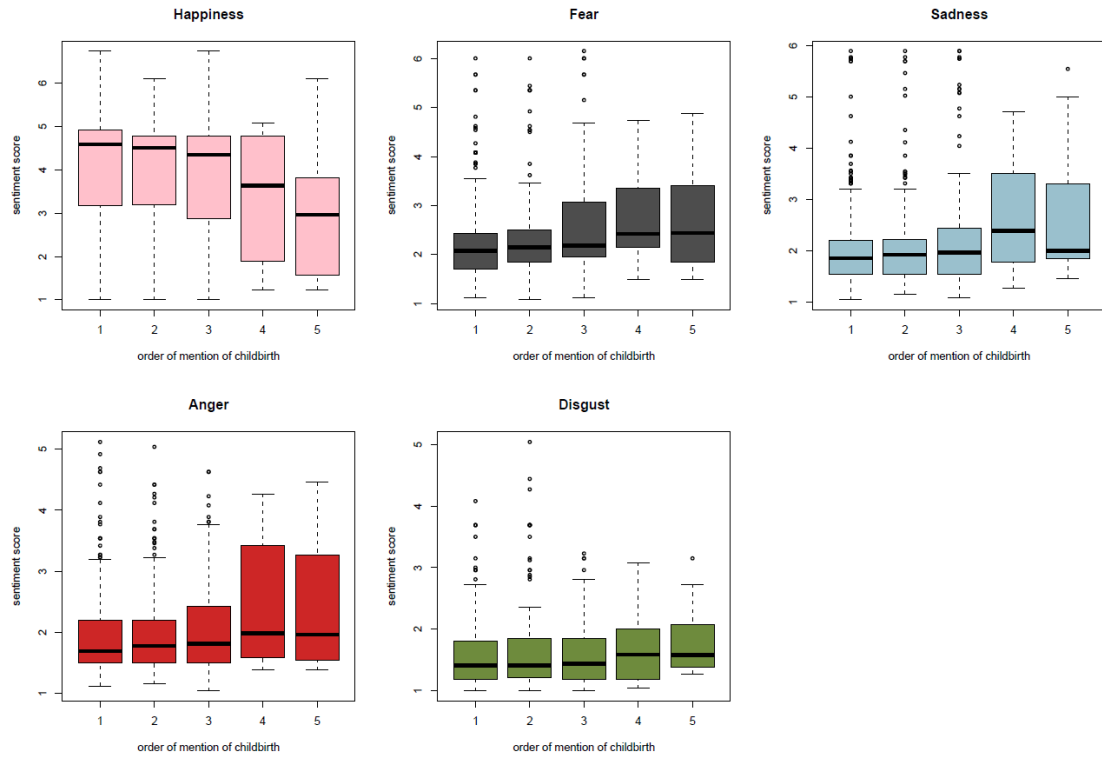Childbirth descriptions' authors locations and gender

The standard procedure of stemming and lemmatization is used with Morfologik dictionary (Polimorfologik 2.1). Sentiment analysis is possible thanks to the Nencki Affective Word List (NAWL) created in Nencki Institute of Experimental Biology of the Polish Academy of Sciences (Riegel et al. 2015). Normative ratings on a scale from 1 to 7 for happiness, fear, sadness, anger & disgust associated with 2902 Polish words were collected from 266 Polish participants (136 women and 130 men).

For modelling the relation between the sentiment scores and authors' characteristics I use tobit model for limited dependent variable with robust SE, as the scores can only fall within the range 1-7 and several authors provide more than one entry.

Preliminary results

The study shows that the happiness encapsulated in the description of childbirth is clearly a generational phenomenon, decreasing with each birth order above three. The positive sentiment towards the childbirth expressed in ego-documents varies also by occupation as different social groups place the description of this event in different contexts of everyday life. Surprisingly, the most optimistic and least anxious about childbirth are those who live from hired labour (agricultural workers and labourers).

Happiness Fear Sadness

sentiment score
order of mention of childbirth

Anger Disgust

sentiment score
order of mention of childbirth

Variable	Estimate	Std. Error	
(Intercept)	4.217	0.438	***
Gender of author (reference female)	-0.077	0.192	
Author birth cohort (reference 'before 1890')			
1890-1900	-0.331	0.191	.
after 1900	-0.384	0.229	.
Order of mention of childbirth (reference 1)			
2	-0.029	0.149	
3	-0.239	0.181	
4	-1.178	0.509	*
5	-0.822	0.243	***
Gender of child (reference female)			
male	-0.011	0.143	
not specified	-0.455	0.241	.
Author's occupation (reference craftsman)			
farmworker	0.720	0.362	*
labourer	0.441	0.403	
peasant	0.581	0.381	
white-collar	0.542	0.446	
Total number of children	-0.019	0.037	
Log(scale)	-0.363	0.085	***
Pseudo-R2	0.1141		
N	107		