Journal of the Patent and Trademark Resource Center Association

Volume 34 Article 3

8-5-2024

How to read a patent: A survey of the textual characteristics of patent documents and strategies for comprehension

Graham Sherriff
University of Vermont, graham.sherriff@uvm.edu

How to read a patent: A survey of the textual characteristics of patent documents and strategies for comprehension

Graham Sherriff

The University of Vermont
graham.sherriff@uvm.edu
https://orcid.org/0000-0003-1454-0705

Abstract

Reading patents is an important activity for inventors and anyone seeking to file or defend a patent, as well as for "exploratory" researchers such as students in a range of disciplines. However, they are notoriously difficult to read. This paper examines the characteristics of patent documents that impair their readability and seeks to identify comprehension strategies and techniques that may alleviate this difficulty. Insights were gathered from a review of the scholarly literature on reading and patent literacy, a survey of patent educators affiliated with the US Patent and Trademark Office (USPTO)'s Patent and Trademark Research Center (PTRC) Program, and interviews with engineering and legal professionals who read patents in the course of their professional practice. The resulting compilation of recommended strategies for patent comprehension is synthesized as a possible basis for best practices for reading patents.

Keywords: patents, patent documents, reading, readability, comprehension

Introduction

One of the foundational principles of any patent system is the dissemination of knowledge. In exchange for a time-limited exclusive right to commercial exploitation of an invention, inventors must agree to the public disclosure of the invention and its workings. This rewards invention and acts as incentive for further experimentation and invention. "Patent information is a valuable and comprehensive source of technical, commercial and legal information that can be used directly for scientific and experimental purposes and as a basis for stimulating the adaptation and improvement of the technology described in patent documents immediately after their publication." (World Intellectual Property Organization, n.d., Introduction) Importantly, up to 80% of the information in patents may be unavailable in any other type of publication. (Asche, 2017)

Reading patents is an essential activity for inventors and anyone seeking to file or defend a patent. It is also necessary for patentability or infringement analysis, in other words for determining the scope of a patent's claims or its novelty. According to Schox (2015), a comprehensive review of prior art should entail reading approximately 500 relevant patents. Also, as one of the patent attorneys interviewed in this study noted, patent analysis and filing strategy need to be grounded in the careful reading and correct interpretation of references, because rejections and amended claims equate to additional legal fees and examination fees.

Reading patents is also a valuable activity for "exploratory" research (Zwicky, 2016, p. 79). For example, engineers, chemists, and entrepreneurs may read patents to understand a technology's state of the art or identify directions for future innovation. Students in STEM fields may research intellectual property (IP) literature to build their knowledge of various technologies and prepare to enter the workforce in roles where they may need to dig deep into technical information. MacMillan (2005) suggests students should acquire the competencies of "patent literacy", based on the student learning outcomes of being able to (1) determine the purpose of a patent, (2) understand the data contained within a patent, (3) relate the

information to the assignment, and (4) cite patents. "It is not enough to be familiar with the search protocols for various patent databases; understanding and interpreting these often unfamiliar, jargon-filled resources are equally important skills" (p. 151). Researchers in a diverse range of non-STEM fields may also find value in patent documents, such as graphic designers, historians, and business researchers (Zwicky, 2019). K-12 students too can learn from the technical information found in patent documents (Carlson & Sullivan, 2004).

For all of these types of researchers, it is usually highly advisable that a significant amount of time be invested in the careful reading of patent documents.[1] However, it is a paradox of the patent system that one of its foundational purposes is the dissemination of technical information across society, yet patent documents have very low readability for almost all readers. They are frequently described as "notoriously" difficult to read (e.g., Mille & Wanner, 2008, p. 1393; O'Toole, 2021, p. 4). In a user study with patent professionals, "lay" users, and other study participants, "Documents were seen as extremely (29%) or somewhat (29%) difficult to read. 39% had some difficulties in finding information." (Suominen et al., 2017, p. 5) Scientific papers and patent applications have lower rates of citations to patents than might be expected, possibly due to the difficulty of reading them, among other reasons.

Patents are a unique type of publication, and their structure, language, and information content can be challenging for anyone not experienced in reading patent documents or not fluent in the technology they describe. As a patent educator who also provides research assistance to members of my library's community and service area, this author has observed how even technically knowledgeable inventors and engineers sometimes have difficulty understanding the information in a patent document.

At the college level, the obstacles to comprehension may be even greater. As an engineering librarian, this author provides patent instruction to undergraduate students conducting exploratory research into the prior art in various fields of engineering and computer science. This instruction has focused on key patent concepts, the features of patent databases, and search techniques - the essential tools and techniques of patentability searching. Judging from conversations with students and the assessments of their research reports, this instruction equips them to find patents in the appropriate technological area. However, it does not support them when it comes to understanding patent documents and patent language. These students are at an early stage in their development as engineers and their scientific and technical knowledge, as well as their familiarity with IP concepts, is not advanced. They struggle to make sense of the information in patent documents or to extract information that they can apply to their design projects.

This paper examines the textual characteristics of patent documents that impair their readability. It surveys the scholarly and professional literature on patent literacy and reading comprehension, and gathers insights from two expert groups: patent educators, specifically librarians who provide patent research assistance in their capacity as representatives of US Patent and Trademark Office (USPTO)-affiliated Patent and Trademark Resource Centers (PTRCs); and patent professionals who read patents in the course of

[1] Is there a reason not to read patents? The literature review for this study identified a single counterpoint. Ouellette (2017) contends that patent research may increase the risk, or at least the fear of a risk, of increased liability for "willful" patent infringement. This concern may have grown since the US Supreme Court's ruling on Halo Electronics, Inc. v. Pulse Electronics, Inc., 579 U.S. 93, which broadened the standard for determining willful infringement.

their professional practice, such as engineers, technology transfer specialists, patent attorneys, and patent examiners. This paper then presents a synthesis of recommended strategies for patent comprehension as a possible basis for best practices for reading patents.

2. Methods

The surveys of patent educators and patent professionals used convenience sampling methods, due to the study's exploratory nature and its focus on qualitative data that might not be generalizable. Subjects were invited to participate based on their relevance to the objectives of the study rather than their representativeness of the general population.

The survey of patent educators was distributed to PTRC representatives in January-February 2024. (See Appendix A.) PTRCs are academic, public, and state libraries that have been designated by the USPTO as resource centers that can assist the public with patent and trademark services. PTRC representatives are trained to instruct inventors, researchers, and others in the use of patent and trademark search tools, application processes, finding legal services, and connecting with USPTO resources and services. They are not able to provide legal assistance (see 4.3.). [(US Patent and Trademark Office, n.d.)]

Questions were designed to elicit PTRC representatives' observations regarding reading difficulties and their experiences and insights regarding how PTRCs can assist patent readers. The survey was approved by the University of Vermont Institutional Research Board (IRB) and exempted from further review (STUDY00002917). An invitation to participate was sent to all persons named as PTRC representatives in USPTO PTRC Program documentation. This was a total of 128 librarians, representing 87 PTRCs (i.e. all PTRCs at that time except the University of Vermont for which the author is the representative).

Twenty-one PTRC representatives participated. Twenty reported the type of library with which they were affiliated: these were a mix of mostly academic libraries (n=14), a smaller number of public libraries (5), and one state library. All 21 PTRC representatives reported working with inventors. Smaller numbers reported working with entrepreneurs or small business owners (17), academic researchers (14), business employees or contractors (7), patent agents or patent attorneys (5), graduate students (3), genealogists and historians (3), undergraduate students (3), post-doctoral researchers (1), and K-12 students (1).

Patent professionals were invited to participate in interviews based on the author's professional connections and recommendations from University of Vermont Engineering faculty. Recruiting concluded after finding participants from a range of work settings and experience levels. Participants were one practicing engineer, one technology transfer specialist at a research-grade university, two patent attorneys ("patent attorney A" was in-house counsel for an industrial corporation, "patent attorney B" was retired), and one patent examiner[2]. Interviews were conducted in April-May 2024. Three were conducted, recorded, and transcribed in MS Teams; two participants participated with written responses. The study used a semi-structured interview protocol: a common set of questions was supplemented with follow-up questions that arose from the conversation and thus varied between interviews. The questions are presented in Appendix B. The survey was approved by the University of Vermont IRB and exempted from further review (STUDY00002975).

[2] The patent examiner's comments represented their individual views and did not necessarily represent the views or opinions of their office.

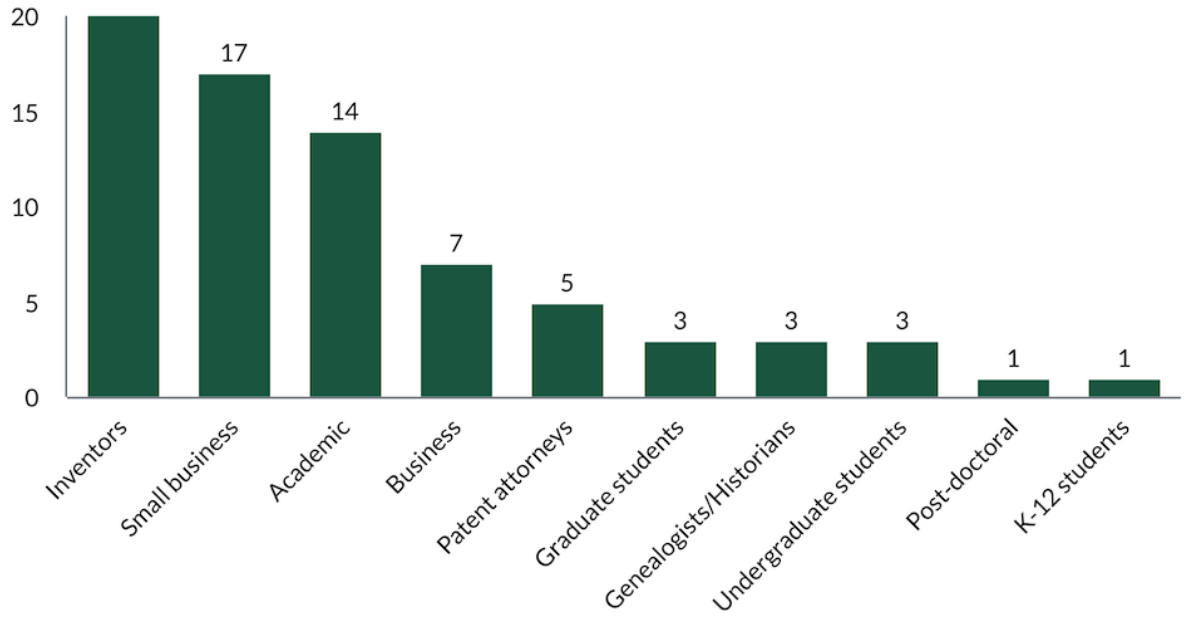


Figure 1. Types of researchers reported by PTRC representatives.

3. The textual characteristics of patent documents

Why are patents so hard to read? A 2023 meta-analysis of studies on reading comprehension concludes that "comprehending texts is a complex cognitive task that is influenced by the linguistic features of the text and various aspects of the reader" (Strohmaier et al., 2023, p. 1). In other words, the readability of a text is determined by the characteristics of the text and by the characteristics of the reader.

This survey of the characteristics of patent documents will use a three-part model of the cognitive process of reading which, per Strohmaier et al., represents the current scholarly consensus on reading comprehension. Together, these three aspects form a model for how a reader converts textual information into a personal mental representation of that information:

- Decoding "lexical units" or determining the meaning of words and word combinations.
- Determining "semantic connections" or interpreting syntax and structure in order to understand the larger meaning of the text.
- Creating a "knowledge structure" that is integrated with the reader's prior knowledge.

The characteristics of patent documents can be mapped to these aspects and so the model provides a framework for analyzing why patents are hard to read.

3.1. Decoding "lexical units"

Readability research in the United States began in the 1930s with the research and writing of literacy advocate William Gray. Gray and Leary (1935) studied 289 elements that can affect readability and concluded that the two most-significant are the frequency of advanced or uncommon vocabulary and the average sentence length as measured in words. Subsequent research has reaffirmed that the presence of complex vocabulary is one of the best predictors of low readability.

STEM texts generally "use technical vocabulary and academic language, which are more difficult to decode, trigger fewer background knowledge and spread less cognitive activation" (Strohmaier et al., 2023, p. 3). The vocabulary in patents is accordingly complex. Technical vocabulary may be advanced or discipline-specific. Compounding this complexity, legal terminology may be obscure or non-intuitive. Several PTRC representatives commented on this, for example "The technical language is not always straightforward." "Not just jargon related to the field of invention, but terms like "proximal" or "embodiment" that most people NEVER use."

Zwicky notes that patent language can be "archaic and obsolete… convoluted and obfuscatory" (2016, p. 80) to avoid unnecessarily limiting the scope of claims, or possibly in order to make it more difficult for a patent to be found in database searches. For example, a search using the keyword "bicycle" may miss patents that refer to "two-wheeled, occupant-propelled vehicles". Claims sections often refer only to a "system", an "object", or a "medium". In this sense, patents may meet the requirement for disclosure of information, but without clarity of information, with the result that readability is low.

Patent readers need to be able to understand the definition of words. Two PTRC representatives and the technology transfer specialist offered the opinion that standardized vocabulary for elements, materials, processes, or functions would be beneficial for readers, but there is no such controlled vocabulary. Exacerbating this situation, dictionary definitions and common usages cannot always be relied on. A patent applicant may define words and phrases in a specification however they may wish. (Segal, 2019). Patent attorney A described how their work is mostly focused on claims, but "you might need to [refer] to the detailed description to understand what the patent team means by a claim term."

Given that a patent reader cannot depend on standard, or standardized, vocabulary, they must fathom the widest breadth of vocabulary that might be used to describe something. The engineer described how they try to brainstorm all possible synonyms and related terms when working with patents. But it is difficult to think beyond what one knows: "Novice patent searchers often have a fixed idea of the terms to describe their invention and if they come across alternative, but relevant, terms in patent results they might discount those results." (PTRC representative)

Legal language can be particularly challenging. "Wordings are careful, resulting in arcane legal jargon". (Suominen et al., 2017, p. 1). For example, transitions in the wording of claims have a nuanced precision: a system "comprising" a specified feature might include additional ones, whereas a system "consisting of" a feature excludes others. Another semantic distinction that might easily be misunderstood or missed altogether is the difference between an example written in the past tense (a "working" example that has been demonstrated) and an example written in the present tense, which is a "prophetic" example - "the inventor's best guess as to what might happen under hypothetical conditions." (Robson, 2001, p. 73)

Many patent researchers lack the legal expertise needed to understand this legal language and its significance. Fifteen of 21 PTRC representatives reported that researchers commonly have difficulty understanding the legal language used in the Claims section. "[Most inventors] have difficulties with the legal language". "Terms often have specific legal meanings that differ from everyday usage; this unique and precise vocabulary might be unfamiliar to individuals who lack legal training." "Patrons at this PTRC are almost entirely unfamiliar with the [Manual of Patent Examining Procedure] and patent law, and are confused by references to specific rules." The

technology transfer specialist echoed this when explaining how their work sometimes involves guiding engineers, scientists, and entrepreneurial interests through legal processes and legal language.

Patent attorney A commented that translations of non-English patents present another language-based difficulty. “Mostly when I have trouble understanding, it's because of machine translations of foreign references. They've gotten better over the past 10 years, but still, sometimes they are hard to understand.”

3.2. Determining “semantic connections”

The second part of Strohmeier et al.’s model of reading comprehension is the determining of “semantic connections”, or making sense of the information represented by the combination of lexical units. This is “particularily [sic] challenging for STEM texts because they are typically informational or expository texts with a [non-linear] logical semantic structure” (3). In a patent document, both the Specifications and Claims sections are expository and non-linear in terms of both time and hierarchy of concepts, similar to other types of scientific and technical texts (Follmer et al., 2018).

3.2.1. The unique characteristics of patents

The patent is a unique format of published information and making sense of them is complicated by the organization, content, and purpose of a patent’s different sections. For novice patent researchers, these are likely to be unfamiliar and non-intuitive. Multiple PTRC representatives observed that they often work with researchers who have not worked with patents before and are unfamiliar with the structure, language, and content of a patent document. The patent examiner’s experience in college was that ”as a Mechanical Engineering student, we very briefly learned about patents but never spent much time at all on the topic even though we all worked on design projects to solve a problem.” Sherriff and Rand (2022) found that senior engineering students scored an average of 60% on a pre-test assessment of their understanding of basic patent concepts, compared with 76% for their understanding of journal articles and conference papers. Kim (2015) proposed guidelines for reading patents because, according to their observations, their electrical engineering colleagues at Samsung Electronics “have little knowledge of how to read or construct a patent” (p. 1).

Several PTRC representatives reported that researchers of all types are often unfamiliar with the organization of a patent document, sometimes referred to as the “anatomy” or “architecture” of a patent. Thirteen of 21 representatives said that researchers often do not know what the various cover-page fields represent, nor the codes and classes that they contain. “I work almost entirely with novice patent searchers. They don't know the meaning of the various classification system acronyms. They don't know the meaning of assignee, field of classification search, and other section headings.” “The cover page looks daunting, with a raft of unexplained terminology and obtuse codes.” PTRC representatives noted that unfamiliarity with classification codes sometimes presents difficulties for distinguishing between utility patents, design patents, and plant patents; and between applications and grants. PTRC representatives also described how researchers are often unfamiliar with the purpose and significance of the main Specifications and Claims sections.

Follmer et al. (2018) observe that readers of scientific and technical texts are sometimes aided by “signaling” of the meaning of different sections. An important example of signaling is the presence of clear and intuitive section headings, which can aid comprehension and navigation. Unfortunately, the signaling within patent documents is not optimal. There are “no explanations of the sections, but they also are not intuitive.” (PTRC)

Participants in Suominen et al.'s user study similarly exhibited confusion about layout and organization, and a preference for more-intuitive subheadings (2017). For some researchers - especially novice patent researchers - field labels may be unfamiliar, such as "Int. Cl." and "U.S. Cl." lens.org's guide to reading a patent (2020) emphasizes the need to understand the purpose of each section, which may not always be self-evident. For example, the "Background" section presents not a fully contextualizing background, but instead "selected art in the field, justifying the invention's need"; the "Summary" is not a summary, but actually a discussion of the invention's novelty; the "Specifications" section is not a comprehensive description, but "drafted both to satisfy the written requirements for patentability as well as to define claim scope" (Specification – description of the invention).

3.2.2. Overwhelming detail

As is true of most STEM publications, patent documents are highly detailed. Patent law's "enablement" requirement stipulates that a patent document must disclose information that is sufficiently detailed "to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same" (Patents Act of 1952, 35 U.S.C. § 112, para. 1). The USPTO's MPEP interprets this as meaning information regarding not just an invention's elements, but also its assembly, functions, and processes - whatever is needed to show its operability. (2022) This level of detail can be overwhelming for some readers. One PTRC representative named "info overload" as the most-frequent difficulty that researchers encounter on the cover page. Exacerbating this difficulty, STEM publications tend to present details with equal importance. It is not always apparent which details are more significant and which are subordinate or secondary. (Howland et al., 1943)

3.2.3. Insufficient detail

Despite the presentation of a high level of detail, some detail that might aid comprehension can still be absent. Also, detail can be present but unclear. Although patent law requires the disclosure of sufficient detail for enablement by "any person skilled in the art", patent documents are sometimes perceived as vague, confusing, obfuscatory, or insufficiently detailed. Both the engineer and patent attorney B expressed frustration that patents can lack sufficient detail. In one survey of researchers in industry, academic, government, and non-profit sectors, vagueness and insufficient detail were reported as two of the most-common reasons for not using patents (Ouellette, 2017). Patent applications are required to be accurate but are not obligated to present information beyond the requirements to show feasibility and enable reproduction by a person skilled in the art. Robson (2001) notes that this can result in the omission of contradictory or unsupportive data. "Data selectivity makes patents a suspect source" (p. 73).

The meaning of patent language can also be unclear because of a lack of textual "cohesion", defined as explicit clues to relationships to elements within and across sentences (Best et al., 2005) Cohesion can take the form of descriptive subheadings, summary paragraphs, connecting language, explanations of difficult terms, and background information - all of which patents typically lack. Where the cohesion of a text is low or impaired, "[t]he meaning of a text often remains fragmented and disconnected without inferences because texts normally do not (or cannot) state all the information relevant to the situations or events" (p. 66). This requires the reader to make inferences about the text. But, as discussed below (3.3.1.), some types of patent reader commonly lack the subject knowledge to make the inferences necessary to fill the information gaps in a patent document.

3.2.4. Complex syntax

STEM texts, including patents, tend to have a complex syntax that makes it more difficult to parse sentences and extract meaning. Also, the non-linear nature of STEM texts adds demands on readers' working memory by increasing the information required to be processed simultaneously. (Strohmaier et al., 2023) PTRC representatives confirmed this, underscoring the complexity of the Claims section: "Parsing the syntax of the claims is difficult at the best of times. I usually make a joke about the claims being written to intentionally confuse readers, and that's barely a joke. The nesting structure is often confusing."

Another notable characteristic of the syntax of patent documents is the length of sentences. As noted above (3.1.), scholarship on readability dating back to Gray and Leary's pioneering work in the 1930s has consistently shown that one of the most significant negative factors is sentence length as measured in words. Patent sentences are typically longer than common language - not least because patent rules typically require each claim to be written in a single sentence. Verberne et al. (2010) found a median length for sentence segments of 22 words in patents[3], compared with a median length for whole sentences of fewer than 10 words in the British National Corpus, while sentences with more than 200 words are not uncommon.

3.3. Creating "knowledge structures"

The third part of Strohmaier et al.'s model of the cognitive aspects of reading is the integration of the reader's understanding of the text with their prior knowledge, creating a new "knowledge structure".

This is challenging because the knowledge represented in STEM texts is "often unfamiliar, difficult to map, and may even conflict with prior knowledge" (Strohmaier et al., 2023, p. 3).

3.3.1. "Skill in the art"

Several studies of reading comprehension suggest that the most significant limiting factor for readers of scientific or technical texts may be an insufficient level of prior domain-specific knowledge (Best et al., 2005; Follmer et al., 2018; Strohmaier et al., 2023). Ideally, this knowledge should be "rich" and "organized" (Best et al., p. 68). Prior knowledge is critical for generating inferences, filling in the gaps in the text's information, and minimizing reliance on that text. This is highly important when the text has a low degree of cohesion, like descriptive subheadings or explanations (see 3.2.3.).

Consequently, prior knowledge is especially significant for the comprehension of patents, whose intended reader has sufficient technical expertise to understand the content of that document. As mentioned, the enablement requirement necessitates the disclosure of only a level of information that is sufficient "to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same" (Patents Act of 1952, 35 U.S.C. § 112, para. 1).[4] Indeed, the MPEP states that "A

[3] The authors of this study describe the patents that they studied as being English-language, but do not specify the issuing office(s).

[4] In U.S. patent law, the enablement requirement takes into consideration "any person skilled in the art to which it pertains, or with which it is most nearly connected". This can be compared with the non-obviousness requirement, which takes into consideration "a person *having ordinary skill* in the art to which the claimed invention pertains" (35 USC § 103, author's emphasis), sometimes abbreviated in legal studies as "Phosita". Pedraza-Fariña and Whalen (2022) reviewed the definitions of these two persons applied in 700 court opinions, as well as conducting an automated text analysis of more than 7,000 cases, and found that most court decisions made no distinction between the two. However,

patent need not teach, *and preferably omits*, what is well known in the art." (US Patent and Trademark Office, 2022, section 2164, author's emphasis)

Who is the person "skilled in the art" anticipated by the enablement requirement and what prior knowledge do they have? Pedraza-Fariña and Whalen (2022) argue that a clear definition has not emerged from court decisions, but court decisions have tended to characterize this person as someone with the skills and background technical knowledge - in all relevant disciplines - to read the patent disclosure and make it or use it with a minimum of experimentation. Accordingly, court decisions have indicated that the person skilled in the art - and therefore capable of understanding a given patent - has an educational level, length of professional experience, and knowledge of sub-disciplines that are appropriate to the technology under consideration. This will vary according to the technology. Meara (2002) illustrates the possible variety in appropriate levels of skill: one court ruling on fly wraps for the legs of horses characterized the person skilled in the art as a person with some formal education but no special skills or training in that art, whereas another ruling on a pharmaceutical patent found the person skilled in the art to be a Ph.D.-bearing organic chemist. Overall, they suggest that the common denominator in federal circuit rulings has been a combination of a completed post-secondary credential and some professional experience.

This, then, is the expected level of skill and knowledge for someone seeking to read and understand a patent to the degree necessary to replicate, use, synthesize, or cultivate the invention that it describes. Some inventors have not completed a post-secondary educational program and would not therefore be considered to meet the definition of a skilled person. Several PTRC librarians affirmed that many patent researchers do not have the technical or scientific expertise needed to understand some content. "Advanced concepts and theories in technical and scientific content can be challenging to grasp, especially for individuals without a strong foundation in the subject matter." "Often the needed level of expertise in a field is well beyond what my patrons have." Several commented that disciplinary knowledge makes a major difference to their comprehension. "STEM professionals seem to be more comfortable understanding the technical or scientific content since they are subject matter experts." The engineer noted that their ability to understand patents developed commensurately with their accumulation of experience in their field.

This problem can be particularly acute in the case of "exploratory" patent research by non-experts, such as students. "It is the nature of education that readers of science texts read them for the purpose of learning new information… The circularity of this phenomenon is such that readers cannot comprehend the text contents at a deep level without learning new concepts or information from the text" (Best et al., 2005, pp. 75-76). PTRC representatives confirmed this is true for students at all levels, from primary to post-secondary. "For the k-12 students [it] is most difficult to understand the technical or scientific content."

It can be an even more acute problem when the technology is advanced, complex, or interdisciplinary. As Meara explains, the more advanced the technology, the higher the levels of skill and knowledge that the skilled person will have. In the words of the MPEP, "The amount of guidance or direction needed to enable the invention is inversely related to the amount of

they also found that a small number of rulings held the non-obviousness Phosita to higher degrees of skill and knowledge than the enablement Phosita, specifically regarding characteristics such as entrepreneurial activity, awareness of market trends, and access to relevant prior art.

knowledge in the state of the art" (US Patent and Trademark Office, 2022, section 2164). But if this person has advanced levels of skill and knowledge, it may be argued that less disclosure and less textual cohesion are required for enablement, because they are more capable of making inferences. And if there is a lesser degree of disclosure and cohesion, the more challenging it is for a person who is not skilled in the art to understand the information presented.

3.4. The situational aspects of reading

The fourth part of Strohmaier et al.'s model of reading comprehension is not cognitive, but situational. The following situational factors can each be an obstacle to comprehension; comments by PTRC representatives confirmed they are common across all the types of researchers that they support, including inventors, college students, academic researchers, engineers, and genealogists.

3.4.1. Affect

Affect is synonymous with mental state and refers to variables such as degrees of interest, motivation, or attention. For example, comprehension can be negatively impacted when a reader does not value cognitively-challenging reading or when they do not see the relevance of a text to their broader need for information. (Follmer et al., 2018)

Disinterest and demotivation can be heightened when the reader is a non-expert and lacks the prior knowledge that facilitates the reading of complex texts - as is often the case for patent researchers. This is very much a problem for students when feeling conscious of a gulf between expectations (whether their instructors' or their own) that they should be able to understand a text and the reality that understanding is not immediately forthcoming. As a result, "Negative feelings about their own efficacy as readers leads to procrastination or avoidance." (MacMillan & Rosenblatt, 2015a, Challenges of Academic Reading) Three PTRC representatives reported that they worked with undergraduate classes where students were often taken aback by the complexity of patent language; one of them speculated that this may be causing some students to avoid patents as an information source and seek out alternatives.

A reader's engagement with a text can of course be affected by external factors - not least, time constraints. Two PTRC librarians observed that time constraints were the most common difficulty for patent researchers; one of them said this was true for all types of researchers. Another stated "these documents can be lengthy and too time consuming to read." Similarly, both patent attorneys, the engineer, and the technology transfer specialist commented that they felt a conscious need to read quickly and efficiently. The engineer stated that they sometimes felt unable to return to parts of a patent document to re-read them, or look up explanations, due to time pressures.

3.4.2. Reading competencies and strategies

The difficulties of reading patents are magnified for non-native speakers of English. One PTRC representative described how the language barrier is sometimes too high for some researchers: "In our case, most of our patrons are Spanish speakers and I need to translate the patent information or search for a Spanish language version of a patent".

Whatever a researcher's degree of fluency in the English language, they may lack appropriate techniques or strategies for comprehending a patent document. None of the patent professionals had received any instruction regarding how to read a patent. Patent attorney A explained that they had had no such instruction at the undergraduate level, nor at the Master's level, nor in nine years as an engineer in industrial manufacturing, nor even at patent law school. "No,

[I was not taught how to read a patent at law school] - it's trial by fire."

Some patent researchers may lack basic reading skills or techniques. According to Department of Education data, 52% of U.S. adults are considered partially or non-literate, reading below the equivalent of a sixth-grade level (U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics, n.d.). Indeed, in U.S. public schools, instruction in reading typically stops at the fifth or sixth grade (Lei et al., 2010). Instruction in reading is also usually absent at the college level, despite the increased complexity of texts in post-secondary education. This is a major gap in information literacy instruction and learning. "Traditional information literacy (IL) instruction [at the college level] focuses on finding, evaluating, and citing materials but seldom addresses how students will actually use the resources they find." (MacMillan & Rosenblatt, 2015b, p. 757). Reading is a prerequisite for several of the ACRL's IL frames (e.g., "Research as Inquiry", "Scholarship Is a Conversation") and is assumed to be a previously-acquired competency. But many students lack college-level reading proficiency. As a result, students feel frustrated and cope by "cherry-picking numbers and other details at random and otherwise misusing the information… but don't always manifest the depth of understanding and integration of knowledge envisaged by instructors." (MacMillan & Rosenblatt, 2015b, p. 757)

4. Strategies for comprehension

The deficiencies in reading instruction described above stand in contrast with the availability of strategies for reading comprehension that have been developed over a long period of time - beginning with Gray's work in the first half of the 20th century - tested, and shown to be effective. This section presents reading strategies that patent researchers can use to overcome or mitigate many of the obstacles to comprehension that were described in the previous section.

4.1 The act of reading: The SQ5R model

Several techniques have been shown to be effective for supporting the act of reading, with positive outcomes for both comprehension and recall. The use of reading strategies may also have a positive effect on the reader's confidence and motivation, which may address the affective dimension discussed above (Artis, 2008).[5]

"Among the most popular methods used in college reading courses" (Artis, p. 131) is "SQ3R" and its variations. This method is a sequence of strategies for self-directed reading that support comprehension, with several metacognitive elements that encourage the reader's monitoring of their comprehension. It was originally formulated in 1941 as "SQ3R": survey, question, read, recite, review. (Robinson, 1941) Since then, several variations have been proposed as research on reading comprehension has pointed to additional strategies that can be included. The model presented here is a "SQ5R" model: survey, question, read, respond, record, recite, review (Lu et al., 2022). This is one of the most-recent iterations of the model and it applies well to this study because of its expanded range of strategies.

4.1.1 Survey

Comprehension should begin not with line-by-line reading, but with a preliminary scan of the whole document. The reader should seek to understand the text's organization and section structure. It is important to "see the big picture" while also

[5] As well as benefiting reading comprehension, the use of reading strategies may also be beneficial for disciplinary learning. Wilson-Lopez et al. (2017) ran a study where K-12 students read texts describing problems that could be solved through engineering, and found that 80.5% of comments that were coded as "comprehension strategy" were also coded as "engineering design process". They concluded that "comprehension strategy instruction and engineering design instruction can be conceptualized as complementary rather than competing" (p. 1).

thinking about "what they need from the various sections" (Artis, 2008, 131). Scanning should also help the reader to consider how well they know the subject matter; and estimate how much time to budget for fully reading the text.

In a patent document, a preliminary scan should focus on the following:

- Title: It is important to take note of the title, while also recognizing that titles usually provide limited information. Several PTRC representatives warned that titles cannot be trusted to provide a clear description of the device, due to the generic and opaque vocabulary typically used.
- Abstract: The abstract should provide a brief summary (maximum 150 words) of the invention's key features, its functions, and the problem it purports to solve. The patent examiner recommended reading the abstract "to give an overview of the invention as a whole." The reader should not seek this kind of information in the "Summary" section, which is a discussion of the invention's novelty. "Often, the summary will discuss advantages of the invention or how it solves the problems existing in the art, such as those presented in the Background" (lens.org, 2020, Specification – description of the invention).
- Drawings: Drawings can take the form of 3D views, block diagrams (showing relationships between parts), flowcharts, section views, or exploded views (which can be helpful for visualizing an invention's assembly). As Kim notes, these can be a "powerful tool for understanding" (Kim, 2015, p.31). Several PTRC representatives and several of the patent professionals, as well as the participants in Suominen et al.'s user study (2017), highlighted the value of drawings as a way to quickly gain some understanding on a general level of what the device is and what are its key features. Patent attorney A commented that "the drawings are essential, [they] are a shortcut to understanding what's going on."[6]

"Reading" patent drawings is also crucial during a close read of the text. In any document, the relationship between text and illustrations creates an additional layer of meaning (MacMillan & Rosenblatt, 2015a). Patent attorney A described how "drawings and specifications are divided even though they should be read in parallel... I'll have one window open with the figures and one with the text… Being able to look at the figures as you're reading goes a long way towards comprehension." The patent examiner's approach went a step further, by marking up drawings with reference numerals from the detailed description of the invention, aiding cross-reference between text and drawings. Suominen et al. found that most of their patent readers valued the ability to navigate cross-references between claims and figures.

- Claims: Kim (2015) recommends that a scan should also include the first claim, though not the entire Claims section. The first claim is "usually the broadest claim [addressing] the overall scope" (p. 31). Likewise, the

[6] Drawings are a critical element for preliminary scanning, but patent readers (and their instructors) should also be aware of the potential for negative effects. Koh (2020) studied the design processes of National University of Singapore (NUS) undergraduate students and found that reading patents can increase "design fixation", an excessive focus on the designs present in the patents that a researcher reviews. Reading patents may also increase "design distraction", the failure to consider design elements that are not present in the patents that a researcher reviews by the researcher. Among NUS students, these behaviors correlated with the presence of patent drawings, which appeared to encourage students to fixate on an exact physical design and design around the corresponding claim. "Drawings can fixate more than written texts" (p. 35). The implication is that drawings can be a valuable source of visual information and technical understanding, but researchers need to seek out a variety of designs and treat any one design as a source of information and inspiration before considering it as a preferred model.

technology transfer specialist emphasized the importance for their work of surveying the independent claim, as a means of quickly establishing the basic intellectual property embodied in the patent.

4.1.2. Question

The second step in the SQ5R method is posing questions that the reader hopes the patent will answer. According to Howland et al. (1943), reading any technical literature should begin with a problem and a purpose, in order to utilize an active thinking process. Active reading might also use predicting or hypothesizing, which, like questioning, creates baseline ideas to be tested, discarded, or reiterated as the reader engages with the text.

Patent attorney A confirmed that they find this method to be effective: "Generally, there are three purposes", encapsulated by three questions: for a prosecution analysis, "does the reference disclose or suggest the [application's rejected] claims?" For an infringement analysis, "does [my company's] product infringe a patent owned by another?" In a more general case, "how [has another] applicant approached the technical problem?"

The patent examiner described how they do not pose questions per se, but they do read with specific goals in mind that guide their reading. "My typical goal in reading a patent is to determine whether the patent constitutes as prior art for an application I'm examining, i.e., is the technology described therein substantially similar to the instant invention or alternatively, is the technology described in the applied-for patent an obvious variant of that disclosed in the prior art. The aim in reading the patent is to determine whether the patent can be used as evidence to reject or allow a patent application."

4.1.3. Read

After the preliminary steps of scanning and posing questions, the third step in the SQ5R method is the actual reading. This should be an active reading process, meaning that the reader should read while performing several activities, including taking note of key concepts, relationships, or examples; monitoring their comprehension; and noting new vocabulary. (Lu et al., 2022)

4.1.4. Respond

The reader should then consider whether the information answers their questions, confirms or disproves predictions, or satisfies their need for information. (Artis, 2008; Best et al., 2005)

4.1.5. Record

Either while reading or afterwards, the reader may find it useful to record some information, to facilitate understanding and to make it easier to return to it at a later time. For example, this might take the form of highlighting or note-taking. However, recording information without modifying it in ways that support integration with prior knowledge is not optimal - as seen in the previous section's discussion of "knowledge structures" (3.3.). Passive recording techniques should be compared with "generative" techniques discussed in the following sub-section (4.1.6.).

4.1.6. Recite

Recording information without modifying it can be helpful but restructuring the information in ways that make sense for the reader have a greater positive impact on comprehension - especially forms of writing, such as annotating or summarizing. "It is well established within adult learning research and instruction that reading, critical thinking, and writing are interconnected" (Artis, 2008, p. 132). "Reciting" information in one's own words requires evaluating comprehension, identifying comprehension gaps, and integrating information into personalized

vocabulary and concepts. (Artis, 2008; Best et al., 2005; Lei et al., 2010; Wilson-Lopez et al., 2017). "Generative" strategies involving a degree of authoring are thus more effective for comprehension than non-generative strategies such as highlighting or simple note-taking. "Reciting" also facilitates the next strategy in the SQ5R sequence, "review" (4.1.7.).

Many participants in this study emphasized the value of annotating and summarizing. PTRC representatives recommended it as a key reading strategy, notably several who include it in their instruction for classes and workshops. The patent examiner commented that "generally I will highlight and/or annotate in pdf and save the document so that I may come back to it later." Patent attorney B considered annotation to be highly valuable for both comprehension and for creating a historical record of their analysis.

Reciting can mean restructuring information for oneself. It can also mean restructuring information in order to communicate it to another person, whether formally in writing or informally, perhaps as a discussion between colleagues. Reciting information for another person's understanding requires careful evaluation of comprehension, integration into personal knowledge, and expression of that knowledge in accessible language. It also creates the opportunity for feedback on this expression's clarity and further examination of both clarity and meaning. In words often attributed to Richard Feynman, "If you want to master something, teach it. The more you teach, the better you learn. Teaching is a powerful tool to learning."[7]

For the patent examiner, discussion with colleagues is an important part of their reading process: "I often confer with colleagues to see if my interpretation is reasonable or if my colleagues have a similar or a different interpretation." Patent attorney A also noted that their process for reading and analysis is in some ways a collaboration, in their case with their supervising attorney, and that the need to articulate and rationalize their interpretation of a patent improves their understanding and decision-making. "No attorney can leave another attorney's writing untouched, so there's usually some revision… We talk about rejections. We discuss our independent readings of the references and the relevance. We agree on a strategy. They provide a draft. I review a comment and then they file so it is a collaborative process."

Another form of reciting that may be beneficial for patent readers is visualizing. The creation of visual images can support both comprehension and the metacognitive act of monitoring comprehension (Wilson-Lopez et al., 2017). One technique for this is sketching. Converting textual information into the reader's own drawings creates a complementary representation of that information that can reinforce understanding. It can also reveal gaps in information or gaps in understanding. Patent attorney A: "If… I'm not understanding what's going on [and] it's an assembly of some kind… I'll follow the language and I'll sketch out the structure that they're describing and that will often help."

A second technique for visualizing is concept mapping: creating a graphical representation that captures the nature and causal direction of relationships between elements. Atherton et al. (2018) argue that "structured graphical representation[s]" (p. 255) can improve the reader's understanding of a patent because a diagram transforms unstructured and uncontrolled natural language patent text into a visual representation of the invention's working principles. They recommend that patent readers first tabulate the invention's geometric features

[7] These words are commonly attributed to Feynman, though the literature review for this study was unable to identify a confirmed source.

and functional interactions. (See Table 1.) These tabulated data can then be used to create a function analysis diagram (FAD). (See Figure 2.) The authors consider this a quick and worthwhile process: 23 minutes for one example. However, they suggest it is most appropriate for mechanical inventions of lesser complexity, and for higher-skilled researchers: "We expect the level of expertise [of someone who might use this technique]… is that of a mechanical engineering graduate level of design expertise with at least 2-5 years of professional experience" (p. 257).

A third possible visualization technique is reformatting the patent text. In particular, the conventional linguistic structure of the Claims is suitable for reformatting in a way that may enhance understanding. Casola and Lavelli (2022), Jiang and Goetz (2024), and Okamoto et al. (2017) in varying ways recommend reorganizing the tree structure of Claims sections, segmenting complex claims into simpler lexical units by parsing the preamble, transition, and body. Different colors can be applied to categorize claim types (e.g., apparatus or method), transactions, and components. Suominen et al. (2017) found that both expert and lay researchers favored "enriching the claims enumeration" (p. 6) with this kind of reformatting and coloring.

Working principle				
	Geometric feature #1	Patent functional interaction term	Functional interaction RFB expression	Geometric feature #2
FGI #1	Flat sheet metal can lid member	Form	Form	Unitary structure
FGI #2	Flat sheet metal can lid member	Provide	Provide	Aperture
FGI #3	Aperture	In the surface of	Locate on	Flat sheet metal can lid member
FGI #4	Flat sheet metal can lid member	Provide	Provide	Gate panel
FGI #5	Gate panel	Underneath	Locate under	Aperture
FGI #6	Gate panel	Close	Close	Aperture
FGI #7	Flat sheet metal can lid member	Underfolded about	Surround	Aperture edge
FGI #8	Flat sheet metal can lid member	As	Form	Outward underfold
FGI #9	Outward underfold	Form	Form	Spacer strip
FGI #10	Spacer strip	At the underside of	Locate under	Flat sheet metal can lid member
FGI #11	Spacer strip	About	Surround	Aperture edge
FGI #12	Flat sheet metal can lid member	Infolded about	Surround	Spacer strip outer edge
FGI #13	Flat sheet metal can lid member	As	Dorm	Inward underfold
FGI #14	Inward underfold	Merge into	Merge	Flat sheet metal can lid member
FGI #15	Flat sheet metal can lid member	Constituting	Provide	Gate panel
FGI #16	Score cut	About	Surround	Gate panel edge
FGI #17	Score cut	Adjacent to	Locate adjacent	Inward underfold
FGI #18	Score cut	Permit	Allowseparation	Gate panel
FGI #19	Gate panel	Severed from	Separate	Spacer strip

Table 1. Tabulated feature and function data for US3334775, a 1967 patent for a gated can lid (Atherton et al., 2018, p. 255). Reproduced under the terms of the Creative Commons CC BY license.

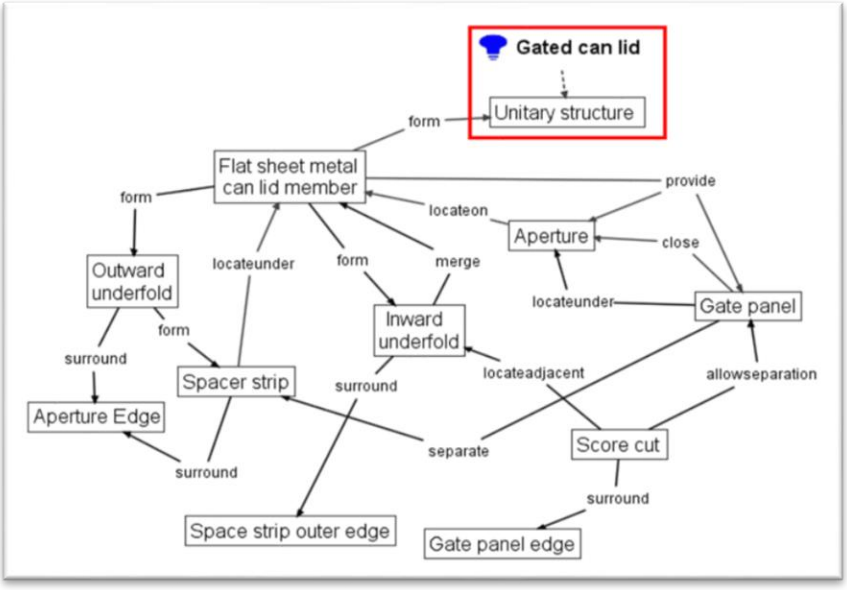


Figure 2. FAD for US3334775, a 1967 patent for a gated can lid. (Atherton et al., 2018, p. 256). Reproduced under the terms of the Creative Commons CC BY license.

Figure 3 shows Okamoto et al.'s suggested model. The authors created a language processing system that appeared to meet the criteria for initial proof-of-concept testing. A robust system of this kind would be highly beneficial for patent readers, though this is based on machine learning, and is therefore subject to several of the limitations and concerns described in section 4.4.4. Until such time as a robust system of this kind can be developed and made publicly available, some patent readers may consider it worthwhile to reformat or recolor a Claims section on their own screen or by hand.

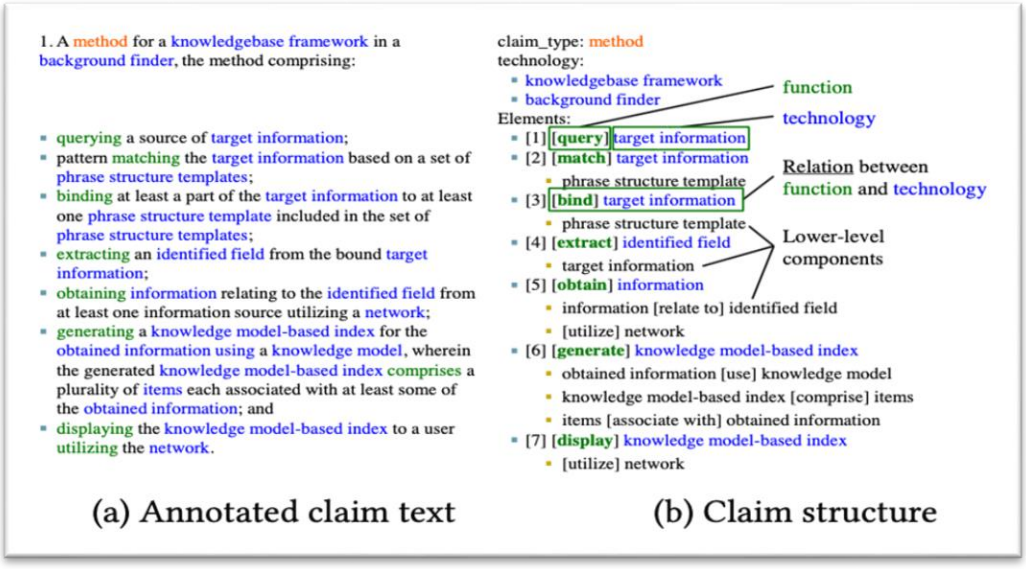


Figure 3. Reformatted and recolored claim section (Okamoto et al., 2017, p. 1). Reproduced with the authors' permission.

4.1.7. Review

The last step in the SQ5R sequence is that the reader should review their reading process and their comprehension by returning to their questions and reflecting. Were they the right questions? Have the questions been adequately answered? If so, the act of reading is complete. Otherwise, if necessary, the reader may reiterate the sequence by returning to the appropriate point in the sequence and repeating steps as necessary. (Artis, 2008)

4.2. Equipping the reader

This section presents various comprehension strategies that are not part of the act of reading. These might be considered auxiliary strategies. They support the reader by equipping them with the cognitive and situational tools to read as effectively as possible.

4.2.1. Addressing the affective dimension

Section 3.4.1. examined affect, or mental state, and how it can impair reading comprehension. Numerous writings on patent literacy and several participants in this study have underscored how negative modes of affect can have a significant impact on patent reading and proposed different measures for alleviating them.

The principal strategy is metacognitive: acknowledge the difficulties of reading and understanding complex texts like patents. “It is important to call attention to that difficulty” and acknowledge the risk of misinterpretation. (O’Toole, 2021, p. 5) "The fact that scholarly reading does take effort and can be frustrating needs to be acknowledged so students don’t feel like they are experiencing unique problems with the material leading them to imagine that they are stupid or unable to complete “real” scholarly work." (MacMillan & Rosenblatt, 2015a, Even More Reading Strategies We Wanted to Include in Our Paper, but …Word Limits.) “Allay students’ initial misgivings about the occasional lack of clarity in patent documents” (MacMillan, 2005, p. 152). Acknowledging these difficulties helps the reader to stay motivated and understand that they do not need to rely on their own knowledge or expertise. “I advise researchers that it is hard (or impossible) to become an instant expert.” (PTRC representative)

One specific recommendation regarding affect is rethinking reading not as “reading” at all and instead recasting it as “translating”. According to MacMillan and Rosenblatt, this relieves the reader of some of the weight of expectation that they should be able to comprehend the text immediately. Instead, the approach of “translating” communicates the need for a more active reading approach and serves to validate the reader’s efforts. (2015a, Addressing the Affective Aspects of Reading)

Similarly, several PTRC representatives recommended assisting patent readers by helping them to plan for research that can be difficult, frustrating, and time-consuming. This can mean understanding reading as an iterative process (as suggested in section 4.1.5.), embracing the need for time, budgeting for that time, and possibly needing to consult with experts. The technology transfer specialist endorsed this approach and shared that, when possible, they block off time in their calendar for patent reading and patent analysis. This earmarks time for these activities, helps them to plan ahead, and helps to create a record of time spent.

4.2.2. Learning the “architecture” of a patent

Patents have a unique and non-intuitive structure, or “architecture” or “anatomy”, which can confuse or overwhelm the reader as they attempt to make sense of a patent’s “semantic connections” (3.2.1.). Accordingly, patent readers need to “Get your bearings” (one of the key tips for reading patents offered by Donald et al., 2018, p. 278). It is

essential to understand the structure and layout of patent documents, including the purpose of each section and the nature of its contents. When asked what reading assistance they would provide to researchers, 15 of 21 PTRC representatives emphasized educating them about the structure and organization of patent documents. Several shared the view that patent documents can be compared in some respects to journal articles (which might be more familiar): both have a conventional structure that is not necessarily temporally or logically linear, but with specific information that can be retrieved from specific locations.

Patent readers thus need to know the meaning of cover page fields and the data they contain (e.g. classification classes and subclasses), and the main sections and their purposes. The graphic layout of fields and sections in official patent documents or on the webpages of patent databases is also important, for both comprehension and navigation. Several publications seek to present a comprehensive explanation of fields and sections, and their non-obvious characteristics. (e.g., Kim, 2015; lens.org, 2020; Meier, 2012) lens.org draws attention to the different scopes of the abstract and the Summary. It also calls on the reader to bear in mind that the Description section has a conventional organization: first, a broad description of the invention and how to use it; possibly, then "preferred embodiments" that are more limited than the broad description and may be a fall-back position of claims if the broader position is not patentable; and lastly, specific examples of how to practice it.

Understanding what each field and section is, what information each one contains, where each is located, and how each one relates to the others, makes it possible to read the patent in a non-linear sequence. (MacMillan & Rosenblatt, 2015a). Unlike some other publication formats that present essential information in the first paragraphs, the important information in patents is spread throughout the document. In the words of one PTRC representative, "I would outline the different sections to guide patrons to the information they are looking for, rather than reading through the entirety of each patent." However, the appropriate non-linear sequence for any given reader will vary according to their purpose for reading the patent. "Students and researchers, assuming that their use cases are research-based and not assessing patentability, are most often able to skip the legal pieces and focus on the technical/scientific content." (PTRC representative). Scientists are advised to "jump" to the abstract, examples, and claims. (Donald et al., 2018)

Non-linear reading supports efficiency. This was strongly emphasized by three of the five patent professionals. In patent attorney A's words: "the practical skill is not just reading, it's efficient reading." As in-house counsel for a corporation, the efficiency of their work and their completion of targets is subject to a high level of scrutiny. They explained the importance to their work of focusing only on the relevant sections of a patent for reasons of time - something likely to be a limiting factor for most patent readers.

Patent attorney A also made the point that non-linear reading is necessary because patent documents require careful reading and the engagement of a high level of intellectual capacity - reflecting the need to engage the lexical vocabulary, semantic understandings, and prior knowledge described above. Consequently, "to start at the beginning of a patent and expect to read it like… an article in a newspaper, this is disastrous because... the reader gets to the important part, then it's most likely that they have spent all of their attention budget on boilerplate [or] preamble information that is otherwise useless or not that informative. It's there for a legal reason to comply with the requirements of

the patent statutes, but it's not actually what the patent is about."

Instead, the purpose or question determines where in the document they begin their reading and the sequence in which they might move around the document to other sections. "The type of the approach to the patent is very much dependent on the question being answered... If it's an infringement question, the first step is the claims - which are at the end, right? …If you start reading the abstract at the beginning and you proceed through the background and you wade through the first few pages, you read through all the brief description of figures and then the 1st 50 paragraphs of the detailed description, you haven't done anything important yet… It's only the independent claims that matter... If the accused device is missing any one of those elements in the independent claim, you're home free, you've done the reading. You can be done in 5 minutes."

Alternatively, when responding to an office action, the starting point could be anywhere. The attorney needs to analyze the basis for the rejection, which can be any part of a reference. "It's not just the claims of the reference, it's the drawings. It's the detailed description that's the background. It's everything."

Another instructive point made by patent attorney A is that readers can use the drawings as a navigation tool for looking up information in a non-linear manner. "The drawings are a good place to start because the drawings are essentially the organizing principle for the application... Each individual drawing is sort of like a table of contents, and the numbers for the individual components within a drawing [are like] the index… The general practice is that the drawing should go from broad to narrow, so you look through the first couple… Each component will have its own reference numbers, so if I get down to… the subsystem I'm most interested in, then I can search that figure number in the detailed description and I go exactly to the paragraph that I'm interested in."

4.2.3. Building vocabulary

Section 3.1. illustrated the challenges presented by the uncommon and sometimes unique lexical units found in patent documents. As in other aspects of patent reading, a metacognitive approach is recommended regarding lexical processing. Patent readers need to keep in mind that one's own vocabulary will not match the vocabulary used in some patents and be alert to the need to check the definitions of words and phrases. This can mean checking by internal reference to a patent's *sui generis* definitions of terms, or by external reference to authoritative technical or legal definitions.

Patent readers also need to build up their vocabulary in the relevant technological areas, by acquiring vocabulary through different media and publications. PTRC representatives also advised familiarizing oneself with the lexical conventions of patents' legal language. For example, readers should know the significance of the different transitions used in claims and the different tenses used in examples (see 3.1.). One representative encouraged readers to build their familiarity with this language by reading more patents: "I always recommend reading more than 5 patents to compare the legal language".

4.2.4. Building knowledge

Building one's technical vocabulary usually goes hand in hand with building one's technical knowledge, which also supports comprehension. Section 3.3. considered how readers who are short on prior technical knowledge will struggle to "fill the gaps" in a patent or make connections between what they read and what they already know. This is especially true for anyone who lacks some of the "skill in the art" that a patent assumes on the part of a reader. To some degree, this can be remedied with more immersion and learning in the relevant technological area. "People with

more experience reading patents have an easier time. Those who have a greater knowledge of their field are slightly more equipped than the average person." (PTRC representative). For readers with minimal knowledge, this may mean starting with tertiary sources that summarize "state of the practice" information, such as textbooks and encyclopedias; or secondary sources such as review articles. For more advanced readers, it may mean diving deeper into primary research and - again - reading more patents.

4.3. The expertise of others

Patent readers can, and may need to, draw on the assistance of experts to support their reading.

PTRC representatives are expert patent educators and available to assist members of the public with a wide range of research services. For some representatives, this role overlaps with responsibilities for teaching patent literacy at the college/university level. However, survey responses showed a nuanced mix of perspectives on reading assistance.

For some PTRC representatives, the need has not arisen: "I don't ever have the occasion to read patents with library patrons." Several PTRC representatives stated that they had rarely received requests for reading assistance because researchers prefer to limit consultations to search techniques. After that "most people don't have the patience" and are eager to continue their research independently."

Several representatives commented that they do not and would not provide assistance with interpreting the meaning of a patent. This is intentional and a necessary precaution because patent educators may not give legal advice or offer information that might be (mis)understood to have legal bearing. (Irvin, 2018; Overhiser et al., 2022) The unauthorized practice of law is prohibited in most US states; in some, it is a criminal offense punishable by imprisonment and/or a fine. Accordingly, PTRC representatives stated "I do not go anywhere near this." "I do not provide interpretations of legal terms." "They're on their own, unfortunately."

Although patent educators cannot assist with the interpretation of individual patents in any situation that might have legal bearing, PTRC representatives in this study expressed empathy for the challenges of reading and may provide reading assistance on a generalized and non-specific level. Generalized reading assistance can take the form of introducing patent readers to the reading strategies presented in this paper. This can be combined with illustrative examples that model these strategies, perhaps in a technological field unrelated to the reader's area of interest. For example, as noted above (4.2.2.), most PTRCs offer detailed guidance on patent "architecture", using consultations and online information to educate researchers about organization and layout.

Patent readers may need to consult with other sources of expertise. "[Reading patents is] extremely difficult. I tell people not to rely on their own knowledge, but to spend time looking up explanations or seeking expert help." (PTRC representative) This might mean consulting with technical or legal specialists. Legal assistance is unlikely to be appropriate or feasible for exploratory research, but for patentability or infringement research, legal assistance is recommended. "The legal language… can be hard to [interpret]... and that's why we recommend having an attorney." "Faculty at my university I will direct to the university patent lawyer, or tech transfer office." (PTRC representatives)

Authoritative explanatory information can also mean reference publications such as dictionaries, encyclopedias, or other published information that readers can use to look up terminology at the point of need. Educators of all kinds who are supporting reading comprehension are well advised to "provide links to resources that can

help explain the methods and terminology" (MacMillan & Rosenblatt, 2015a, Decoding the Technical Elements of Texts). In the case of patents, this addresses the fact that many readers may need to acquire "lexical units" for comprehension. The patent examiner commented that patent vocabulary generally does not present difficulties for comprehension to them, but "if a new term or concept is being used in a patent that I haven't encountered before, then I may research that term or concept."

In Suominen et al.'s user study (2017), both experienced patent readers and "lay" persons - though especially the "lay" persons - highlighted the value of having access to definitions. Several suggested that, for convenience and efficiency, patents would ideally have definitions embedded within the document. In the absence of embedded definitions, access to technical, scholarly and legal reference resources is important. Rutwik and Avantika's non-reviewed investigation (2023) concluded that using a large language model (LLM)-based chatbot for explanations of key terms in patents was "impressive", though they also found that asking it for synonyms of selected terms returned responses that were too general to be helpful.

4.4. Accessible language

An alternative - or complement - to obtaining explanations or interpretations of patent language can be converting patents into language that is more accessible for a wider range of readers. IN Suominen et al.'s study (2017), "The laypeople strongly agreed that patents should and could be made more readable." (6. Discussion) In this area, AI-based technologies may offer new solutions.

4.4.1. Translation

When reading a patent issued in a language in which the reader does not have fluency, it is valuable to have a good-quality translation. Machine translations can contain awkward phrasings or odd word choices that confuse the text's meaning. Patent attorney A expressed appreciation for having access to a patents database that provides higher-quality machine translations of non-English-language patents. They were critical of Google Patents' machine translation, though they thought this had improved in recent years. They liked the quality of machine translations in Patsnap, though this is a proprietary product that may be beyond the budgets of many libraries, organizations, and businesses.

4.4.2. Simplification

A second form of text conversion is simplification: replacing original language with simpler language, without loss of informational content. This includes paraphrasing, removing jargon, rephrasing technical terms, and shortening sentences. When Suominen et al. (2017) asked researchers for their thoughts on a range of possible features that might make it easier to understand patent information, both expert and lay readers rated "non-binding clarifications" the most highly. This referred to plain-English translations or summaries of the entire patent document, while avoiding changes that might affect the legal scope of the claims.

In the preface to a review of patent language-processing technologies, Jiang and Goetz (2024) agree that rewriting patent documents for improved readability has the potential to enhance comprehension, accessibility to non-expert readers, efficiency, the dissemination of information, and, ultimately, innovation and the protection of IP rights - while noting that the costs in expertise, time and money make it prohibitive to create a corpus of patent texts written by humans. Instead, substantial progress is being made in the field of automated language processing, including text simplification, due to artificial intelligence-based technologies. "At first sight, recent large models for generative tasks and language processing appear to be a perfect match for the patent literature" (Jiang & Goetz, 2024, p.

3) due to their precise language, consistent document structure, and availability to the public.

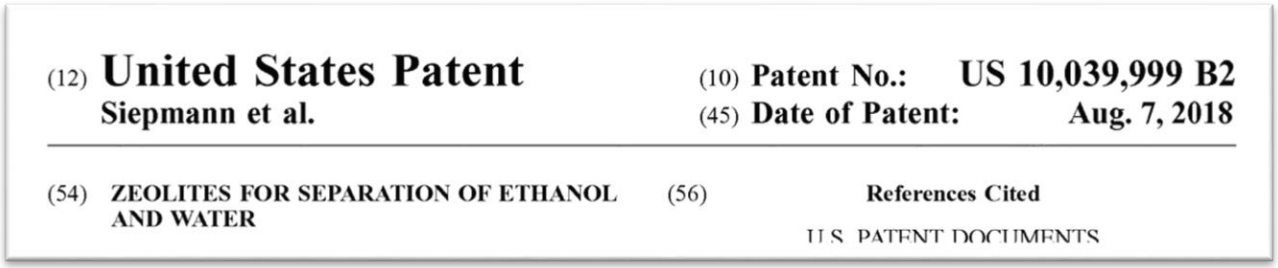
According to Casola et al., (2023) most work on automated simplification has focused on the Claims section, but "since the legal scope needs to remain unchanged, modifying the text presentation is preferred to rephrasing" (p. 1045). In other words, there is a focus on reformatting the Claims section in ways similar to the approach developed by Okamoto et al. (see Figure 3) and others. There thus appears to be more scope for applying sentence simplification techniques to the Description section. This could be done at the lexical level, by translating unusual words into simpler ones or adding explanations for advanced concepts. The authors also propose a method for simplification at the syntactic level, replacing long and complex sentences with shorter ones with simplified structures. This entails using a paraphrasing system trained on "general-domain" text such as the English and Simple English Wikipedia language corpora. In their example, "In interaction with the component secured on the cylinder head, the radial shoulder of the pressure medium distributor now prevents the camshaft from migrating axially further into the cylinder head." becomes "The radial shoulder of the pressure medium distributor prevents the camshaft from moving further into the cylinder head when interacting with the component secured on the cylinder head." (p. 1047)

Measured by sentence length, simplification is possible. "The process is, however, hard to control and might be error-prone." (p. 1046) Possible errors include the loss of elements, the removal of important adjectives, the reversal of relationships between elements, and hallucination of sentences.

4.4.3. Summarization

The "non-binding clarifications" liked by the researchers in Suominen et al.'s user study (2017) included both simplified versions and summaries. Unlike simplified versions of entire documents, summaries can be cost-effective for production by humans.

Patent database Derwent Innovations Index (DII), a product licensed by science analytics and IP company Clarivate, has a team of subject matter experts that adds "enhanced" titles and abstracts to each patent record. (Clarivate, 2021) Each title identifies the nature of the invention and its novelty. For example, DII provides US10039999B2 "Zeolites for separation of ethanol and water" with the expanded title "Method of separating ethanol from mixture including ethanol and water involves contacting mixture with sorbent or membrane including zeolites". (See Figures 4 and 5.) The custom abstracts describe the invention's novelty, its uses, and its advantages over prior art, in a structure that is consistent across the database. The "enhanced" title and abstract use language that tends to be more self-explanatory and, again, more consistent across different patents. Researchers still need to have some understanding of the relevant technical area, but the additional information can be beneficial for both comprehension and searching. One PTRC representative commented that DII's accessible language and consistent structure are a "huge help to my academic users, in terms of triaging which patents to dig more deeply into". Indeed, Clarivate markets DII as "designed for use by the patent non experts" (Clarivate, 2021). Casola and Lavelli (2022) also highlight DII as an example of non-automated summarization.



(12) **United States Patent**
Siepmann et al.

(10) **Patent No.: US 10,039,999 B2**
(45) **Date of Patent: Aug. 7, 2018**

(54) **ZEOLITES FOR SEPARATION OF ETHANOL AND WATER**

(56) **References Cited**

U.S. PATENT DOCUMENTS

Figure 4. USPTO document for patent US10039999B2. Public domain.

As in the case of simplification, AI is making possible summarization technologies and these are in a state of rapid development. Some PTRC representatives, as well as the engineer, expressed interest in the implications for comprehension: “Requesting 'explain to me like I am stupid' of a generative AI/LLM, like ChatGPT, is one of their best uses when it comes to patents. People like to hear that they can ask ChatGPT, etc., to explain blocks of text copied from a patent.” (PTRC representative)

At the same time, AI-based language processing is not yet a mature technology and there are reasonable concerns about the quality and validity of its outputs. Unvalidated assessments by two patent law firms indicate that ChatGPT can generate a broad summary by paraphrasing text, but it may omit essential features and tends to omit any summarization of novelty. [(Rutwik & Avantika, 2023; Sharma, n.d.)]

Method of separating ethanol from mixture including ethanol and water involves contacting mixture with sorbent or membrane including zeolites

Jump to ↓ Patent family ↓ Chemical Information

Inventors: SIEPMANN J I; BAI P; TSAPATSIS M

Patent Assignees: SIEPMANN J I(SIEP-Individual)
BAI P(BAIP-Individual)
TSAPATSIS M(TSAP-Individual)

Derwent Primary Accession Number: 2016-07748F

Indexed: 2023-08-10

Abstract: NOVELTY - Method of separating ethanol from mixture including ethanol and water involves contacting mixture with sorbent or membrane including greater than or equal to 1 zeolites selected from FER-type zeolite, MRE-type zeolite, MTT-type zeolite, ATN-type zeolite, ZON-type zeolite, OWE-type zeolite, ESV-type zeolite, CGF-type zeolite, MAZ-type zeolite, CDO-type zeolite, STI-type zeolite, AFO-type zeolite, AFI-type zeolite, ATO-type zeolite, and/or EZT-type zeolite in which greater than or equal to 1 fraction

Expand to show full abstract

Figure 5. DII record for patent US10039999B2. Reproduced with the permission of Clarivate.

As explained by Jiang and Goetz, summarization can be "extractive" or "abstractive", and both methods have weaknesses. "Extractive methods refer to selecting and extracting key phrases or sentences directly from the text of the patent. The goal is to retain the most significant and representative parts of the original document without altering the text." (36) This method is simpler to implement and it preserves more original wording and meaning; but original wording may perpetuate the original difficulties for comprehension. Also, extractive summaries tend to have less coherence and may have confusing syntax. Abstractive summaries are rephrasing of the patent's language. These should be more concise and more coherent, but the processing method requires "deeper understanding of the text" and risks distorting facts or meanings.

4.4.4. A note on the state of the art in generative AI

While some PTRC representatives and some patent professionals, especially the engineer, were positive about the potential for using AI-based technologies to aid patent reading, others expressed concerns. For example, "I worry that's a niche that AI will try to fill." (PTRC representative) As seen, AI-based language processing is advancing rapidly in a range of applications that might be useful to patent reading. However, while "At first sight, recent large models for generative tasks and language processing appear to be a perfect match for the patent literature" due to their precise language, consistent document structure, and availability to the public, "the field of machine learning on patents and patent-related aspects is still underdeveloped and not highly prominent yet." (Jiang & Goetz, 2024, p. 3)

There are multiple obstacles to the development of AI-based text generators for patents - including some that are also obstacles to human comprehension. These include lexical complexity, neologisms, and the vagueness of some wording. Developing abstractive methods that retain the precision and accuracy of the original language is also a complex task.

Further improvement of these technologies should be anticipated, not least if generative LLM can be trained on an IP-focused language corpus, including well-structured patent data and target summaries in "plain English". (Casola & Lavelli, 2022; Jiang & Goetz, 2024) To date, almost all testing of patent language processing has been based on regular, non-IP language corpuses. One exception, Bai et al's "PatentGPT" prototype, suggests that an IP-trained LLM chatbot can return superior results on tasks such as summarization, simplification, and explanation. (Bai et al., 2024) It would also be beneficial to train a patent-oriented LLM on relevant technical and legal literature and documentation that enable the LLM to contextualize patent tasks. This would support abstractive methods, similar to how "rich and organized" prior knowledge helps humans to fill in the gaps in patent information. To this end, Bai et al. trained PatentGPT on not just patents (47.7%), but also academic papers (22.9%), web pages (10.8%), file wrappers (6.4%), news (4.6%), books (3.0%), Wikipedia (2.3%), litigation (1.2%), and miscellaneous other sources.

5. Conclusion

Patents are difficult to read and notoriously so. This survey has sought to create a map of the wide range of difficulties that patent readers typically encounter. It illustrates how many of these difficulties are inherent in the textual characteristics of patents, or technical documents more generally. In all of the three cognitive processes of reading - decoding lexical units, determining semantic connections, and creating knowledge structures - the text of patent documents presents major obstacles. They are numerous, though some of the most impactful are the complexity of words and phrases (which can

reflect either a high degree of precision or a high degree of vagueness); the unique organization and purpose, or "architecture", of patents; and the fact that many readers of patents do not have the "skill in the art" or prior knowledge attributed to a patent's hypothetical audience.

In addition, the typical patent reader is likely to be experiencing common situational circumstances that can compound the difficulties in these cognitive aspects. These include demotivation, awareness of time constraints, and sub-optimal language or reading competencies.

This study collected data from PTRC representatives who work with many types of patent researchers, technical professionals, and legal professionals. Their observations indicate that most of these obstacles to successful reading difficulties are experienced across all types of patent readers, regardless of their experience or purpose.

Fortunately, there are strategies for reading comprehension that readers of patents can apply. Many of these have emerged from extensive work on reading techniques and they have been shown to have a positive impact on comprehension. Indeed, the data in this study suggest that while all types of patent reader experience cognitive and situational difficulties, some read more effectively or efficiently because they have acquired appropriate comprehension strategies.

This survey has sought to compile strategies for patent comprehension and trace how they correspond to the difficulties of patent reading. Each element in the SQ5R method supports reading comprehension, though some have particular relevance to the idiosyncrasies of patent documents, such as surveying and visualizing. "Equipping" strategies are similarly important: for example, acknowledging and analyzing the frustrations of reading patents can be a pathway to mitigating negative affect; while learning the architecture of patents can reduce disorientation and facilitate non-linear reading. Meanwhile, AI-based technologies are creating new possibilities for converting patent texts into more-accessible language, though the reliability of machine-generated text will need to be carefully evaluated. Taken as a whole, the synthesis of reading strategies presented in this paper may be a basis for best practices for reading patents.

Reading strategies have the potential to enable and enrich the work of patent researchers. This study emerged from a recognition that college students often lack the necessary reading skills to make sense of patent information and consequently fail to integrate that information into their design projects. As this paper has discussed (3.4.2.), reading comprehension is foundational to other information literacy competencies. For this reason, William Gray, the pioneer of scholarship on readability, promoted the slogan that "every teacher is a teacher of reading." (Moore et al., 1983, p. 424) It can be argued that reading comprehension requires more attention and more-purposeful instruction and assistance than it typically receives.

With regards to patents, the stakes for reading comprehension include not just students' academic and professional success, but the advancement of society as a whole. Successful reading comprehension leads to the circulation of patent information and the technological concepts that it embodies. In this light, reading is a linchpin in the dissemination of innovation and promotion of "the Progress of Science and useful Arts". (U.S. Const. art. 1, § 8.)

References

Artis, A. B. (2008). Improving marketing students' reading comprehension with the SQ3R method. *Journal of Marketing Education*, *30*(2), 130–137. https://doi.org/10.1177/0273475308318070

Asche, G. (2017). "80% of technical information found only in patents" – Is there proof of this [1]? *World Patent Information*, *48*, 16–28. https://doi.org/10.1016/j.wpi.2016.11.004

Atherton, M., Jiang, P., Harrison, D., & Malizia, A. (2018). Design for invention: Annotation of functional geometry interaction for representing novel working principles. *Research in Engineering Design*, *29*(2), 245–262. https://doi.org/10.1007/s00163-017-0267-2

Bai, Z., Zhang, R., Chen, L., Cai, Q., Zhong, Y., Wang, C., Fang, Y., Fang, J., Sun, J., Wang, W., Zhou, L., Hua, H., Qiu, T., Wang, C., Sun, C., Lu, J., Wang, Y., Xia, Y., Hu, M., … Tu, C. (2024). *PatentGPT: A large language model for intellectual property* (arXiv:2404.18255). arXiv. https://doi.org/10.48550/arXiv.2404.18255

Best, R. M., Rowe, M., Ozuru, Y., & McNamara, D. S. (2005). Deep-level comprehension of science texts: The role of the reader and the text. *Topics in Language Disorders*, *25*(1), 65. https://journals.lww.com/topicsinlanguagedisorders/abstract/2005/01000/deep_level_comprehension_of_science_texts__the.7.aspx

Carlson, L. E., & Sullivan, J. F. (2004). Exploiting design to inspire interest in engineering across the K-16 engineering curriculum. *International Journal of Engineering Education*, *20*(3), 372–378. https://www.ijee.ie/articles/Vol20-3/IJEE2502.pdf

Casola, S., & Lavelli, A. (2022). Summarization, simplification, and generation: The case of patents. *Expert Systems with Applications*, *205*, 117627. https://doi.org/10.1016/j.eswa.2022.117627

Casola, S., Lavelli, A., & Saggion, H. (2023). Creating a silver standard for patent simplification. *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1045–1055. https://doi.org/10.1145/3539618.3591657

Clarivate. (2021). *Discover Derwent Innovations Index for easy search and discovery of patent content: User Guide*. Clarivate. https://clarivate.com/webofsciencegroup/wp-content/uploads/sites/2/dlm_uploads/2021/12/Derwent-Innovations-Index_User-Guide-Dec-2021.pdf

Donald, K. E., Kabir, K. M. M., & Donald, W. A. (2018). Tips for reading patents: A concise introduction for scientists. *Expert Opinion on Therapeutic Patents*, *28*(4), 277–280. https://doi.org/10.1080/13543776.2018.1438409

Follmer, D. J., Fang, S.-Y., Clariana, R. B., Meyer, B. J. F., & Li, P. (2018). What predicts adult readers' understanding of STEM texts? *Reading and Writing*, *31*(1), 185–214. https://doi.org/10.1007/s11145-017-9781-x

Gray, W. S. & Bernice E. Leary. (1935). *What makes a book readable, with special reference to adults of limited reading ability; an initial study*. The University of Chicago Press. https://hdl.handle.net/2027/mdp.39015031032009

Howland, H. P., Jarvie, L. L., & Smith, L. F. (1943). *How to read in science and technology*. Harper.

Irvin, D. (2018). The Patent Office in the library. *Public Services Quarterly*, *14*(4), 392–398. https://doi.org/10.1080/15228959.2018.1487360

Jiang, L., & Goetz, S. (2024). *Artificial intelligence exploring the patent field* (arXiv:2403.04105). arXiv. https://doi.org/10.48550/arXiv.2403.04105

Kim, D. G. (2015). *A rhetorical approach to patents: A guide to reading a patent* [M.A., San Diego State University]. https://www.proquest.com/docview/1708672804/

Koh, E. C. Y. (2020). Read the full patent or just the claims?: Mitigating design fixation and design distraction when reviewing patent documents. *Design Studies*, *68*, 34–57. https://doi.org/10.1016/j.destud.2020.02.001

Lei, S. A., Rhinehart, P. J., Howard, H. A., & Cho, J. K. (2010). Strategies for improving reading comprehension among college students. *Reading Improvement*, *47*(1), 30–42. https://www.ingentaconnect.com/content/prin/rimp

lens.org. (2020, November 11). *How to read a patent*. Lens.Org. https://support.lens.org/knowledge-base/how-to-read-a-patent/

Lu, P., Schroeder, S., Burris, S., Rayfield, J., & Baker, M. (2022). The effectiveness of a metacognitive strategy during the reading process on cognitive allocation and subject matter retention. *Journal of Agricultural Education*, *63*(2), 201–218. https://doi.org/10.5032/jae.2022.02201

MacMillan, D. (2005). Patently obvious: The place for patents in information literacy in the sciences. *Research Strategies*, *20*(3), 149–161. https://doi.org/10.1016/j.resstr.2006.06.004

MacMillan, M., & Rosenblatt, S. (2015a). They've found it. Can they read it? Adding academic reading strategies to your IL toolkit. *ACRL 2015*. ACRL 2015, Portland, OR. https://alair.ala.org/bitstream/handle/11213/17916/TheyveFoundIt.pdf

MacMillan, M., & Rosenblatt, S. (2015b, January 15). Challenges & Strategies. *Adding Reading Strategies to Your IL Toolkit*. https://readingstratsacrl2015.wordpress.com/challenges-of-academic-reading-2/

Meara, J. (2002). Just who is the person having ordinary skill in the art? Patent law's mysterious personage. *Washington Law Review*, *77*(1), 267. https://digitalcommons.law.uw.edu/wlr/vol77/iss1/7/

Meier, J. J. (2012). Intellectual property: Patents: Important information literacy standards for interdisciplinary science courses. In J. R. Davidson & K. O'Clair (Eds.), *The Busy Librarian's Guide to Information Literacy in Science and Engineering* (pp. 93–104). American Library Association. https://alastore.ala.org/content/busy-librarians-guide-information-literacy-science-and-engineering

Mille, S., & Wanner, L. (2008, May 28). Making text resources accessible to the reader: The case of patent claims. *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*. International Conference on Language Resources and Evaluation, Marrakech, Morocco. http://www.lrec-conf.org/proceedings/lrec2008/pdf/352_paper.pdf

Moore, D. W., Readence, J. E., & Rickelman, R. J. (1983). An historical exploration of content area reading instruction. *Reading Research Quarterly*, *18*(4), 419–438. https://doi.org/10.2307/747377

Okamoto, M., Shan, Z., & Orihara, R. (2017). Applying information extraction for patent structure analysis. *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 989–992. https://doi.org/10.1145/3077136.3080698

O'Toole, J. (2021). Questioning authority: Patents and source evaluation in an era of misinformation. *Journal of the Patent and Trademark Resource Center Association*, *31*(1). https://tigerprints.clemson.edu/jptrca/vol31/iss1/3

Ouellette, L. L. (2017). Who reads patents? *Nature Biotechnology*, *35*(5), 421–424. https://doi.org/10.1038/nbt.3864

Overhiser, S., Barrett, S., Miller, J., Mittelbach, S., Raider, R., Ptrc, D., Reinman, S., Murphey, M., & Ptrc, O. (2022). *Guide for best practices for Patent and Trademark Resource Centers (PTRCs)*. PTRC Association. https://ptrca.org/files/Guide%20for%20Best%20Practices%20for%20Patent%20and%20Trademark%20Resource%20Centers%20(PTRC)%20August%202022.pdf

Patents Act, 35 U.S.C. § § 112 (1952). https://uscode.house.gov/view.xhtml?req=granuleid:USC-prelim-title35-section112&num=0&edition=prelim

Pedraza-Fariña, L., & Whalen, R. (2022). The ghost in the patent system: An empirical study of patent law's elusive "skilled artisan." *Iowa Law Review*, *108*(1), 247–302. https://ilr.law.uiowa.edu/print/volume-108-issue-1/the-ghost-in-the-patent-system-an-empirical-study-of-patent-laws-elusive-skilled-artisan

Robinson, F. P. (1941). *Effective study*. Harper & Brothers. http://archive.org/details/in.ernet.dli.2015.224377

Robson, H. (2001). How to read a patent. In Harry Robson & K. P. Lillerud (Eds.), *Verified syntheses of zeolitic materials* (p. 73). Elsevier Science. DOI: 10.1016/B978-044450703-7/50114-9

Rutwik & Avantika. (2023, October 27). Exploring ChatGPT's capabilities & limitations in prior art search. *GreyB*. https://www.greyb.com/blog/chatgpt-for-

prior-art-search/

Schox, J. (2015). *Not so obvious: An introduction to patent law and strategy* (4th edition). CreateSpace Independent Publishing Platform.

Segal, D. (2019). *One hundred patents that shaped the modern world*. Oxford University Press.

Sharma, R. (n.d.). Putting ChatGPT to the patent analysis test [UnitedLex]. *Insights*. https://unitedlex.com/insights/putting-chatgpt-to-the-patent-analysis-test/

Sherriff, G., & Rand, D. (2022). Prior art research in the capstone design experience: A case study of redesigned online and in-person instruction. *Issues in Science and Technology Librarianship*, *100*. https://doi.org/10.29173/istl2598

Strohmaier, A. R., Ehmke, T., Härtig, H., & Leiss, D. (2023). On the role of linguistic features for comprehension and learning from STEM texts. A meta-analysis. *Educational Research Review*, *39*, 100533. https://doi.org/10.1016/j.edurev.2023.100533

Suominen, H., Ferraro, G., Nualart, J., & Hanlen, L. (2017). User study for measuring linguistic complexity and its reduction by technology on a patent website. *Proceedings of the 34th International Conference on Machine Learning*. https://openresearch-repository.anu.edu.au/handle/1885/250650

U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. (n.d.). *Highlights of the 2017 U.S. PIAAC Results Web Report* (NCES 2020-777). https://nces.ed.gov/surveys/piaac/current_results.asp

US Patent and Trademark Office. (2022). *Manual of Patent Examining Procedure (9th ed.)*. US Patent and Trademark Office. https://www.uspto.gov/web/offices/pac/mpep/index.html

US Patent and Trademark Office. (n.d.). *Patent and Trademark Resource Centers* [Text]. https://www.uspto.gov/learning-and-resources/patent-trademark-resource-centers

Verberne, S., D'hondt, E. K. L., Oostdijk, N. H. J., & Koster, C. H. A. (2010, March 28). Quantifying the challenges in parsing patent claims. *21*. 1st International Workshop on Advances in Patent Information Retrieval (AsPIRe-2010), Milton Keynes, UK. https://repository.ubn.ru.nl/handle/2066/84168

Wilson-Lopez, A., Gregory, S., & Larsen, V. (2017). Reading and engineering: Elementary students' co-application of comprehension strategies and engineering design processes. *Journal of Pre-College Engineering Education Research (J-PEER)*, *6*(2). https://doi.org/10.7771/2157-9288.1116

World Intellectual Property Organization. (n.d.). *R&D, Innovation and Patents*. https://www.wipo.int/patent-law/en/developments/research.html

Zwicky, D. (2016). Balancing conflicting service priorities as an academic patent librarian. *Public Services Quarterly*, *12*(1), 77–85. https://doi.org/10.1080/15228959.2015.1128372

Zwicky, D. (2019). Thoughts on patents and information literacy. *Journal of the Patent & Trademark Resource Center Association*. https://docs.lib.purdue.edu/lib_fsdocs/216

7. Appendix A: Survey of patent reading assistance

Background information

1. What is the name of your library?
2. What type of library hosts your PTRC?
    - academic
    - public
    - special
    - state
    - other
3. What type of patent researchers do you work with?
    - academic researchers
    - business employees or contractors
    - entrepreneurs or small business owners
    - inventors
    - patent agents or patent attorneys
    - other

Reading difficulties

4. What reading difficulties specific to the cover page/patent data/bibliographic information have you observed?
5. What reading difficulties relating to technical or scientific content have you observed?
6. What reading difficulties specific to legal language or information have you observed?
7. Have you observed differences in the reading needs or experiences of different types of patent researcher (e.g. inventors, engineers, students, etc.)?

Assistance for reading a patent document

The following questions pertain to both person-to-person assistance and off-the-shelf guides, e.g., webpages, paper handouts, etc.

8. What assistance do/would you provide regarding how to read a patent for comprehension?
9. What assistance do/would you provide regarding how to read a patent to evaluate its relevance to the reader's interest?
10. What assistance do/would you provide regarding how to read a patent to evaluate its legal significance or legal implications?
11. Is there any other assistance or advice that you provide that might aid a patent reader?

Further information

12. Is there anything else you would like to share about patent researchers' needs and experiences regarding reading patent documents?

8. Appendix B: Interviews with patent professionals

Preliminary questions:

1. Please state your job title and your place of work.
2. How many years have you worked as a practicing engineer or patent professional?
3. Do you self-identify as a member of any underrepresented groups or identities?

Working with patents:

4. To get us started, please tell me about your work: your main responsibilities, activities, and areas of expertise?
   Please describe the role that patents have in your work. In what situations do you work with patents?
5. Before working as a professional engineer, what was your knowledge or familiarity with patents? Did you learn about patents and patent literature in your education or professional development?
6. Besides hands-on work with patents, have you done anything during your professional career to develop your knowledge of patents? For example, any training provided by an employer or any self-directed learning?

Reading patents:

7. Do you typically have a purpose in mind when reading a patent? What do you aim to accomplish?
8. Do you have any strategies for reading a patent document, making sense of it, or assessing its usefulness for your work? What works well for you?
9. The "SQ5R" model of reading combines several different techniques that can aid the reading and comprehension of a text. I'd like to ask whether you use any of these techniques when reading a patent. (You may have already mentioned some of them.)
   - Surveying: skimming for an overview of main ideas, drawings, length, predicting time, whether to close read, etc. (Follow-up: How do you decide whether a patent might be worth reading closely?)
   - Questioning: identifying questions to be answered
   - Reading: reading, eg. annotation or highlighting; monitoring comprehension
   - Recording: making notes, summarizing
   - Responding: considering how the document answers the questions
   - Reciting: eg. reporting or discussing with colleagues
   - Reviewing: returning to the document or to notes to check comprehension and plan next steps

10. To what extent do you experience any difficulty understanding any of the following: vocabulary, syntax, technical information or scientific principles, gaps in the technical specifications, legal implications?
11. When you experience difficulty understanding [area of difficulty], how do you resolve this? For example, what would you do if you encountered unfamiliar vocabulary?

Anything else:

12. Is there anything else you would like to share about your work with patents or how you read them?

9. Acknowledgements

The author thanks every patent educator and patent professional who contributed their time and expertise to participating in this study; Dustin Rand for assisting with the recruitment of interview participants; and Suzanne Reinman and Jim Miller for providing feedback on a draft version of this paper.

10. Biographical info

Graham Sherriff (he/him) is an instructional librarian at the University of Vermont with responsibility for services supporting the College of Engineering and Mathematical Sciences, and representative for the Burlington, VT, PTRC.
ORCID: 0000-0003-1454-0705