

Article

Large-Scale Analysis of the Medical Discourse on Rheumatoid Arthritis: Complementing with AI a Socio-Anthropologic Analysis

Mario Santoro ^{*,†}  and Christine Nardini ^{*,†} 

Istituto per le Applicazioni del Calcolo, Consiglio Nazionale delle Ricerche, Via dei Taurini, 19, 00185 Roma, Italy

* Correspondence: mario-santoro@cnr.it (M.S.); christine.nardini@cnr.it (C.N.)

† These authors contributed equally to this work.

Abstract

The medical discourse entails the analysis of the modalities, which are far from unbiased, by which hypotheses and results are laid out in the dissemination of findings in scientific publications. This gives different emphases on the background, relevance, robustness, and assumptions that the audience takes for granted. This concept is extensively studied in socio-anthropology. However, it remains generally overlooked within the scientific community conducting the research. Yet, analyzing the discourse is crucial for several reasons: to frame policies that take into account an appropriately large screen of medical opportunities; to avoid overseeing promising but less walked paths; to grasp different types of representations of diseases, therapies, patients, and other stakeholders; to understand how these terms are conditioned by time and culture. While socio-anthropologists traditionally use manual curation methods—limited by the lengthy process—machine learning and AI may offer complementary tools to explore the vastness of an ever-growing body of medical literature. In this work, we propose a pipeline for the analysis of the medical discourse on the therapeutic approaches to rheumatoid arthritis using *topic modeling* and *transformer*-based emotion and sentiment analysis, overall offering complementary insights to previous curation.

Keywords: medical discourse; large language models; topic modeling; AI; rheumatoid arthritis; disease modifying anti-rheumatic drug; physical therapies; vagus nerve stimulation



Academic Editor: Rüdiger Pryss

Received: 15 April 2025

Revised: 27 October 2025

Accepted: 19 November 2025

Published: 23 November 2025

Citation: Santoro, M.; Nardini, C. Large-Scale Analysis of the Medical Discourse on Rheumatoid Arthritis: Complementing with AI a Socio-Anthropologic Analysis. *J* **2025**, *8*, 45. <https://doi.org/10.3390/j8040045>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The medical discourse [1,2] highlights an aspect of communication generally neglected by the community producing the scientific results. In particular, it entails the analysis of the modalities by which hypotheses and results are presented when communicating and disseminating findings, a process that can introduce conscious or unconscious bias. Exploration of the discourse is therefore important for decision-making based upon the scientific evidence presented in publications. This, in fact, has also an impact on policy definitions, funding orientation and it is important to better understand how scientific writing is conditioned by contingencies like time and culture.

We here propose the analysis of the medical discourse on therapies for the treatment of rheumatoid arthritis (RA), an exemplar inflammatory, autoimmune, non-communicable disease (NCD). NCDs are among the most deadly and burdensome maladies that affect societies worldwide [3], and for this reason, NCDs' control is among the objectives of

the WHO Sustainable Development Goal (SDG) 3.4. To perform our analysis, we explore a spectrum of clinical approaches divided into three large categories: pharmacological (PHA) and non-pharmacological, with the latter including unstandardized (USTD) and experimental (EXP) approaches. In short, PHA therapies directly target the immune system response (innate, adaptive, and trained) while USTD and EXP consider other therapeutic modulators, i.e., the autonomic nervous system (ANS), the gut intestinal (GI) microbiome, and the elicitation of wound healing (WH) processes, as proposed in [4].

PHA approaches are by far the most assessed and recognized category, i.e., the gold standard in widespread practice. Leading scientific societies endorse this approach, including the American College of Rheumatology (ACR), as well as the European and Asia-Pacific League for Rheumatoid Arthritis (EULAR and APLAR, respectively). PHA targets the immune system response and includes a large set of drugs controlling inflammatory symptoms. These drugs fall into several subcategories: non-steroidal anti-inflammatory drugs (NSAIDs); paracetamol or morphine(-derived) analgesics; corticosteroids to counteract the degeneracy of the disease; disease-modifying anti-rheumatic drugs (DMARDs) to interfere or block the pro-inflammatory endogenous activity [5–7] in a generic or targeted manner (conventional and biologic cDMARDs and bDMARDs, respectively).

Experimental (EXP) therapies are mostly concerned with the attempts to control inflammation by interacting with the ANS (mostly via the vagus nerve stimulation—VNS) and the GI microbiome, via diet, nutraceuticals, antibiotics, and fecal microbial transplant—FMT, as well as antibiotics.

Finally, unstandardized approaches (USTD) include wound-healing modulators that are represented by physical stimuli (optic, mechanic, magnetic, and electric). This approach, although extremely promising under certain medical specializations (see, for instance, RA [8–10] but also Alzheimer’s Disease [11,12]) continues to suffer from limited standardization, evidence-based research, guidelines, funding, and overall public recognition [13].

Our current work builds on recent research [14] where we performed an analysis of the medical discourse based on the manual curation of two exemplar articles per subcategory of therapy, totaling 28 articles (see Section 2.2 and File S1). In that work, our conclusions highlighted some interesting, non-trivial patterns, including, for instance, the lack of awareness among USTD practitioners of the scientific evidence underlying their therapies, as well as the bias in approaching the same therapies (electrostimulations), depending on the traditional versus modern medical perspective adopted. However, one of the main limitations of that study was the reduced amount of literature analyzed (for details, see Section 2.1 and Tables S1 and S2).

To overcome this, we exploit here two main approaches that are effective in performing textual analyses on a large amount of data. First, *structural topic modeling* (STM, [15,16]) is chosen to identify the main themes discussed in the literature. STM automatically uncovers latent thematic structures within the corpus and offers a more sophisticated approach compared to keywords identification. STM was created for social sciences [16,17] and successfully used in biomedical literature analyses [18–20]. Although the approach is fully mathematically controlled, human-in-the-loop can be included in data pre-processing to guarantee that results are also clearly interpretable, as largely discussed by STM authors in an application of political sciences to open-ended surveys [17].

Second, we use *transformers* [21] to analyze the topics’ content. In this study, we used transformers specifically fine-tuned on translation, language, and emotion recognition. Our approach improved transformers’ explainability [22] by applying this technology to granular data, i.e. to single sentences. This allows interpretation of smaller chunks of information, making them less prone to ambiguity. We consider this approach very well adapted to text in English (where short sentences are preferred) and scientific literature,

where there is less of a need to construct wavy emotional flows, and where sentence-granularity in the analysis does not affect the overall understanding of the paragraphs or sections.

In particular, we developed an original pipeline based on Semanticase (<http://www.semanticcase.it> (accessed on 26 October 2025)) developed as a commercial tool (PiazzaCopernico, 2023) by author MS. Details about its implementation can be found in Supplementary Materials File S1 as they have not been discussed elsewhere.

Thanks to this approach, we can explore the totality of the material selected in [14], i.e., 204 articles (the *corpus*, Section 2 and Table S3), to address the following research question: “How does the medical discourse on rheumatoid arthritis therapies differs, in thematic content and emotional framing, between pharmacological (PHA), experimental (EXP), and unstandardised (USTD) approaches?”

Overall, this work complements and integrates the previous study. In particular, the current computational approach could successfully identify several subcategories not given a priori (i.e., not given as variables in the dataset), such as cDMARD and bDMARDs. Additionally, this approach was able to merge subcategories that clinical practice keeps separate, namely electroacupuncture and VNS. Furthermore, emotions analysis discovered a similarity between PHA and EXP versus USTD, a finding consistent with the previous analysis in [14] and justified by *temporal legitimation* (i.e., therapies recognition for PHA and EXP is based on the most recent scientific discoveries, while USTDs take their strength from a long dated, empirical, and historical practice). The combination of these results suggests that, in the medical discourse, approval based on temporal legitimation has a strong emotional component, a bias that may negatively affect therapeutic opportunities for patients.

2. Materials and Methods

2.1. Materials

Materials are described extensively in [14] and the accompanying supplementary materials. They have been used here as is to guarantee comparability of the findings, and the detailed process of selection is also reported for ease of consultation in File S1. Briefly, the selection emerges from a literature search on PubMed performed by three-tier queries using Medical Subject Headings (MeSH [23]). The first tier is a general query on Arthritis (Rheumatoid/therapy”[MeSH] AND Review” [Publication Type]”) and the latter tiers specialize (level I) in the mechanism modulating inflammation [4] and (level II) their physical nature. Level I includes the GI microbiome, the ANS, and wound healing. Level II is taken from the list of stimuli (optic, mechanical, electric, magnetic) routinely used in wound-healing testing [24]. As covariates in the dataset we used the three categories EXP, USTD and PHA, thus dividing the articles into 3 groups.

From the original corpus (500 articles along with their PubMed Identifiers, PMIDs) we have removed 31 duplicates and retained only the 347 articles with abstracts (to avoid publications other than scientific articles). Finally, due to varying levels of accessibility, only 204 articles (58.8% of the originally identified PMIDs) could be fully downloaded and converted into plain text from the original PDF format using Semanticase. The list of 500 articles along with PubMed queries and all categories and subcategories can be found in Table S1. The list of the 28 manually curated articles in [14] in Table S2, and the list of the final 204 PMIDs analyzed in this article can be found in Table S3.

In the following section, we present an overview of the methods adopted (Section 2.2), with further details reported in Supporting Materials File S1. Results pertain to automatic topic modeling (Section 3.1), with a discussion on the identification, semantics, and relations among the topics. We then approach the sentiment analysis (Section 3.2), which is useful in

gaining a general sense of the differences that characterize the three categories under this unusual perspective for the scientific literature. This section also acts as an introduction to the more sophisticated emotions analysis (Section 3.3), where we focus on three dominant emotions: approval, disapproval, and neutral.

2.2. Methods

2.2.1. Sentence Splitting

Our methodological pipeline decomposes every document into individual sentences to facilitate a more granular text handling and transparent analysis. We used for this PySBD [25,26] a lightweight, rule-based Python 3.9 library for Sentence Boundary Disambiguation (SBD) that works out-of-the-box without additional training data or external dependencies. This approach is now (contextually to this work) part of the Semanticase toolbox.

2.2.2. Corpus Pre-Processing

Corpus creation for any specific domain necessitates an iterative approach to refine the data and to progressively identify inconsistencies. This step includes the management of the following: (i) multilingual articles; (ii) frequent but non-meaningful acronyms or other types of symbols; (iii) synonyms that inflate the statistics on words' frequency and, from there, the inference of semantics.

Regarding multilingual articles, the sentence-level strategy is crucial to identifying all and only sentences that are not in English. Identification and translation are both performed via transformer models in Semanticase, and finally, sentences are reassembled in the original order to reconstruct the entire document in English.

For acronyms, symbols, and synonyms, the corpus undergoes pre-processing to optimize data quality. This starts with text normalization by converting all text to lowercase. Next, a cleaning routine based on regular expressions removes extraneous characters and symbols. Finally, to improve domain-specific understanding, human-in-the-loop synonyms identification and substitution is implemented. Similarly, meaningless symbols and acronyms are manually removed.

2.2.3. Words Statistics

We employed a statistical technique known as *keyness* analysis [27] to explore the distribution of words throughout the corpus and identify terms characteristic of specific categories. Keyness analysis quantifies the differential occurrence of words between a target group and a reference group, with a χ^2 -statistic assessing the statistical significance (Yates' correction applied to mitigate bias introduced by small sample sizes in individual documents).

This study's target group corresponds to one category (i.e., PHA, EXP, USTD) while the reference includes all other documents combined. Words that exhibit statistically significant higher frequency within a particular category potentially reflect the thematic focus of that category. For a more in-depth exploration of keyness analysis, we refer the readers to [28].

2.2.4. Topic Modeling

Our study employs topic modeling to uncover latent thematic structures within the corpus. The categories EXP, PHA, and USTD are the covariates, and the topic characterization relies on two statistics.

The first, topic *prevalence*, refers to the degree to which a particular topic is represented within a document or across the entire corpus. It measures how frequently or prominently a topic appears in a text, and using STM, we measure topic by topic whether there is a stronger or weaker association with a specific category.

The second, *topic content*, deals with the specific N-grams (set of N consecutive words, with $N = 1, 2, 3$) that define a particular topic. It represents the vocabulary that characterizes a concept or a theme. STM defines the probability for each word in the topic to be associated to each covariate (PHA, EXP, USTD). This allows us to describe a situation where different words or N-grams (i.e., a specific vocabulary) are characteristic and descriptive of a given topic, category-wise. The algorithm automatically identifies the number of topics in Semanticase, within a human-given range, here set to [10–19].

After topic modeling, Semanticase computes *inter-topic correlations* to construct an aggregative hierarchical clustering tree that visually represents the relations among (sub)topics.

2.2.5. Sentiment Landscape Through Sentence-Level Analysis

We employ the sentence splitting defined above and the *sentiment* analysis framework within Semanticase. This pipeline relies on the `sentimentr` R package version 2.9.0 [29] to categorize words and phrases into positive, negative or neutral. The package uses pre-defined sentiment lexica and includes rules to account for valence shifters such as negations, intensifiers, and adversative conjunctions.

2.2.6. Emotional Landscape Through Sentence-Level Analysis

We exploit the emotion analysis (more nuanced than the tripartite—neutral, positive, negative—sentiment analysis) to better understand the feelings landscape within the corpus at the sentence level. We used a pre-trained classification pipeline built upon the *Huggingface* transformer [30] *roberta-base-go-emotions* model [31] that, based on a comprehensive GO Emotions dataset [32], allows the recognition of 27 distinct emotional states plus a neutral category (Table S4).

The model outputs a predicted probability score for each sentence. As a result, a sentence may have positive probabilities for more than one emotion reflecting its inherent polysemy. Assuming that a single sentence is unlikely to harbor many strong emotions concurrently, we establish a threshold of 0.3, limiting to a maximum of three the number of emotions can be incorporated into one sentence (scores > 0.3).

Additionally, the *sent* index is computed to enable comparative analysis across papers' categories. This index assigns a unique, sequential integer to each sentence within a paper, normalized by the total number of sentences in that paper. To visualize the flow of emotions, we plot the smooth profile (`geom_smooth` in `ggplot2` [33]) of *sent* and *emotion* scores on the x- and y-axis, respectively, and each article is presented separately using the category as color.

3. Results and Discussion

3.1. Topic Modeling

The analysis identifies 10 topics (indexed 1–10), whose hierarchical structure, in Figure 1, is almost fully nested. Each topic is included in the topic above (Topic 3 is included in Topic 2, Topic 1 is included in Topic 10, etc.), with the exception of Topics 1 and 8. This is in line with a structure that proceeds by progressive specialization of one main, dominant topic (Topic 6), rather than by the identification of a certain number of independent topics of similar relevance, which would be represented by multiple topics at the same height (as seen with Topics 1 and 8).

This structure is also in line with the expected topics' proportions shown in Table 1, where we observe that Topic 6 dominates (39 expected documents), followed by specialization in Topics 5, 4, 9, and 7 (with 30, 26, 25, and 21 expected documents respectively), followed by Topic 10 (16 expected documents) and Topic 2 (13 expected documents), finally followed by Topics 3, 1, and 8 (11, 10, and 9 expected documents, respectively).

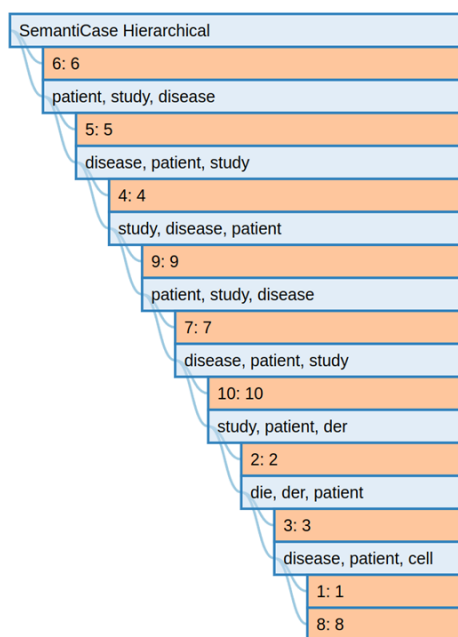


Figure 1. Topics hierarchical clustering as produced by Semanticase. Construction starts from Topics 8 and 1 (highest correlation) and reports the words shared by the two topics with highest probability (disease, patient, cell).

Table 1. Topics proportion by number of documents in the corpus.

Topic	Proportion	N. Documents
6	0.194707	39
5	0.149315	30
4	0.127531	26
9	0.123913	25
7	0.103302	21
10	0.082576	16
2	0.064592	13
3	0.054142	11
1	0.050930	10
8	0.048994	9

To discuss the semantics of the topics, in Table 2, we pair each topic index with an intuitive labeling (column *Label*). These labels are extrapolated by Prob, Frex, and a selection of 10 to 15 words per topic (details and complete list of statistics can be found in File S1 Section S5.1 and Table S5). This approach enables a discussion where the mathematical structure of the topics hierarchy describes the (power) relations among their semantic content and the associated therapies. With this intuitive labeling in place, in fact, it is easier to observe that the hierarchical structure is in line with the observed dominance of the mainstream clinical approach to RA (overlapping with category PHA, synonym of *conventional* medicine).

Table 2. Topic labels.

Topic	Words	Prob	FREX	Label
6	Leflunomice, prednisone, ankle arthroplasty, methotrexate	Leflunomide, prednisone, placebo, methotrexate, drug, combination	Prednisone, ankle arthroplasty, selective jak1, auranofin	Conventional
5	Nanoparticles, curcumin, bee venom	Curcumin, inflammation, cytokine, cancer, activation	Nanoparticle, dendrimer	Nanotech
4	Still disease, erythema, GI microbiome	Microbiota, gut, probiotic, still disease	Still and sjorgen disease, anular erythema, gut health	GI microbiome
9	Biosimilar, methotrexate	Infliximab, methotrexate, etanercept, patient, drug, trial	Biosimilar, sirukumab, methotrexate	Biologics
7	PUFA	Fatty acid, oil, fish, dietary	PUFA, fish, supplementation, marine-derived, MTX-related toxicity	Nutraceuticals
10	VNS, acupuncture	Acupuncture, nerve, stimulation, vagus, trial	Non-pharmacological, non-surgical	Electrostimulation
2	Laser, meta analyses	Laser, trial, cochrane, placebo	Laser, review	Laser therapy
3	-	-	-	German language
1	Fever inflammation	Rheumatic fever, infection, reactive	Rheumatic, heart, fasciitis, myositis	Fever
8	Angiotensin, amyloid, toll receptor, lupus	Patient, receptor, inflammation, autoimmune	Aptamer, doi, fem	Misc

The PHA-associated scientific narrative includes a relatively fixed Section 1. This section sets the problem with a conventional clinical definition and etiology of the disease (i.e., according to ACR, EULAR, and APLAR). Within this hierarchy, all other therapeutic approaches appear to be nested, with terms recalling innovation (EXP) that include in Topic 5 nanomedicine, as well as -on lower ranks- VNS and curcumin (expected in the nested Topic 7), then GI-associated therapies (Topic 4) and biologics (Topic 9) and finally nutraceuticals (Topic 7). From this point on, non-pharmacological/physical therapies are identified, including acupuncture and VNS (Topic 10) and laser therapies (Topic 2). Follows Topic 3 with German written articles where automatic translation fails, which further includes Topics 1 and 8, more difficult to label, relatively heterogeneous and less prevalent (see again how this mirrors the amount of identified articles in Expected Proportion in Table 1).

Finally, we compared these results to our previous work [14]. We can observe that several of the subcategories we defined a priori (reported in Table 3 for convenience) have indeed been automatically identified. In particular: Dys (dysbiosis, Topic 4) and Diet (Topic 7) appear as distinct topics. Further, the current automatic approach was able to discriminate between cDMARDs and bDMARDs (Topics 6 and 9, respectively), which were merged in our previous work under the AI—anti-inflammatory drug—label. Interestingly VNS, EL (electrotherapies), and AP (acupuncture), manually and distinctly labeled in our previous work, are merged here under Topic 10: the automatic approach recognizes the many commonalities existing among these therapies—(electro)stimulation by needles—despite the minimal communication existing among the medical specialties.

Additional topics (that were not discussed in [14]) are concerned with nanotechnology (Topic 5) and fever (Topic 1).

Finally, FMT (fecal transplant), US (ultrasound), MASS (massage), EM (electromagnetism), and AB (antibiotics) did not emerge. For FMT, US, MASS, and EM, this can be attributed to the small amount of literature available (they were indeed supported by 0, 5, 2, and 1 articles, respectively, in our original work).

Regarding AB, it is relevant to recall that anti-bacterial agents have been promoted in two main ways. Early approaches repurposed accidental observations on the positive effects of antibiotics on RA comorbidities. Then, more recently, in the “microbiome era”, AB were used to control the intestinal bacterial population. These represent two different discourses for the same therapy, and this has likely challenged the STM automatic recognition, which was unable to identify AB per se. Furthermore, the literature regarding the early approach is very old (earliest articles being from 1968) and could not be properly downloaded for PDF manipulation. Overall, Topic 4 describes the most recent application.

Table 3. Listing of subcategories identified manually in [14].

Category	Label	Extended MeSH
PHA	AI	Anti-Inflammatory Drug Therapy
EXP	VNS	Vagus Nerve Stimulation
EXP	Dys	Dysbiosis Therapy
EXP	AB	Anti-Bacterial Agents
EXP	Diet	Dietary Supplements
EXP	FMT	Fecal Microbiota Transplantation
USTD	US	Ultrasonic Therapy
USTD	Mass	Massage
USTD	AP	Acupuncture Therapy
USTD	EL	Electric Stimulation Therapy
USTD	LLT	Low Laser Therapy
USTD	EM	Electromagnetic Phenomena

3.2. Sentiment Analysis

Regarding sentiments (Figure 2), PHA presents with a relatively large plateau on the slightly negative sentiment side, EXP is definitely negative, and USTD presents a slightly bimodal positive behavior.

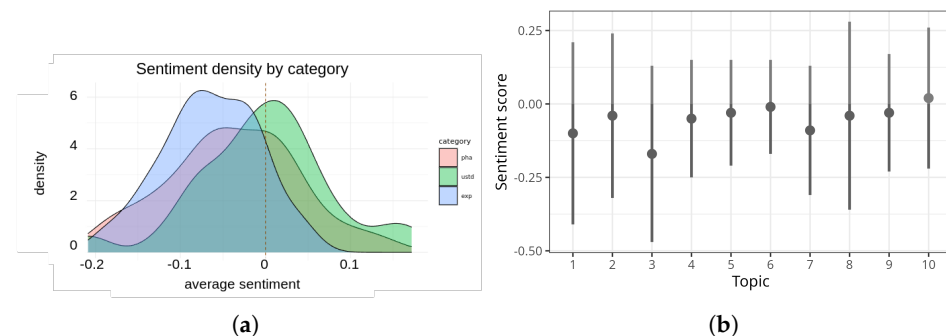


Figure 2. Sentiment estimation visualizations. (a) The sentiment distribution across the literature corpus by category values. (b) The sentiment estimation by topic. Points represent the mean value, while the bars range from the minimum to the maximum values.

When assessing how sentiments are distributed by topic (Figure 2b), upon removal of the uninformative (German language) Topic 3, we observed that the most negative average sentiment is shared by fever and nutraceuticals (Topics 1 and 7). These are followed by a milder negative sentiment in all other topics, with the exception of (nerve) electrostimulation (Topic 10). It is interesting to notice that such (average) positive sentiment is associated with a topic shared by EXP and USTD, i.e., electrostimulation, describing the same type of physical stimulus, but surrounded by a very different narrative (experimental, innovative VNS, versus traditional electroacupuncture).

3.3. Emotions Analysis

When comparing the whole range of normalized emotions (Table 4), the three categories present two shared and dominant emotions, i.e., *neutrality* and *approval*, probably unsurprisingly for a scientific document.

When looking for differences, few key observations arise. First, USTD shows a narrower range of emotions. Then, while EXP and PHA share the same variability in range, they present two complementary emotions. PHA lacks *love* and *disgust*, possibly in line with a more neutral scientific writing. EXP lacks *nervousness* and *annoyance*, possibly in agreement with excitement for novelty, which is a hallmark of these articles. In USTD, the lack of breadth is backed by a third relatively relevant emotion: *disapproval*.

Finally, the sentence-wise emotion analysis enables tracking the flow of dominant sentiments from the beginning to the conclusion of each document (see Methods Section 2.2.6). When looking at the narrative progression over the course of the whole article (Figure 3), we observe that USTD presents a pattern that is fairly unique.

Specifically, *neutrality* (Figure 3c) appears to build and grow over the course of the writing in both PHA and EXP, while USTD loses *neutrality* over the first half of the discussion. Further, the lack of *neutrality* in PHA and EXP is compensated by *approval* (Figure 3a) that sets the tone at the beginning of the articles, while the decrease in neutrality is compensated in USTD by *disapproval* (Figure 3b). In other words, PHA and EXP build their scientific writing by setting the tone with *approval* first, before moving on to *neutrality*, while USTD starts with *neutrality* that quickly loses ground for *approval/disapproval* to go finally back to *neutrality*. It has to be noted that *disapproval* dominates over *approval*. This possibly suggests a type of reporting that (needs to) work in defensive rather than proactive mode.

Table 4. Ranked normalized emotions—top ranking emotions in bold, missing emotions in italics—by category.

Emotion	PHA	USTD	EXP
neutral	9.52×10^{-1}	9.48×10^{-1}	9.53×10^{-1}
approval	2.48×10^{-2}	2.12×10^{-2}	2.27×10^{-2}
confusion	5.95×10^{-3}	6.81×10^{-3}	5.75×10^{-3}
disapproval	5.22×10^{-3}	1.28×10^{-2}	7.05×10^{-3}
curiosity	3.84×10^{-3}	3.54×10^{-3}	3.76×10^{-3}
disappointment	2.12×10^{-3}	1.84×10^{-3}	1.71×10^{-3}
sadness	1.54×10^{-3}	1.50×10^{-3}	1.75×10^{-3}
admiration	1.13×10^{-3}	9.53×10^{-4}	1.19×10^{-3}
optimism	1.10×10^{-3}	9.53×10^{-4}	5.26×10^{-4}
realization	5.65×10^{-4}	4.76×10^{-4}	8.18×10^{-4}
caring	5.18×10^{-4}	4.76×10^{-4}	7.60×10^{-4}
gratitude	2.51×10^{-4}	1.09×10^{-3}	2.53×10^{-4}
amusement	1.25×10^{-4}	1.36×10^{-4}	7.79×10^{-5}
surprise	1.25×10^{-4}	1.36×10^{-4}	4.48×10^{-4}
excitement	1.10×10^{-4}	0.00	7.79×10^{-5}
fear	9.41×10^{-5}	0.00	3.89×10^{-5}
joy	9.41×10^{-5}	1.36×10^{-4}	7.79×10^{-5}
desire	4.71×10^{-5}	0.00	1.95×10^{-5}
annoyance	3.14×10^{-5}	6.81×10^{-5}	0.00
nervousness	3.14×10^{-5}	0.00	0.00
remorse	1.57×10^{-5}	0.00	1.95×10^{-5}
disgust	0.00	6.81×10^{-5}	3.89×10^{-5}
love	0.00	0.00	3.89×10^{-5}

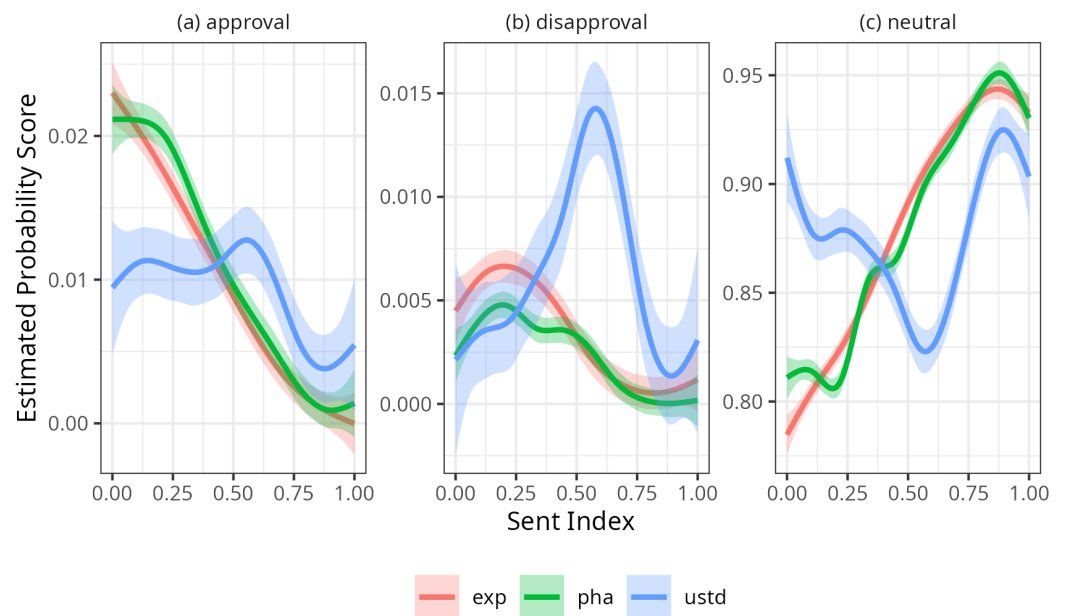


Figure 3. Dominant emotions: Approval, Disapproval, and Neutral in panels (a–c), respectively.

Although very different in substance, we can briefly discuss how these results complement the ones obtained in [14]. To ease the comparison, we report here (Figure 4) the results obtained for the systemic analysis that represents the three categories under study in terms of six so-called socio-anthropological variables. In this representation, EXP and USTD appear to be more similar than EXP and PHA. In particular, three out of six variables have a shared value, namely the *spatialization* of the disease, both in terms of the *etiology*

and of the *therapy* (i.e., where the origin and the treatment of RA are supposed to be located, in terms of body geography). This similarity does not emerge from the emotions analysis, since only for *optimism* we observe the same (stable) trend for EXP and USTD (dropping in PHA). However, this emotion contributes negligibly to the overall analysis (see all plots in Figure S1).

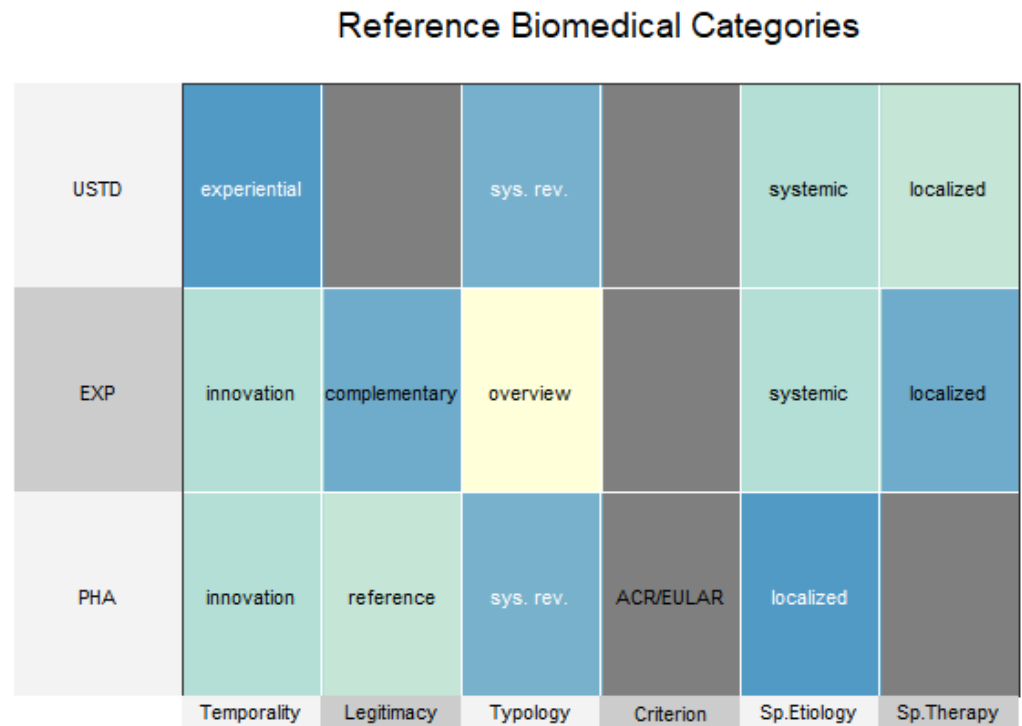


Figure 4. Reproduced with permission from [14]. Representation of the 3 categories in terms of 6 socio-anthropological variables.

Differently, PHA and EXP share the same approach in *temporality* (temporality indicates how the proposed therapy relates to time, i.e., whether its solidity descends from long-dated experience in the past or from cutting-edge discoveries shaping the future), which is *innovative*, versus *experiential* for USTD. We can speculate that this aspect is captured by the emotions analysis. We could argue that medical approaches based on cutting-edge scientific discoveries (a requirement of modern medicine) legitimates the authors to drop *neutrality* (which is another requirement of scientific writing) in favor of *approval* to assertively remark the legitimacy of the discussion that follows. Conversely, USTD, which relies on *experience* as a guarantee for legitimacy, is forced to back up this weakness with *neutrality*, at least to set the tone, i.e., at the beginning of the article. In summary, while PHA and EXP, leveraging on scientific innovation, can drop one of the commitments and hallmarks of scientific writing (i.e., *neutrality*), USTD must make up for this original sin by opening with *neutrality*.

4. Conclusions

This analysis represents an automatic approach to the medical discourse and complements the semi-automated work presented in [14].

Among the advantages of the current analysis, we observe that topic modeling reproduced, without supervision, the majority of the subcategories decided a priori in our previous work (although an absolute ground truth does not exist), provided a sufficient number of articles are available. Based on this availability, indeed, the automatic approach clearly distinguishes conventional and biologic drugs (all merged previously in the cate-

gory anti-inflammatory, AI). Intriguingly, electrostimulation was identified in one topic only (Topic 10), including both electroacupuncture and VNS, that belongs to two different categories (USTD and PHA, respectively). This indicates that the two approaches share an amount of factual commonalities that automatic approaches can identify, despite the fact that limited to no communication among the two research areas artificially segregates them. Finally, emotions analysis was able to discover a similarity between PHA and EXP and versus USTD that we could relate to the temporal legitimization of the approaches, suggesting that the medical discourse is strongly biased by these variables. Although these results can by no means directly orient medical policies, this type of work is a fundamental pre-requisite to enable a discussion in this direction, for example to support regulation of USTD therapies in medical practice.

Among the challenges, we encountered a significant burden in downloading scientific articles using automated scripts/services/API. Other efforts have already assessed the potential benefit and lost opportunities descending from enabling such sourcing [34,35]. Owing to the incomplete openness of articles, we could fully download only 204 articles (58.8% of the selected PMID). Although solving this issue is beyond the scope of this work, we wish to contribute to this heated debate and corroborate the requests of the Open Science movement: not only human but also machine accessibility represents an issue. Indeed, the article's PDF/HTML formats need accurate methods to transform it into simple, machine-processable text. At the source, the articles are in WYSIWYG Microsoft Word-like, LaTeX, or Markdown formats that are more easily and precisely processable.

Finally our study presents numerous limitations. Sentence splitting to overcome bias and enhance transparency is well adapted to the analysis of text in English (that privileges short sentences) and scientific genre (where sentences are supposedly complete units of information). This implies that the transferability of our method is strongly affected by these constraints, which, however, cover a large part of the scientific literature. Similarly, pre-processing with expert human-in-the-loop represents both a guarantee of more control and transparency on the automated process, as well as an additional potential source of bias. This subtle balance is currently one of the main challenges in LLM-driven approaches. Additionally, we were not able to fully reproduce the socio-anthropological characterization performed in [14]. This is possibly due to the insufficient number of articles and to the complexity of the concepts captured by human labeling in previous work, adding up to the limitations of AI systems recently shown for other types of hard tasks [36].

Overall, however, we believe that the integration of structural semantic topic modeling, accompanied by sentiment and emotion analysis, presents a very original approach offering novel, rarely walked paths and returning important insights, beyond traditional medical specialties and cultural stereotypes, for the ultimate well being of patients.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/j8040045/s1>. File S1: Corpus and Semanticase details; Figure S1: Emotion analysis for all 27+1 emotions; Supporting Table.xlsx with Table S1: Table containing queries and full list of 500 articles originally returned; Table S2: List of 28 articles manually processed in the socio-anthropological analysis; Table S3: List of the 204 PMID processed in the automatic analysis; Table S4: Full list of the analyzed emotions; Table S5: Semanticase Topic Words.

Author Contributions: Conceptualization, M.S. and C.N.; Methodology, M.S. and C.N.; Software, M.S.; Validation M.S. and C.N.; Formal Analysis, M.S. and C.N.; Data Curation, M.S. and C.N.; Writing—Review and Editing, M.S. and C.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Hedgecoe, A. Schizophrenia and the Narrative of Enlightened Geneticization. *Soc. Stud. Sci.* **2001**, *31*, 875–911. [[CrossRef](#)]
- Hedgecoe, A.; Martin, P. The Drugs Don't Work: Expectations and the Shaping of Pharmacogenetics. *Soc. Stud. Sci.* **2003**, *33*, 327–364. [[CrossRef](#)]
- Saha, A.; Alleyne, G. Recognizing noncommunicable diseases as a global health security threat. *Bull. World Health Organ.* **2018**, *96*, 792–793. [[CrossRef](#)]
- Maturo, M.G.; Soligo, M.; Gibson, G.; Manni, L.; Nardini, C. The greater inflammatory pathway-high clinical potential by innovative predictive, preventive, and personalized medical approach. *EPMA J.* **2020**, *11*, 1–16. [[CrossRef](#)]
- Lau, C.S.; Chia, F.; Dans, L.; Harrison, A.; Hsieh, T.Y.; Jain, R.; Jung, S.M.; Kishimoto, M.; Kumar, A.; Leong, K.P.; et al. 2018 update of the APLAR recommendations for treatment of rheumatoid arthritis. *Int. J. Rheum. Dis.* **2019**, *22*, 357–375. [[CrossRef](#)]
- Smolen, J.S.; Landewé, R.B.M.; Bijlsma, J.W.J.; Burmester, G.R.; Dougados, M.; Kerschbaumer, A.; McInnes, I.B.; Sepriano, A.; van Vollenhoven, R.F.; de Wit, M.; et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. *Ann. Rheum. Dis.* **2020**, *79*, 685–699. [[CrossRef](#)]
- Fraenkel, L.; Bathon, J.M.; England, B.R.; St Clair, E.W.; Arayssi, T.; Carandang, K.; Deane, K.D.; Genovese, M.; Huston, K.K.; Kerr, G.; et al. 2021 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis. *Arthritis Care Res.* **2021**, *73*, 924–939. [[CrossRef](#)]
- Koopman, F.A.; Chavan, S.S.; Miljko, S.; Grazio, S.; Sokolovic, S.; Schuurman, P.R.; Mehta, A.D.; Levine, Y.A.; Faltys, M.; Zitnik, R.; et al. Vagus nerve stimulation inhibits cytokine production and attenuates disease severity in rheumatoid arthritis. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 8284–8289. [[CrossRef](#)]
- Koopman, F.A.; Schuurman, P.R.; Vervoordeldonk, M.J.; Tak, P.P. Vagus nerve stimulation: A new bioelectronics approach to treat rheumatoid arthritis? *Best Pract. Res. Clin. Rheumatol.* **2014**, *28*, 625–635. [[CrossRef](#)]
- Koopman, F.A.; van Maanen, M.A.; Vervoordeldonk, M.J.; Tak, P.P. Balancing the autonomic nervous system to reduce inflammation in rheumatoid arthritis. *J. Intern. Med.* **2017**, *282*, 64–75. [[CrossRef](#)]
- Stepanov, Y.V.; Golovynska, I.; Zhang, R.; Golovynskyi, S.; Stepanova, L.I.; Gorbach, O.; Dovbynychuk, T.; Garmanchuk, L.V.; Ohulchansky, T.Y.; Qu, J. Near-infrared light reduces β -amyloid-stimulated microglial toxicity and enhances survival of neurons: Mechanisms of light therapy for Alzheimer's disease. *Alzheimer's Res. Ther.* **2022**, *14*, 84. [[CrossRef](#)] [[PubMed](#)]
- Chen, L.; Xue, J.; Zhao, Q.; Liang, X.; Zheng, L.; Fan, Z.; Souare, I.S.J.; Suo, Y.; Wei, X.; Ding, D.; et al. A Pilot Study of Near-Infrared Light Treatment for Alzheimer's Disease. *J. Alzheimer's Dis.* **2023**, *91*, 191–201. [[CrossRef](#)] [[PubMed](#)]
- Paparozi, V.; Hooshmandabbasi, R.; Ravoni, A.; Ma, Y.; Manni, L.; Koh, T.J.; Maake, C.; Guarnieri, T.; Lai, D.; Zablotskii, V.; et al. Anti-inflammatory effects of physical stimuli: The central role of networks in shaping the future of pharmacological research. *Br. J. Pharmacol.* **2025**. [[CrossRef](#)] [[PubMed](#)]
- Nardini, C.; Candelise, L.; Turrini, M.; Addimanda, O. Semi-automated socio-anthropologic analysis of the medical discourse on rheumatoid arthritis: Potential impact on public health. *PLoS ONE* **2022**, *17*, e0279632. [[CrossRef](#)]
- Vayansky, I.; Kumar, S.A. A review of topic modeling methods. *Inf. Syst.* **2020**, *94*, 101582. [[CrossRef](#)]
- Roberts, M.E.; Stewart, B.M.; Tingley, D.; Airoldi, E.M. The structural topic model and applied social science. In Proceedings of the Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation. Lake Tahoe, NE, USA, 9–10 December 2013; Volume 4, pp. 1–20.
- Roberts, M.E.; Stewart, B.M.; Tingley, D.; Lucas, C.; Leder-Luis, J.; Gadarian, S.K.; Albertson, B.; Rand, D.G. Structural Topic Models for Open-Ended Survey Responses. *Am. J. Political Sci.* **2014**, *58*, 1064–1082. [[CrossRef](#)]
- Ebadi, A.; Xi, P.; Tremblay, S.; Spencer, B.; Pall, R.; Wong, A. Understanding the temporal evolution of COVID-19 research through machine learning and natural language processing. *Scientometrics* **2020**, *126*, 725–739. [[CrossRef](#)]
- Chen, X.; Chen, J.; Cheng, G.; Gong, T. Topics and trends in artificial intelligence assisted human brain research. *PLoS ONE* **2020**, *15*, e0231192. [[CrossRef](#)]
- Kompa, B.; Hakim, J.B.; Palepu, A.; Kompa, K.G.; Smith, M.; Bain, P.; Woloszynek, S.; Painter, J.L.; Bate, A.; Beam, A. Artificial Intelligence Based on Machine Learning in Pharmacovigilance: A Scoping Review. *Drug Saf.* **2022**, *45*, 477–491. [[CrossRef](#)]
- Tunstall, L.; Von Werra, L.; Wolf, T. *Natural Language Processing with Transformers*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2022.
- Zhao, H.; Chen, H.; Yang, F.; Liu, N.; Deng, H.; Cai, H.; Wang, S.; Yin, D.; Du, M. Explainability for Large Language Models: A Survey. *ACM Trans. Intell. Syst. Technol.* **2024**, *15*, 1–38. [[CrossRef](#)]
- Coletti, M.H.; Bleich, H.L. Medical Subject Headings Used to Search the Biomedical Literature. *J. Am. Med. Inform. Assoc.* **2001**, *8*, 317–323. [[CrossRef](#)] [[PubMed](#)]

24. Stamm, A.; Reimers, K.; Strauß, S.; Vogt, P.; Scheper, T.; Pepelanova, I. In vitro wound healing assays—state of the art. *BioNanoMaterials* **2016**, *17*, 79–87. [[CrossRef](#)]
25. Sadvilkar, N.; Neumann, M. PySBD: Pragmatic sentence boundary disambiguation. *arXiv* **2020**, arXiv:2010.09657. [[CrossRef](#)]
26. Chai, C.P. Comparison of text preprocessing methods. *Nat. Lang. Eng.* **2023**, *29*, 509–553. [[CrossRef](#)]
27. Bondi, M. Perspectives on keywords and keyness. In *Keyness in Texts*; John Benjamins: Amsterdam, The Netherlands, 2010; pp. 1–20.
28. Gabrielatos, C. Keyness analysis: Nature, metrics and techniques. In *Corpus Approaches to Discourse*; Routledge: Oxfordshire, UK, 2018; pp. 225–258.
29. Rinker, T. *Sentimentr: Calculate Text Polarity Sentiment*. Available online: <http://github.com/trinker/sentimentr> (accessed on 27 October 2025).
30. Jain, S.M. Hugging face. In *Introduction to Transformers for NLP: With the Hugging Face Library and Models to Solve Problems*; Apress: Berkeley, CA, USA, 2022; pp. 51–67.
31. Lowe, S. A Model Trained from Roberta-Base on the Go_emotions Dataset for Multi-Label Classification. 2023. Available online: https://huggingface.co/SamLowe/roberta-base-go_emotions (accessed on 27 October 2025)
32. Demszky, D.; Movshovitz-Attias, D.; Ko, J.; Cowen, A.; Nemade, G.; Ravi, S. GoEmotions: A Dataset of Fine-Grained Emotions. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020.
33. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016.
34. Himmelstein, D.S.; Romero, A.R.; Levernier, J.G.; Munro, T.A.; McLaughlin, S.R.; Greshake Tzovaras, B.; Greene, C.S. Research: Sci-Hub provides access to nearly all scholarly literature. *eLife* **2018**, *7*, e32822. [[CrossRef](#)]
35. Buehling, K.; Geissler, M.; Strecker, D. Free access to scientific literature and its influence on the publishing activity in developing countries: The effect of Sci-Hub in the field of mathematics. *J. Assoc. Inf. Sci. Technol.* **2022**, *73*, 1336–1355. [[CrossRef](#)]
36. Shojaee, P.; Mirzadeh, I.; Alizadeh, K.; Horton, M.; Bengio, S.; Farajtabar, M. The Illusion of Thinking: Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity. *arXiv* **2025**, arXiv:2506.06941. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.