Report

A corpus linguistic approach to childhood public health

Julia Dratva   **- ZHAW/ Department Gesundheit**
Noah Bubenhofer   **- Universität Zürich** (Department Angewandte Linguistik until August 2019)
Julia Krasselt   **- ZHAW/Department Angewandte Linguistik**
Suzanne Suggs   **- Università Svizzera Italiana**

Background:

Digital transformation has provided new technologies for communication and knowledge transfer, heavily used by the public to gain information about issues of relevance, such as health. Recent studies have shown that the Internet is frequented more and more by pregnant women and new parents as a source for health information [1-4]. The current generation of parents is predominantly digital native and therefore it is not surprising that they increasingly access the internet to obtain information about their child's health and development, too. This is also reflected by the increasing number of websites designed for parents [5] which offer pediatric information but also the possibility to exchange with other parents, and more generally to access a wide range of relevant topics on children, health, and parenthood [4]. Our own study on parental digital use indicated that 90% of all parents with children <2 years of age use digital media to inform themselves on child health and development [6]. About half reported to access websites for parents frequently (47%). According to these parents the most frequented website in the German speaking region is Swissmom (www.swissmom.ch). 84% of the study sample indicated it to be their preferred site [6]. Much of the online information is however complex and requires a higher literacy level [7]. Many parents may resort to peer information on websites, as they have difficulty understanding common pediatric health information [8] [9]. Nevertheless, parents present insecurity to evaluate and act upon the digital information found [10, 11].

Exploitation of digital data in public health, especially in the pediatric public health, is still rather new and little applied in Switzerland. The project aims at improving the expertise in digital data analyses using linguistic and qualitative methods in a collaborative manner. The three partners represent three different professional backgrounds, as well two different language regions in Switzerland; they thrive for a sustainable collaboration.

Objectives:

Main aim of the project was to establish a collaboration between public health, linguistics and health communication researchers at SSPH+ and identify content and method-specific interfaces between linguistics and public health.

Further, to develop a rich text corpus, the SWISSMOM Corpus, to apply novel corpus analytical methodologies thereby advancing the expertise in all partners and providing a data base for further interprofessional scientific use.

Doing so, we addressed the following questions:

a) What are the topics parents share in the Internet and are most concerned about;
b) What are vaccination specific topics and questions of parents, and are these addressed by Swiss official and socially relevant public health bodies, e.g. FOPH or Swiss Society of Pediatrics

The choice of the Swissmom website was based on a common interest in pediatric public health and a rich parental blog on the website with age-specific and health related fora.

Collaborative Agenda and Output

While some of the work is ongoing, the aims and aspiration of developing a sustainable collaboration and exchange of expertise across two SSPH+ members were very successful.

1. Workshops

Three workshops took place between the partners between July 2019 and June 2020.

Workshop 1: 14th of November 12:30 - 17:30

- Introduction to data, content and possibilities
- Introduction to linguistic techniques
- Discuss possible analyses and next steps

Workshop 2: 29th of January 2020 14:30 – 17:00
- Discuss domains/taxonomies based on topic labels
- Validation with quantitative data,
- Next analytic steps

Online Workshop 3: 8th of May 2020
- Main topic analyses: progress, next steps, paper planning
- Presentation Master thesis: current status

2. MSc. thesis
A medical MSc. student joined the group from the start and is analyzing the question on vaccination discourse in the Swissmom fora. The project thus has an outreach beyond the partners and provided a training opportunity for a public health interested medical student.
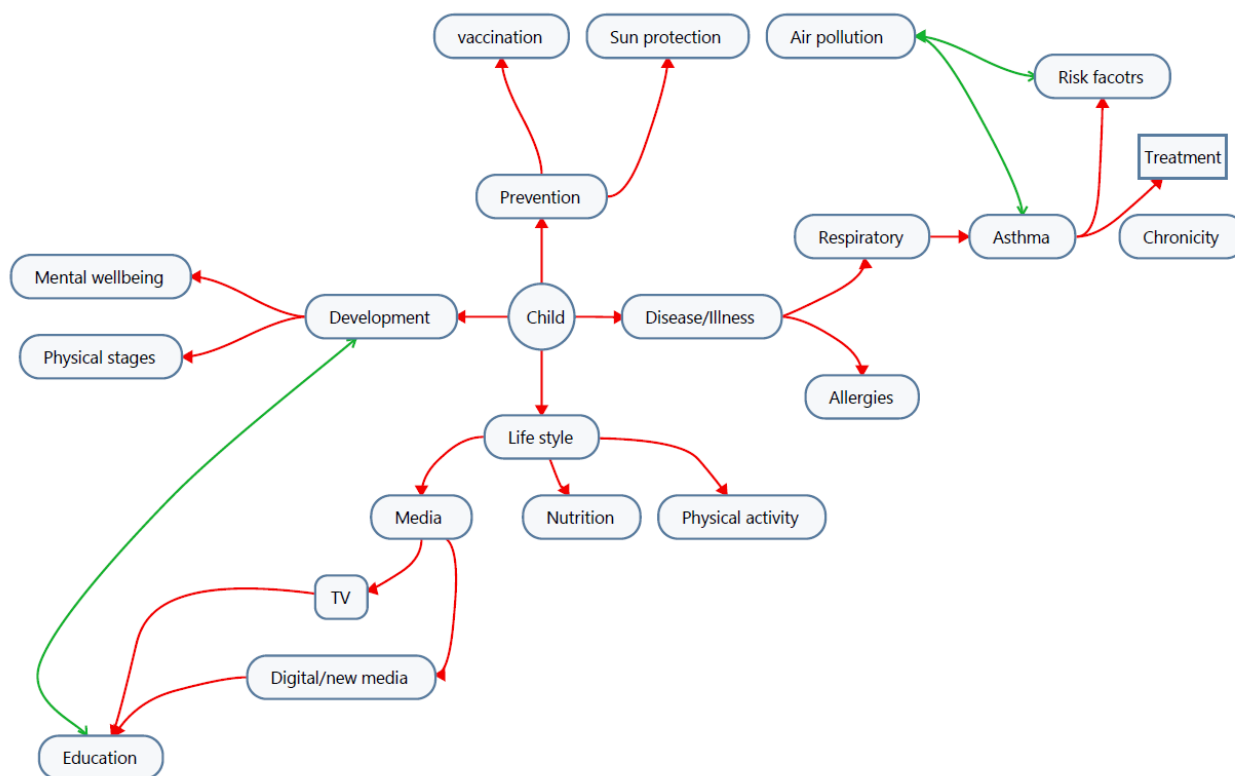
3. Corpus Swissmom
The full available content (ranging from 2014 to the moment of the crawling in September 2019) was crawled and scraped in order to get a well-structured data set with rich meta data such as the person characteristics provided by the parent. The corpus contains six subfora on child health related topics, summing up to approx. 5000 threads, which are made up of a total of approx. 105'000 individual postings. The corpus is stored on a ZHAW internal server and not accessible to the public.

Using corpus linguistic methodology, the topics, semantic fields and their distribution were analyzed. Starting point of the analyses was a data driven topic modeling to reveal thematic clusters and their distribution in the data [12]. In an interactive and interprofessional process the thematic clusters were discussed, named and evaluated with qualitative iterative method, and in a stepwise process we derived an understanding of how the methodology can provide relevance, variability, seasonality and temporality of the topics parents address.

4. Concept of visualizing qualitative results and linguistic topics for future research
Using linguistic and public health concepts we identified the following approach as useful in the analyses and visualization of topical analyses. Figure 1 shows how topics can be attributed to certain areas of relevance in health and public health (deductive approach) and how the topics expand, develop and interact (inductive approach, red arrows). The model also shows the overlap, respectively the net of themes across the different topics and areas (green arrows).

PH_Topic_Mapping_v0.mmap - 17.07.2020 - Win10User

Figure 1: Preliminary deductive and inductive mapping of topics of the Swissmom Corpus

5. Successful collaboration: grant application COVIDisc

Building on the previous exchange and resulting understanding we were rapidly able to put together a convincing SNF project for the COVID-19 call in March including further linguistic colleagues. COVIDisc (www.zhaw.ch/covidisc) investigates the COVID-19 discourse and perception of the discourse in young adults (15 – 34 years) using similar methodologies and analytic approaches.

6. Scientific Paper

The MSc. Thesis will be written up as a scientific paper coauthored by the group. Analyses are ongoing.

Conclusion:

We thank SSPH+ für the funding to establish this working group on linguistic public health research. The exchange has been profitable and educational.

References:

1. Lupton, D. and S. Pedersen, *An Australian survey of women's use of pregnancy and parenting apps.* Women Birth, 2016. **29**(4): p. 368-75.
2. McDaniel, B.T., S.M. Coyne, and E.K. Holmes, *New mothers and media use: Associations between blogging, social networking, and maternal well-being.* Maternal and child health journal, 2012. **16**(7): p. 1509-1517.
3. Wainstein, B.K., et al., *Use of the Internet by parents of paediatric patients.* Journal of paediatrics and child health, 2006. **42**(9): p. 528-532.
4. Bernhardt, J.M. and E.M. Felter, *Online pediatric information seeking among mothers of young children: results from a qualitative study using focus groups.* Journal of medical Internet research, 2004. **6**(1).
5. Plantin, L. and K. Daneback, *Parenthood, information and support on the internet. A literature review of research on parents and professionals online.* BMC family practice, 2009. **10**(1): p. 34.
6. Jaks, R., et al., *Parental digital health information seeking behavior in Switzerland: a cross-sectional study.* BMC Public Health, 2019. **19**(1): p. 225.
7. Mcinnes, N. and B.J. Haglund, *Readability of online health information: implications for health literacy.* Informatics for health and social care, 2011. **36**(4): p. 173-189.
8. Kumar, D., et al., *Parental understanding of infant health information: health literacy, numeracy, and the Parental Health Literacy Activities Test (PHLAT).* Acad Pediatr, 2010. **10**(5): p. 309-16.
9. Yin, H.S., et al., *Assessment of health literacy and numeracy among Spanish-Speaking parents of young children: validation of the Spanish Parental Health Literacy Activities Test (PHLAT Spanish).* Academic pediatrics, 2012. **12**(1): p. 68-74.
10. Plantin, L. and K. Daneback, *Parenthood, information and support on the internet. A literature review of research on parents and professionals online.* BMC Family Practice, 2009. **10**(34).
11. Jaks, R., et al., *Parental digital health information seeking behavior in Switzerland: a cross-sectional study* BMC Public Health, accepted 2019.02.08 - in print, 2019.
12. Blei, D., A. Ng, and M. Jordan, *Latent dirichlet allocation.* Journal of Machine Learning Research, 2003. **3**: p. 993 - 1022.
13. Mikolov, T., et al., *Efficient Estimation of Word Representations in Vector Space.* arXiv:1301.3781 [cs], 2013.