Self-Improvement in the Age of Big Data

A language and content analysis of self-improvement apps

Phd. Candidate: Beáta-Hajnalka Berecki

Supervisor: Professor Hanna Orsolya Vincze, Ph.D.

Doctoral School of Political and Communication Sciences

Department: Communication Sciences

Babeş-Bolyai University

This thesis aims to analyze self-improvement applications (which are part of the wider phenomenon of self-improvement and the quantified self) through their broader and linguistic specificity. It also investigates the impact of the digital medium on self-improvement practices.

The reason for choosing the topic of self-improvement is that in recent years, due to technological advancements, digital data that people can track about themselves has taken an exponential route both in terms of variety and quantity. From an individual's or user's perspective, the data revolution offers the opportunity to quantify more and more aspects of his life: gadgets and devices allow tracking of the hours slept, calories eaten, steps walked, minutes meditated, and hours concentrated. Data can thus give more accurate insights into the world - and the individual's life. In other words, within self-improvement, there was a significant shift in the ability to gain self-knowledge through data analysis in the last decade. Previously, it was challenging to track and measure daily activities, interactions, and emotions objectively. But as a result of technological developments, users gained new ways to leverage data and technology to gain self-knowledge and monitor their progress. Integrating data-driven approaches into self-improvement practices comes with the possibility of enhanced self-awareness through numbers, more practical goal setting, and a more specific way of tracking the cultivation of habits.

Currently, a big variety and number of self-improvement applications are available for download. These apps are designed with the aim to help users improve their mental and physical well-being, cultivate positive mind states, and reduce the harmful impact of negative

emotions in their life; furthermore, they often aim to help cultivate new behaviors, develop new habits, or focus on enhancing productivity.

In the objectives part of the thesis, the research questions were listed and grouped into two main categories: questions related to the broader purpose of the apps (the user needs it addresses, promised outcomes) based on textual descriptions, and questions related to the language and discourse specific to self-improvement apps (prevalent themes, linguistic features, patterns, persuasive strategies) - also compared to that of self-improvement books.

To gain insights, the study quantitatively analyzes the texts of prominent self-improvement smartphone apps, including app descriptions and user reviews, using the Linguistic Inquiry and Word Count (LIWC) software. App descriptions and user reviews are valuable sources of language data that, in large numbers, can be analyzed using techniques such as natural language processing (NLP). These types of analyses provide quantitative insights into the language used in self-improvement apps, such as the frequency of certain words or phrases, the sentiment of user reviews, and the prevalence of different marketing messages. LIWC has been widely used in various fields, including psychology, communication, and marketing, to understand how language reflects psychological and social states.

The present study aims to contribute to a better understanding of the language and mechanisms of self-improvement apps, as well as the role of big data and technology in shaping these tools by conducting discourse analysis. By dissecting language patterns and mechanisms, the research aims to shed light on different communicative strategies used in self-improvement apps. In recent years several studies and books mentioned in this thesis have explored the topic of self-improvement and its discourses; these focused either on books, ideas, or concepts related to it. The technologies present in self-improvement apps, the effects, and the philosophy behind has also been researched; however, the communicative aspects have not received much attention so far.

In the theoretical part of the thesis, a short history of self-improvement concepts is presented together with the current state of the self-improvement industry. Important concepts related to self-improvement are overviewed, including self-tracking, lifelogging, data doubles, datafication, and personal analytics.

This thesis also overviews important studies related to self-improvement books (topics, purposes, language, register), which, although a different medium, are closest to self-improvement apps in terms of language/discourse.

In the theoretical part, Marshall McLuhan's medium theory[1] is presented, as the thesis builds on this theory: the content of self-improvement can also be analyzed through the medium (mobile apps) and compared to the more traditional genre of self-improvement books (drawing on the concepts of old and new media/print vs. digital media). Another theory upon which the present thesis builds is the media effects theory, more specifically the preference-based effects models as described by Cacciatore et al. (2016)[2]. Current technologies and trends in self-improvement are also presented, focusing on gamification and persuasive technology, as well as interdisciplinary self-improvement projects at the intersection of art and science.

The thesis also presents self-improvement apps and their communicative aspects, overviewing the dominant communication styles according to previous research, as well as important aspects related to user experience, and questions related to privacy and security.

In the Methodology part, the app selection process and criteria are presented. The selection process resulted in N=92 apps, from Google Play Store and Apple App Store. From each app ten reviews were exported, the ones marked as most useful, resulting in N=870 reviews (some apps did not have reviews available at the time of collection - according to the guidelines, this means that the app has a new version). Another source of data were the app descriptions themselves (N=92).

The methodology devised for this study aimed to leverage the power of NLP techniques, which allow the processing of larger volumes of data, and provide insights into the discourse/content surrounding self-improvement apps. In some cases, this was also complemented with manual analysis. The tool selected is called Linguistic Inquiry and Word Count (LIWC); this was developed based on the theory that language reflects underlying psychological and cognitive processes. It is a dictionary-based approach that is based on lexical semantics, studying the meaning of words and their relationships as described by

---

[1] McLuhan, M. (1967): *The Medium is the Massage*, Republished edition (2001), Gingko Press.
[2] Cacciatore, M., Scheufele D. A., Iyengar, S. (2016): The End of Framing as we Know it ... and the Future of Media Effects, *Mass Communication and Society*, 19, 7–23, https://www.tandfonline.com/doi/full/10.1080/15205436.2015.1068811

Tausczik and Pennebaker (2010).[3] The LIWC dictionary contains word categories that capture various aspects of human experiences, such as emotions, social interactions, cognitive processes, and linguistic style. The thesis also builds on the research and findings of Tausczik and Pennebaker (2010) focusing on the social and psychological meaning of words (such as pronouns, verb tenses, indicators of dominance, and status dynamics).

In the results part, the findings related to the research questions are presented, investigating the broader purpose of self-improvement apps, topic and discourse characteristics, and similarities with and differences from self-help books. LIWC was used for all parts of the results, and it was combined with manual analysis when required (e.g., in the case of the main app functions). The broader purpose of the apps discusses the user needs that are addressed by self-improvement apps and the proposed benefits of these apps based on texts. To accomplish this, a manual analysis was conducted by categorizing apps based on the benefits outlined in their respective app descriptions. For this, the common elements of these descriptions were identified. After the analysis of description structures, the apps included in the study were categorized based on the main perceived benefits they offer to users in their descriptions. This categorization process involved a manual examination of the app descriptions, focusing on identifying the main keywords that highlight the primary functionalities and features. The main app categories identified were: sleep tracking, goal setting, meditation/mindfulness, relaxation, journaling/cognitive diaries, health tracking, mood tracking, affirmations, self-care/mental health.

Another objective of this study was to explore whether different types of self-improvement texts reflect similar topics. In order to address this objective, the main functions of self-improvement books (based on previous research) were also compared to those found in self-improvement apps. By comparing these two mediums, the aim was to identify any potential overlap or divergence in the thematic content they offer. The results reveal that, while some core concepts remain similar, the shift to self-improvement apps has brought forth new features and functionalities that shape the way these topics are presented and interacted with. The wording and framing within the app context often prioritize actionable steps and quantifiable progress, aligning with the medium's interactive and goal-oriented nature.

---

[3] Tausczik, Y. R., & Pennebaker, J. W. (2010): The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods, Journal of Language and Social Psychology,

Word count analysis in both app descriptions and user reviews was performed with LIWC-22. By analyzing them separately, unique vocabulary and language patterns present in each category were identified. A comparative analysis was also conducted to determine if there are any shared words between the app descriptions and user reviews.

The topics prevalent in app descriptions were surfaced with the help of LIWC; these results show those app descriptions are keyword-heavy texts that align with common themes and goals of self-improvement, which often focus on achieving greater well-being and personal growth. In user reviews the most common words can be grouped into the following topics: positive sentiments, emphasis on time and daily routines, focus on specific self-improvement goals, emphasis on tracking and monitoring progress.

Checking the most used words in these two types of texts, it can be observed that there's a consistent percentage that is present in both app descriptions and user reviews. The shared word usage between app descriptions and user reviews suggests a consistency in the themes and topics addressed by self-improvement apps. It indicates that users' experiences and feedback align with the app's intended focus and objectives as described in the promotional content.

In self-improvement books, according to the results of Koay's (2019)[4] analysis, the *you* pronoun had a high frequency compared to other genres (Corpus of Contemporary American English, British National Corpus). In the case of app descriptions, *you* and *your* climbed even higher on the list - and the frequency was already unusually high in the case of the book corpus, suggesting that this word is an important concept of both self-improvement apps and self-help books. The high frequency of you and your in the app descriptions indicates a reader-oriented approach (rather than solely topic orientation).

The function of the pronoun *you* from a grammatical perspective also shows differences between self-improvement books and apps: in books, the most common occurrences of the *you* pronoun were in the subject position, suggesting that individuals are encouraged to actively take charge of their lives to strive for personal growth and improvement. In apps, the results of the manual grammatical analysis show that the proportion of *you* appearing in the subject position decreases; instead, it becomes common to find constructions where *you* is the

[4] Koay, J. (2019): *Persuasion in Self-improvement Books*, Postdisciplinary Studies in Discourse, Palgrave Pivot, Cham. https://doi.org/10.1007/978-3-030-12149-5_3

object from a grammatical perspective. This change suggests a shift in focus from solely emphasizing the reader's agency to a more collaborative or supportive relationship between the user and the app. What can be deducted is that self-improvement apps aim to position themselves as active facilitators in the user's self-improvement journey.

The results highlight that in the case of self-improvement app descriptions, questions play a similar role to those found in self-improvement books (stimulating readers' reflection and introspection, introducing new ideas). They are utilized to engage potential users, encourage active participation, introduce new ideas, and prompt reflection.

Credibility strategies show differences in self-improvement books and apps; while in books creating a credible persona and an authoritative persona are the most common strategies, in the case of apps, credibility markers change as the medium and format change. While self-help books are narrative-driven and long-form content, apps provide an interactive and technology-driven experience. App descriptions (and apps) do not have a known author and rely on other factors to establish credibility. The credibility strategies identified are: including user reviews in the description, social proofs, and transparent privacy policies.

Looking at the persuasive strategies applied in app descriptions, it can be observed that self-monitoring is the most employed strategy, followed by simulation and rehearsal, and personalization and tailoring on the third place. Reduction and tunneling are the least employed approaches. Applied to user reviews, we get the same list: self-monitoring is the most prevalent strategy, followed by simulation and rehearsal, personalization and tailoring, tunneling, and reduction - meaning that user reviews echo the persuasive strategies applied in app descriptions.

A LIWC analysis was performed on both app descriptions and user reviews, generating scores for all available LIWC variables, including the four main variables: Clout, Analytical thinking, Authenticity, and Tone. The Analytical thinking variable in LIWC-22 measures the extent to which a text contains words associated with analytical thinking, including words related to cause and effect, insight, and certainty. The presence of Clout words within a text indicates that the author is communicating with a sense of confidence and expertise. The Authenticity score refers to the extent to which a text reflects genuine personal experiences and emotions and identifies written content that is personal and sincere in nature. Emotional tone is also among the four primary summary variables used to assess the tone of written

messages, measuring the extent to which a text contains words associated with different emotions, including positive and negative emotions.

Comparing the Clout scores of self-improvement texts (app descriptions and user reviews) to the average Clout scores of other texts (Twitter, blogs, conversations, New York Times articles, others), it became apparent that self-improvement texts have more extreme values: app descriptions are on the high extreme of the Clout variable, while User reviews are on the Low extreme of the Clout variable, compared to various other text types as well, from microblogging to heritage media. This disparity suggests that self-improvement texts exhibit a more polarized distribution in terms of perceived influence and authority. When compared to other texts, it can be observed that user reviews have the highest Authenticity score from all text types and are closest to the Authenticity score measured in Blogs and Conversations. App descriptions, in the other hand, have the lowest authenticity score, along with texts from the New York Times. The Analytic score of app descriptions is closest to texts from the New York Times; the professionally-written nature of both types of texts implies intentional consideration of language use and structure. User reviews are closest to Blogs from the Analytical score perspective; while user reviews are generated by individuals sharing their opinions and experiences, blogs are typically written by individuals who express their thoughts and insights on various topics. The similarity in analytic scores between user reviews and blogs indicates that both text types may involve a certain degree of critical thinking, reasoning, and language complexity. When comparing to other texts (Twitter, conversations, blogs, NYT, others), it can be observed that both app descriptions and user reviews of self-improvement apps score higher on Tone than the others. The higher Tone scores in both app descriptions and user reviews indicate a prevalent language style in self-improvement texts that is focused on positivity. This aligns with the goal of self-improvement apps to inspire and uplift.

The thesis ends with conclusions part, summing up the implications of the thesis as well as its limitations (no validation of the authenticity of the reviews, limitations of automated content analysis), and showing some future research directions (creating a self-improvement dictionary in LIWC, supplementing the research with qualitative methods, interdisciplinary approaches).

Keywords: self-improvement apps, quantified self, self-tracking, LIWC software, content analysis, app descriptions, user reviews

Table of contents