BILINGUAL PUBLISHING GROUP

Forum for Linguistic Studies

https://journals.bilpubgroup.com/index.php/fls

ARTICLE

Semantic Sense in Medical Communication: A Collocational Analysis of 'Cancer' and 'Patient'

Khalid Rokan Mansoor [1] **,** ***Afsaneh Shokri*** [2] **,** ***Khorshid Mousavi*** [3*]

[1] *English Department, Al Turath University, Baghdad 10013, Iraq*

[2] *Academic Department of Foundation Studies (FS), Global College of Engineering and Technology, Muscat P.O. Box 2546, Oman*

[3] *Department of English Language, The Islamic University, Najaf 54003, Iraq*

ABSTRACT

In this study, the semantic distribution of the clinical collocations "cancer" and "patient," with an emphasis on their medical applications, was analyzed using Hoey's lexical priming model. A corpus of 1000 medical research articles from leading journals, encompassing approximately 2,576,035 words, was utilized. AntConc software identified high-frequency words, and the semantic loads of collocations of "cancer" and "patient" were categorized using Xiao and McEner's labels. A qualitative analysis further examined patterns of semantic load, clinical collocations, and contextual factors. Cross-validation with a second rater achieved high inter-rater reliability (Kappa = 0.983, $p = 0.000$). The findings revealed that "patients" typically carry a neutral semantic prosody, reflecting their association with various medical contexts, interventions, and cases without inherently positive or negative connotations. In contrast, "cancer" predominantly carries a negative semantic prosody, strongly linked to serious health risks, adverse outcomes, and mortality. These results underscore the importance of context in medical terminology for effective communication. This study supports Hoey's theory that repeated exposure to specific contexts primes lexical items for clinical collocation. It emphasizes how context influences the emotional tone of medical communication. The findings have important clinical implications, particularly for enhancing doctor-patient communication, promoting empathetic interactions, and reducing misunderstandings in medical settings.

Keywords: Cancer; Clinical Collocations; Chemical Health Risks; Medical Science; Patient; Semantic Sense

*CORRESPONDING AUTHOR:

Khorshid Mousavi, Department of English Language, The Islamic University, Najaf 54003, Iraq; Email: kh.mousavi2024@gmail.com

ARTICLE INFO

Received: 23 April 2025 | Revised: 26 May 2025 | Accepted: 10 June 2025 | Published Online: 18 December 2025
DOI: https://doi.org/10.30564/fls.v7i12.9662

CITATION

Mansoor, K.R., Shokri, A., Mousavi, K., 2025. Semantic Sense in Medical Communication: A Collocational Analysis of 'Cancer' and 'Patient'. Forum for Linguistic Studies. 7(12): 1903–1914. DOI: https://doi.org/10.30564/fls.v7i12.9662

1. Introduction

Cancer is one of the most significant health challenges facing the world today, known for its high prevalence and mortality rates[1]. The World Health Organization (WHO) identifies cancer as a leading cause of death globally, with millions of new cases diagnosed each year. In 2012 alone, there were approximately 14 million new cases and 8.2 million cancer-related deaths reported[2]. The number of new cases is expected to rise by 70% over the next two decades, making it more important than ever to develop effective tools for prevention, detection and treatment of this disease[3]. New research into cancer draws on existing knowledge reported in scientific literature. Relevant literature has grown rapidly in both size and complexity. There are over 3 million citations related to 'cancer' in PubMed (www.ncbi.nlm.nih.gov/pubmed). As many as 151,872 were added over the past year. The high number of potentially relevant articles is a valuable source on to which new research can build but at the same time poses a challenge for scientists. While applications such as Google Scholar (scholar.google.com) and PubMed Advanced Search (www.ncbi.nlm.nih.gov/pubmed/advanced) can be of help, they fall short in providing all and only the information of interest. Accurate representation of cancer-related information in medical literature is essential for educating healthcare professionals, supporting patients, and enhancing clinical practices.

One method for the true representation of cancer-related information in medical literature is the concept of *semantic prosody* a concept rooted in corpus linguistics. First introduced by Louw[4], semantic prosody refers to the connotative meaning that emerges from the frequent co-occurrence of words, offering insight into how language shapes perception and interpretation[5]. Though extensively studied in general linguistics, its application in medical discourse remains underexplored. It offers a unique lens through which to analyze language use in medical communication. Before moving to the nature of semantic prosody, it is necessary to add that semantics and syntax form the fundamental structures of all human languages and are universally present[6] and within the umbrella term of semantic, semantic prosody play the important role[7]. The concept of semantic prosody was originally outlined by Louw[4]. It describes the characteristics of a word in terms of some aspects of its semantic context. The context has implications for the meaning of a word since the prosody becomes part of the word meaning[8]. Semantic prosody, which is based on the study of co-occurrences or collocations, is an important topic in corpus-based studies in different disciplines including medicine[9]. Though the term "semantic prosody" is somewhat controversial, it has gained wide acceptance and has been investigated in many disciplines[10,11]. Importantly, Milojkovic and Louw[11] argue that semantic prosody is science, not just a concept.

There has been a heated discussion on semantic prosody in the last few decades[12]. Put simply, semantic prosody is connotative meaning unreported by traditional semanticists as it is inferred from habitual lexical co-occurrences in a contemporary corpus of natural language data. Semantic prosody in medical communication is particularly relevant in clinical and academic contexts, as it shapes how healthcare professionals and researchers communicate. Words like *cancer* and *patient* carry significant emotional and cognitive weight in medical settings. Analyzing their semantic prosody can provide insights into how language influences attitudes, decision-making, and even patient care. Furthermore, understanding semantic prosody can enhance the clarity and precision of medical writing, benefiting both healthcare practitioners and patients. Three are a few studies in the medical science and text analysis of medical corpus touching semantic prosody[13–15]. However, it can be referred to the studies about the analysis of medical texts, which have several practical applications, leveraging various techniques and technologies. One significant application is the use of natural language processing (NLP) to extract clinical concepts from unstructured data, such as electronic health records (EHRs) that can enhance the richness of patient information available for research and treatment planning[16]. Another application involves the classification of medical texts using machine learning and heuristic approaches, which can improve the categorization of medical information into topics or disease types, enhancing the interpretability and performance of classification models[17]. Additionally, automated processing and analysis of medical passages using programming languages like Java can improve the efficiency of handling large volumes of medical data, aiding in diagnosis and treatment planning[18]. Privacy-preserving techniques for keyword-based classification of unstructured medical text data also play a crucial role in maintaining data security while enabling

collaborative analytics[19]. These applications demonstrate the potential of medical text analysis to improve healthcare outcomes and research capabilities.

Regardless of the fact that semantic prosody has been widely studied in general linguistics, its application in the field of medical science remains underexplored. Investigating the semantic prosody of high-frequency medical terms such as *patient* and *cancer* can offer valuable insights into their usage and implications in scientific literature. This study seeks to bridge the gap between corpus linguistics and medical communication by examining the semantic prosody of *patient* and *cancer* in medical research articles. The findings of this study contribute to a deeper understanding of language use in medical contexts, helping researchers and clinicians communicate more precisely and effectively. Specifically, it focuses on how semantic prosody is reflected in the collocations of key terms commonly found in medical literature. The study examines how researchers convey implicit meanings through the use of the terms of "patient" and "cancer," highlighting the nuances and connotations inherent in medical communication.

By analyzing these terms, the research aims to uncover the ways language shapes perceptions and communication within professional medical settings. This study seems to be thought-provoking as it is the first to specifically investigate the relationship between "cancer" and "patient" in medical texts. Moreover, it uniquely explores their connection to Chemical Health Risks, providing essential insights into how these terms address patient needs.

This research classifies patient concerns, transforms unstructured medical texts into actionable data, and enhances patient-centered communication. Overall, it lays the groundwork for improved healthcare delivery and risk management in both medical and environmental health contexts.

This research seeks to address important questions in order to improve the clarity and emotional impact of medical communication, ultimately enhancing patient care and professional education. This study seeks to fill this gap by focusing on the semantic prosody of the words "cancer" and "patient" in scientific articles, particularly in the field of chemical health hazards. The aim of the study is to show how language in medical texts implicitly influences the audience's perception, transforming fragmented texts into coherent insights, and thereby helping to enhance patient-centered care, improve clinical communication, and develop more effective risk management strategies.

The main objective is to apply these findings to elevate the quality of medical communication, enabling healthcare professionals to use language more effectively. This will help them build trust, provide better emotional support, and offer clear, empathetic explanations during discussions about diagnosis and treatment. Additionally, this research facilitates the conversion of unstructured medical texts into structured data, allowing for large-scale analysis and better clinical decision-making.

2. Research Method

This section outlines the methodological framework adopted for the present study. It is organized into three main components: the Corpus, which describes the selection and characteristics of the data; the Procedure, which details the steps taken to collect, process, and prepare the data for analysis; and the Data Analysis, which explains the analytical tools and techniques employed to investigate the research questions. This structured approach ensures transparency and replicability while aligning with established practices in corpus-based discourse analysis.

3. Corpus

The data for this study comprised 1000 research articles (RAs) from the field of medical sciences. These articles were sourced from five leading academic databases: Elsevier, Sage, Springer, Taylor & Francis, and Wiley Online Library. The selection of these databases was based on consultation with field experts and consideration of journal Impact Factors (MIFs) ranging from 1 to 5 to ensure the findings' generalizability and representativeness of the medical research discourse community. Articles were published between 2010 and 2024 to provide a contemporary and comprehensive view of the linguistic trends in medical communication. The articles were selected through random sampling, ensuring a balanced representation of native and non-native authors. The final corpus consisted of approximately 2,576,035 words. To ensure that the chosen articles aligned with the study's objectives, we applied specific inclusion criteria. These criteria required that the texts be published in reputable journals with a strong peer-review process. Additionally, the arti-

cles needed to include distinct sections, such as Introduction, Methods, Results, and Discussion, to facilitate a systematic analysis. Also, the corpus needed to belong to subfields relevant to clinical and patient-centered medical research to maintain contextual consistency.

4. Procedure

Corpus Corporation handled the study's layout and design. Based on Hoey's lexical priming framework, a comparison of the use of clinical collocations of two words, "cancer" and "patient," as high-frequency words in the corpus was conducted, and the results were qualitatively and descriptively assessed. Additionally, the design was exploratory because, as Dornyei [20] notes, it is employed to tackle a research challenge when there are few or no prior studies to draw on. The emphasis is on acquiring new information and becoming more familiar with a problem so that it can be investigated later or when it is still in the early stages. Exploratory research is adaptable and can address various kinds of research inquiries (what, why, how).

To conduct the current research, the corpus was converted into plain text files (txt format) for analysis, excluding tables, references, and unrelated sections to focus solely on textual content relevant to the research questions. The study's framework was designed using a corpus-based methodology handled through Corpus Corporation. The semantic dispersal analysis or the medical intervention in text analysis centered on Hoey's *Lexical Priming* framework to explore the clinical collocational patterns and semantic prosody of two high-frequency terms in medical communication: *cancer* and *patient*. This design was exploratory and aimed to gather insights into an under-researched area to inform future investigations. To identify usage patterns, we determined the frequency of clinical collocations using AntConc software [21]. AntConc was selected for its versatility and reliability in corpus linguistics analysis, particularly due to its ability to efficiently handle large datasets. The software supports various analytical functions, including frequency analysis, clinical collocation identification, concordance generation, and keyword extraction, all of which are essential for identifying linguistic patterns in a specialized field like medical communication. Additionally, its user-friendly interface and free access make it a preferred tool in linguistic research. AntConc is especially suitable for analyzing medical texts due to its robust capabilities for identifying patterns in technical language and specialized vocabulary. This allows researchers to uncover nuanced semantic relationships and collocational behaviors in domain-specific contexts. besides, AntConc's ability to handle specialized corpora allows for a detailed investigation of technical terms like *cancer* and *patient* within the unique context of medical communication.

The articles were pre-processed to isolate content from the primary sections, ensuring consistent analysis. High-frequency words were identified through preliminary frequency counts, with *cancer* and *patient* emerging as the most prominent terms in the corpus. To analyze semantic prosody, affective context labels were applied to categorize the prosody of clinical collocations into three categories of positive semantic prosody (contexts reflecting favorable or supportive connotations), neutral (contexts with no evident positive or negative affect), and negative (contexts reflecting unfavorable or distressing connotations) [5]. The affective meaning of each clinical collocation was determined by examining how its referents and complements functioned semantically within the surrounding context. Special attention was given to noun groups and their clinical collocational patterns, as guided by Hoey's model, emphasizing the mental associations individuals develop through repeated lexical encounters.

5. Data Analysis

To ensure the validity and reliability of the data analysis during secondary data analysis, 20% of the clinical collocations were independently re-analyzed by a second and a third researcher (a Ph.D. in TEFL and a medical science expert). These researchers were briefed on the study's goals and methodologies. Using Cohen's Kappa, the inter-rater reliability was assessed, yielding a high agreement (Kappa = 0.983, $p = 0.000$), confirming the robustness of the findings.

The processed corpus (plain text files) was loaded into AntConc. This corpus excluded non-textual elements such as tables, figures, and references to ensure the focus remained on linguistic content. The *Word List* tool in AntConc was used to generate a frequency list of all words in the corpus. This feature ranks words based on their frequency of occurrence, allowing the identification of high-frequency terms

relevant to the research. *Cancer* and *patient* were identified as the most frequent terms in the corpus, reflecting their centrality in medical research discourse. The high-frequency words were then examined in their clinical collocational contexts to determine their semantic prosody. The *Keyword List* tool was employed to identify keywords—terms that are statistically more frequent in the medical corpus compared to a general reference corpus. This comparison highlights domain-specific terms, ensuring that the analysis focuses on vocabulary relevant to medical communication and diagnosis.

The *Collocates* tool was used to identify the co-occurring words (clinical collocations) of *cancer* and *patient*. This step involved setting specific parameters, such as the span (number of words before and after the target word), to extract meaningful linguistic relationships. A span of ±5 words was used, as recommended in prior corpus studies, to capture relevant clinical collocational patterns. Concordance lines for *cancer* and *patient* were generated to analyze their contextual usage. This step involved examining surrounding words and phrases to understand the semantic and pragmatic implications of the terms in medical discourse. The *Clusters/N-Grams* tool was utilized to study the semantic dispersion of *cancer* and *patient* across different sections of the articles (e.g., Introduction, Results, Discussion). This helped identify whether these terms were predominantly associated with positive, neutral, or negative contexts based on their co-texts.

6. Results

Semantic prosody refers to the consistent affective coloring or emotional tone that certain words or phrases carry within a particular context. In this section, some sentences containing the collocations of 'patient' were selected, and they were analyzed in terms of their semantic prosody. Furthermore, the researcher looked at the semantic preferences of each collocation. Based on Stubbs' definition, semantic preference focuses on the relation between a lemma or word-form and a set of semantically related words. The corpus analysis revealed that the word "patient(s)" appeared 1273 times. Furthermore, the word 'cancer' appeared 989 times. Among these, specific collocations like "cancer patients" (occurred 57 times), "patients with cancer" (34 times), "polytrauma patients" (9 times), and "patients with PE" (11 times) can be observed. The following **Table 1** shows the collocations and frequencies for a sample of data.

Table 1. Sample of Key Collocations and Frequencies.

Collocation	Frequency	Semantic Prosody
Cancer patients	117	Negative
Patients with cancer	98	Negative
Polytrauma patients	26	Negative
Patients with PE	25	Neutral
Not all patients	21	Neutral
One patient	47	Neutral
10–25% of patients	1	Neutral
33–68% of patients	1	Neutral
Lung cancer patients	36	Neutral
Head and neck cancer	14	Negative
Esophageal cancer	5	Negative
Cancer of ovary/uterus/prostate	8	Negative
Renal cancer	16	Negative
Fatal cancer	9	Negative
Colorectal cancer	23	Negative
Non-small cell lung cancer	2	Negative/Neutral nuance

Note: This table supports the interpretive emphasis on certain semantic patterns discussed below.

1. "However, curative NETLM resection is only possible in about 10–25% of patients due to tumor size, location, and other patient/tumor characteristics"[22].

"10–25% of patients": The collocation "of patients" indicates a quantitative aspect, emphasizing the population of individuals undergoing curative NETLM resection. The semantic prosody here is neutral, focusing on numerical representation. The lemma "patients" is associated with terms

related to medical treatment and conditions, such as "curative," "resection," "tumor," and "characteristics." The semantic preference here is related to medical interventions, procedures, and characteristics of medical conditions.

2. "Trauma to the mesentery is usually associated with other injuries that may change the medical care of polytrauma patients"[23].

"Polytrauma patients": The collocation "polytrauma patients" suggests a specific subgroup of patients who have experienced multiple traumatic injuries. The semantic prosody leans towards a negative connotation, highlighting the severity and complexity of their medical condition. The lemma "patients" is associated with terms related to traumatic injuries, such as "polytrauma," "mesentery," and "injuries." The semantic preference is related to traumatic medical conditions and injuries.

3. "However, EVAR may not always be the optimal treatment option since not all patients are eligible for it. A hostile neck, consisting of severe angulation, a short, reverse taper and severe calcification and thrombus, remains a leading anatomical limitation of EVAR"[24].

"Not all patients": The collocation "not all patients" implies variability among individuals in terms of eligibility for a specific treatment option. The semantic prosody is neutral, indicating diversity in patient characteristics and medical needs. The lemma "patients" is associated with terms related to treatment eligibility and options, such as "optimal treatment," "eligible," and "hostile neck." The semantic preference is connected with the variability in treatment options and eligibility criteria.

4. "It has an influential role before therapy planning and before targeted therapy for treatment of lung cancer to obtain tissue for molecular testing in lung cancer patients"[25].

"Lung cancer patients": The collocation "lung cancer patients" specifies the group of individuals affected by lung cancer. The semantic prosody is neutral, focusing on the medical condition under consideration in therapy planning and molecular testing. The lemma "patients" is associated with terms related to cancer diagnosis and treatment, such as "therapy planning," "targeted therapy," and "molecular testing." The semantic preference is related to cancer diagnosis, treatment, and testing.

5. "One patient who underwent only angiography of the external carotid artery without super-selective angiography of the branches was excluded"[26].

"One patient": The collocation "One patient" refers to a singular individual within a medical context. The semantic prosody is neutral, indicating a specific case or instance relevant to the discussion. The lemma "patient" is associated with the singular term "one," indicating a specific instance or case within a medical context. The semantic preference is related to individual cases or instances.

6. "Complications from fiducial marker insertion have been reported, including pneumothorax in 33–68% of patients as well as bleeding"[27].

"33–68% of patients": The collocation "of patients" again indicates a subset of individuals affected by a medical procedure or intervention. The semantic prosody is neutral, specifying the proportion experiencing complications from fiducial marker insertion. The lemma "patients" is associated with terms related to medical procedures and complications, such as "fiducial marker insertion," "pneumothorax," and "bleeding." The semantic preference is related to procedural outcomes and complications.

7. "It is unclear whether mortality in patients with PE treated with full anticoagulation is a result of progressive right ventricle failure or of subsequent embolism"[28].

"Patients with PE": The collocation "patients with PE" specifies individuals diagnosed with pulmonary embolism (PE). The semantic prosody is neutral, indicating the medical condition under consideration regarding mortality and treatment. The lemma "patients" is associated with terms related to medical conditions and treatments, such as "PE," "anticoagulation," and "mortality." The semantic preference is connected with pulmonary embolism (PE) diagnosis, treatment, and outcomes.

The other high frequent word in the corpus was cancer. The term along with the collocations were analyzed in some sentences. Following is an elaboration on these terms and its related semantic load.

1. "Nasopharyngeal carcinoma (NPC) is one of the most common *head and neck cancer* in Southeast Asia, especially in southern China."

The term "cancer" here is associated with a specific type of cancer prevalent in a certain region, emphasizing its commonality and potential impact on public health. This collocation directly relates to a specific type of cancer. The context is clinical and indicates prevalence, contributing to a negative semantic prosody due to its association with illness and the need for medical attention. In other words, this term is often used in scientific and medical texts as an indicator of negative connotation in connection with serious and dangerous diseases such as cancer. Besides, by highlighting the prevalence of NPC in Southeast Asia, the sentence underscores its significance as a public health issue. The inclusion of the specific region intensifies the negative semantic load by drawing attention to the localized impact of the disease, implicitly addressing the need for targeted healthcare resources and interventions.

2. "In our study, we retrospectively compared patients who underwent RE with prior coiling with patients without prior coiling with regard to the appearance of early and late toxicities according to the *National Cancer Institute*'s Common Terminology Criteria for Adverse Events"

The term "cancer" is part of an institutional name, which does not directly convey a negative sense but is associated with the context of adverse events, indirectly hinting at negative outcomes related to cancer therapy. The collocation is neutral in its institutional context. However, since it is related to cancer research and adverse events, it indirectly carries a negative connotation due to the association with cancer therapy and side effects. Although "National Cancer Institute" is a neutral reference to an organization, the broader context of "early and late toxicities" linked to treatment outcomes introduces a negative nuance. This framing emphasizes the hardships faced by patients undergoing cancer therapies, reflecting the complexity of managing side effects.

3. "*Esophageal cancer* is the sixth leading cause of cancer-related death worldwide."

The semantic sense is starkly negative, as the statement highlights esophageal cancer's high mortality rate. By explicitly referencing death statistics, the sentence underscores the severe consequences of this disease. The global scale of the problem amplifies the negative connotation, portraying esophageal cancer as a critical health challenge requiring significant attention and resources. In other words, the terms highlight the severity and mortality associated with esophageal cancer, clearly conveying a negative sense or impact.

4. "In this last case, back pain might be due to colon disease, pelvic inflammation, or *cancer of the ovary*, uterus, or prostate."

The sentence lists cancer as a possible serious cause of back pain, reinforcing its negative connotation. This collocation lists various cancers affecting reproductive organs, carrying a negative load due to the association with serious and potentially life-threatening diseases. To say differently, the inclusion of cancer among possible causes of back pain suggests its role as a hidden yet severe threat. This portrayal reinforces the fear and anxiety surrounding cancer diagnoses, as well as the need for thorough medical evaluation to rule out life-threatening conditions.

5. "Renal cell carcinoma (RCC) is the third most common genitourinary tumor, representing 2% of all *adult cancers*."

The context indicates prevalence and categorization of cancer, implying its widespread impact and seriousness. Adult "cancers" is a broad term that categorizes cancers affecting adults. The context implies the prevalence and impact on the adult population, thus carrying a negative sense. The statistic "2% of all adult cancers" contextualizes RCC's prevalence, drawing attention to its impact despite its relatively low incidence compared to other cancers. This highlights the broader health implications of RCC and the importance of awareness and early detection. The semantic prosody is negative due to the mention of renal cell carcinoma as a significant health concern within the adult population.

6. "The aim of this study is to measure the radiation dose in patient and staff following chemoembolization of HCC and to estimate the risk of *fatal cancer* and deterministic effect."

The term "fatal cancer" explicitly emphasizes mortality and severe risk, reinforcing a strong negative load. The term "fatal cancer" has a strong negative sense, explicitly emphasizing the deadly nature of certain cancers, which underscores the seriousness and mortality associated with the disease.

7. "Aggressive treatment of lung tumors by resection or

ablation has been established in various entities, such as *colorectal cancer*, melanoma, *renal cancer*, sarcoma, *non-small cell lung cancer (NSCLC)* and others."

The semantic prosody is predominantly negative, as it underscores the severity of cancers requiring aggressive treatment. However, the mention of treatment options introduces a neutral or slightly positive nuance, suggesting efforts to combat these conditions. Listing multiple types of cancers, including "colorectal cancer," "melanoma," and "renal cancer," emphasizes the diverse and serious nature of these diseases. The use of "aggressive treatment" reflects the challenges of addressing these conditions, reinforcing the negative semantic prosody. However, the mention of "resection or ablation" suggests a sense of hope and progress in treatment strategies, balancing the overall tone. Overall, the word "cancer" in the provided sentences predominantly carries a negative semantic load, reflecting its association with disease, risk, and mortality.

7. Discussion

As stated, the essential goal of the current research was to analyze two high frequent collocations of medical terms that were 'caner' and 'patient' based on Hoey's lexical priming model. The analysis reveals that the word "patients" usually has a neutral semantic prosody, primarily used in medical contexts related to interventions and individual cases. In contrast, the term "cancer" predominantly carries a negative connotation, reflecting its association with serious health risks, adverse outcomes, and mortality. This distinction highlights the significance of context in shaping the emotional tone and implications of medical terminology.

In medical settings, "patients" typically maintains a neutral tone because it is often linked to a variety of medical conditions, treatments, and procedural details, without carrying any inherently positive or negative connotations. While "patients" conveying descriptive and quantitative information such as "10–25% of patients" and "33–68% of patients" remains neutral, the term "cancer" embodies a significantly negative load due to its strong associations with severe health challenges and mortality.

Phrases like "polytrauma patients" and "patients with PE" (pulmonary embolism) specify distinct groups based on their medical conditions; however, the term "patients" itself continues to serve as a neutral description of individuals receiving care. The phrase "one patient" refers to a specific instance, sustaining a neutral perspective within the broader medical context. Furthermore, terms like "lung cancer patients" link "patients" to specific medical situations, providing clear descriptive information about the nature of their conditions or treatments while retaining a neutral tone.

Phrases such as "esophageal cancer is the sixth leading cause of cancer-related death worldwide" and "risk of fatal cancer" emphasize the life-threatening nature of cancer, which contributes to its negative perception. The "National Cancer Institute" links cancer to the adverse consequences of treatments, further reinforcing its negative associations. Additionally, the statement "back pain might be due to cancer of the ovary, uterus, or prostate" identifies cancer as a serious potential cause of symptoms, which again highlights its negative implications.

The findings of this study align with the research conducted by Feng et al.[29]. Similar to the current study, Feng et al.[29] introduced a connotation lexicon—a new type of lexicon that categorizes words based on their connotative polarity. This includes words with positive connotations (e.g., award, promotion) and those with negative connotations (e.g., cancer, war). The results suggest that understanding the connotation of words often requires common sense and world knowledge.

Additionally, terms like "cancer" are commonly associated with negative connotations, which contrasts sharply with words that evoke positive sentiments. The current study's findings are supported by Perez et al.[30], who stated that this negative connotation can influence patient perceptions and experiences, potentially worsening feelings of helplessness and distrust in medical settings. Furthermore, the language used in cancer-related information can position patients in a passive role, contributing to negative perceptions and potentially impacting health outcomes[31].

These findings highlight the importance of using careful language in medical communications to mitigate negative connotations and improve patient engagement and understanding.

The study supports Hoey's theory of lexical priming, which suggests that terms gain meaning through their contexts and co-texts. Our findings highlight the significance of collocates and their positions within corpora. Understanding

both general and specific collocations is vital for academic reading and writing, impacting job opportunities and academic success [32]. A lack of collocation awareness can hinder academic text construction. The results can assist researchers in recognizing variations in collocation use across disciplines and improve the natural flow of language in scholarly writing using authentic concordance lines.

Surprisingly, Bi's [33] analysis yielded similar results. By comparing Chinese students' writing with that of native English speakers, he explored the use of frequency, collocation, and semantic prosody of two sets of synonyms. While the collocational patterns varied significantly between the two groups, the semantic prosody remained the same. The findings are consistent with previous research on semantic prosody by Louw [4], Hunston [10], Sinclair [34], and Hoey [35]. These analyses have indicated that the affective meaning of a term can only be fully understood when it is used in the context of its routine collocations.

The current study suggests that the terminology used to describe individuals with cancer can have significant psychological impacts. For instance, terms like "cancer patients" versus "people with cancer" can influence how individuals perceive themselves and their illness. The term "cancer patients" may lead to a perception of being defined by the disease, which can increase feelings of helplessness and stigma. This, in turn, may exacerbate psychological distress, including anxiety and depression, which are common among those affected by cancer. In contrast, using "people with cancer" emphasizes the individual first, helping to maintain a sense of identity beyond the illness. This person-centered language may lead to better psychological outcomes by reducing stigma and promoting a more holistic view of the individual.

Overall, the psychological impact of cancer is profound, affecting emotional well-being, social interactions, and overall quality of life. Proactive psychological interventions and supportive language can help mitigate these effects, promoting better mental health and coping strategies during the cancer journey.

The implications of this study also extend to clinical communication. Understanding the semantic nuances of terms like "cancer" and "patients" can guide healthcare providers in their choice of vocabulary when communicating with patients. For example, replacing negatively framed phrases with more neutral or positive expressions could help reduce patient anxiety and create a more supportive therapeutic environment. Improved communication may lead to greater patient trust and adherence to medical advice, potentially enhancing treatment outcomes. Instead of emphasizing "risks" or "fatal outcomes," physicians could focus on preventive measures or treatment options, thereby fostering a sense of agency and hope in their patients.

In educational settings, these findings have practical applications for teaching English to medical students. By integrating semantic analysis into medical English courses, educators can train students to recognize the emotional undertones of medical vocabulary and use context-appropriate language when addressing patients. Role-play activities and case studies focused on effective communication strategies can help future healthcare professionals develop both linguistic precision and empathy. Additionally, corpus-based tools can be utilized in classrooms to highlight the frequent collocates of medical terms, enabling medical students to practice using them accurately and sensitively.

In summary, this research underscores the importance of raising awareness among healthcare professionals about the implications of language use in clinical settings. Incorporating semantic analysis into communication training programs for doctors and nurses can enhance patient-centered care. Addressing the emotional weight of medical terminology may reduce misunderstandings and foster more constructive interactions, ultimately improving overall healthcare quality.

8. Conclusions

The primary aim of this research was to analyze the clinical terms "cancer" and "patient" using Hoey's lexical priming model. The analysis showed that the term "patients" generally carries a neutral semantic tone, focusing on medical contexts, interventions, and individual cases. This neutrality stems from its frequent associations with a wide range of medical conditions, treatments, and procedural details that do not inherently possess positive or negative meanings. In contrast, the term "cancer" usually conveys a negative connotation due to its strong connections with serious health risks, adverse outcomes, and mortality. This distinction highlights the importance of context in shaping the emotional

tone and implications of medical terminology, which is vital for effective communication in healthcare settings.

To enhance communication between healthcare providers and patients, medical training programs should incorporate lessons on semantic prosody into clinical communication courses. For example, clinicians should be encouraged to use language that highlights treatment and recovery, such as "manageable condition" or "recovery plan", rather than terms like "cancer" or "fatal", which may induce fear and anxiety. Additionally, role-playing scenarios and simulations could be included in training to help practitioners practice delivering sensitive information in ways that are both accurate and empathetic.

In teaching medical English, educators should emphasize the emotional and contextual nuances of terminology. This approach could include exercises that help medical students recognize the potential emotional impact of their word choices and adjust their language to create a more supportive and reassuring environment for patients. For example, students could analyze medical texts to identify patterns in semantic prosody and practice using alternative phrasing that promotes patient comfort and trust.

Future research could explore the effects of specific lexical choices on patient satisfaction, trust, and treatment adherence. Clinical studies might examine how patients emotionally and behaviorally respond to different terminologies during diagnosis and treatment discussions. For instance, does framing a diagnosis with neutral or positive language lead to better patient outcomes?

Additionally, cross-cultural studies could investigate how the semantic prosody of medical terms varies across languages and cultural contexts. Such research could provide valuable insights into developing culturally sensitive communication strategies in multilingual healthcare settings. Longitudinal studies may also look at how the use and understanding of medical terminology evolve among healthcare professionals throughout their careers, identifying critical stages for targeted training interventions. Future studies may consider examining how collocational patterns vary by country of origin or healthcare system structure, as differences in access, terminology (e.g., the role of "healthcare providers"), and public health discourse could influence linguistic choices in scientific writing. Finally, it is suggested that researchers investigate how patients perceive and respond to different medical terms healthcare providers use. Understanding the patient perspective could lead to more patient-centered communication strategies and enhance patient satisfaction and outcomes.

Author Contributions

Conceptualization, K.R.M., K.M., and A.S.; methodology, K.R.M., K.M., and A.S.; software, K.M., and A.S.; validation, K.R.M., K.M., and A.S.; formal analysis, K.R.M., K.M., and A.S.; investigation, K.R.M., K.M., and A.S.; resources, K.R.M., K.M., and A.S.; data curation, K.R.M., K.M., and A.S.; writing—original draft preparation, K.R.M., K.M., and A.S.; writing—review and editing, K.M., and A.S.; visualization, K.R.M. and A.S.; supervision, K.R.M., K.M., and A.S.; project administration, K.R.M., K.M., and A.S. All authors have read and agreed to the published version of the manuscript.

Funding

This work received no external funding.

Institutional Review Board Statement

Not applicable. The study did not require ethical approval.

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

Data Availability Statement

Data available on request due to privacy/ethical restrictions.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Mignone, A., Grand, A., Fiori, A., et al., 2019. Semalytics: A semantic analytics platform for the explo-

ration of distributed and heterogeneous cancer data in translational research. Database. baz080. DOI: https://doi.org/10.1093/database/baz080

[2] Wild, C.P., Stewart, B.W. (Eds.), 2014. World Cancer Report 2014. World Health Organization: Geneva, Switzerland.

[3] Soerjomataram, I., Bray, F., 2021. Planning for Tomorrow: Global Cancer Incidence and the Role of Prevention 2020–2070. Nature Reviews Clinical Oncology. 18(10), 663–672.

[4] Louw, B., 1993. Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In Text and Technology: In Honour of John Sinclair. John Benjamins: Amsterdam, Netherlands. pp. 156–159.

[5] McEnery, T., Xiao, R., Tono, Y., 2006. Corpus-Based Language Studies: An Advanced Resource Book. Taylor & Francis: London, UK.

[6] Mansoor, K.R., 2017. Investigating Semantic and Syntactic Equivalence in the Translation Process. Editura Politehnica: Timişoara, Romania.

[7] Azizmohammadi, F., Barjesteh, H., 2020. Studies in Ethnopragmatics, Cultural Semantics, and Intercultural Communication: Ethnopragmatics and Semantic Analysis, Lauren Sadow, Bet Peeters & Kerry Mullan (Eds.) (2019), Springer, ISBN 978-981-32-9982-5. International Journal of Society, Culture & Language. 8(1), 126–131.

[8] Starcke, B., 2008. I Don't Know–Differences in Patterns of Collocation and Semantic Prosody in Phrases of Different Lengths. In Language, People and Numbers. Corpus Linguistics and Society. Brill: Amsterdam, Netherlands. pp. 199–216.

[9] Jiang, G., Solbrig, H.R., Chute, C.G., 2011. Quality Evaluation of Cancer Study Common Data Elements Using the UMLS Semantic Network. Journal of Biomedical Informatics. 44(Supplement 1), S78–S85.

[10] Hunston, S., 2007. Semantic Prosody Revisited. International Journal of Corpus Linguistics. 12(2), 249–268.

[11] Milojkovic, M., Louw, B., 2017. Towards a Corpus-Attested Definition of Creativity as Accessed Through a Subtextual Analysis of Student Writing. In Essential Competencies for English-Medium University Teaching. Springer: Singapore. pp. 125–141.

[12] Liu, Y., Tang, Y., Cao, Z.B., et al., 2019. Results from the China 2018 Report Card on Physical Activity for Children and Youth. Journal of Exercise Science & Fitness. 17(1), 3–7.

[13] Baker, S., Silins, I., Guo, Y., et al., 2016. Automatic Semantic Classification of Scientific Literature According to the Hallmarks of Cancer. Bioinformatics. 32(3), 432–440.

[14] Miranda-Escalada, A., Farré, E., Krallinger, M., 2020. Named entity recognition, concept normalization and clinical coding: Overview of the cantemist track for cancer text mining in Spanish, corpus, guidelines, methods and results. In Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020), Málaga, Spain, 23 September 2020; pp. 303–323.

[15] Shah-Mohammadi, F., Finkelstein, J., 2024. Addressing Semantic Variability in Clinical Outcome Reporting. BioMedInformatics. 4(4), 2173–2185.

[16] Yada, S., Nishiyama, T., Wakamiya, S., et al., 2025. Utility analysis and demonstration of real-world clinical texts: A case study on Japanese cancer-related EHRs. PLoS One. 20(3), e0320522.

[17] Cui, M., Bai, R., Lu, Z., et al., 2019. Regular Expression Based Medical Text Classification Using Constructive Heuristic Approach. IEEE Access. 7, 147892–147904.

[18] Semchyshyn, V., Mykhalyk, D., 2023. Automated processing and analysis of medical texts. In Proceedings of the 3rd International Workshop on Information Technologies: Theoretical and Applied Problems 2023, Ternopil, Ukraine, Opole, Poland, 22–24 November 2023; pp. 300–305.

[19] Pépin, I., Alaca, F., Zulkernine, F., 2024. Privacy-Preserving Multi-Party Keyword-Based Classification of Unstructured Text Data. In Proceedings of the 20th International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT), Abu Dhabi, United Arab Emirates, 29 April–1 May 2024; pp. 203–210.

[20] Dornyei, Z., 2007. Research Methods in Applied Linguistics. Oxford University Press: Oxford, UK.

[21] Anthony, L., 2011. AntConc (Version 3.2.4w) [Computer Software]. Waseda University: Tokyo, Japan.

[22] Ludwig, J.M., Ambinder, E.M., Ghodadra, A., et al., 2016. Lung Shunt Fraction prior to Yttrium-90 Radioembolization Predicts Survival in Patients with Neuroendocrine Liver Metastases: Single-Center Prospective Analysis. Cardiovascular and Interventional Radiology. 39(7), 1007–1014.

[23] Hughes, T.M.D., Elton, C., 2002. The Pathophysiology and Management of Bowel and Mesenteric Injuries Due to Blunt Trauma. Injury. 33(4), 295–302.

[24] Le, T.B., Moon, M.H., Jeon, Y.S., et al., 2016. Evaluation of Aneurysm Neck Angle Change After Endovascular Aneurysm Repair Clinical Investigations. Cardiovascular and Interventional Radiology. 39(5), 668–675.

[25] Nestle, U., Schimek-Jasch, T., Kremp, S., et al., 2020. Imaging-based target volume reduction in chemoradiotherapy for locally advanced non-small-cell lung cancer (PET-Plan): A multicentre, open-label, randomised, controlled trial. 21(4), 581–592.

[26] Masuda, K., Yamazoe, S., Baba, A., et al., 2021. Superselective intra-arterial chemoradiotherapy using altered blood flow compared to conventional systemic chemoradiotherapy for locally advanced oral squamous cell carcinoma: A single-center retrospective study.

Oral Radiology. 37(4), 700–706.

[27] Schnell, J., Beer, M., Eggeling, S., et al., 2018. Management of Spontaneous Pneumothorax and Postinterventional Pneumothorax: German S3-Guideline. Zentralblatt für Chirurgie. 143(S01), S12–S43.

[28] Klok, F.A., Ageno, W., Ay, C., et al., 2022. Optimal follow-up after acute pulmonary embolism: A position paper of the European Society of Cardiology Working Group on Pulmonary Circulation and Right Ventricular Function, in collaboration with the European Society of Cardiology Working Group on Atherosclerosis and Vascular Biology, endorsed by the European Respiratory Society. European Heart Journal. 43(3), 183–189.

[29] Feng, S., Bose, R., Choi, Y., 2011. Learning General Connotation of Words Using Graph-Based Algorithms. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, Edinburgh, UK, 27–31 July 2011; pp. 1092–1103.

[30] Perez, G.K., Mutchler, J., Yang, M.S., et al., 2016. Promoting quality care in cancer patients with Limited English Proficiency: Perspectives of medical interpreters. Psycho-Oncology. 25(10), 1241.

[31] Aasen, E.M., Crawford, P., Dahl, B.M., 2020. Discursive construction of the patient in online clinical cancer pathways information. Journal of Advanced Nursing. 76(11), 3113–3122.

[32] Selmistraitis, L., 2020. Semantic preference, prosody and distribution of synonymous adjectives in COCA. GEMA Online Journal of Language Studies. 20(3), 1–18.

[33] Bi, Z., 2019. A semantic prosody analysis on two synonymous pairs in English Native Speakers' and Chinese learners' writings. English Language Teaching. 12(8), 14–19.

[34] Sinclair, J., 2004. Trust the Text: Language, Corpus, and Discourse. Routledge: London, UK.

[35] Hoey, M., 2005. Lexical Priming: A New Theory of Words and Language. Routledge: London, UK.