2018

Patent Clutter

Janet Freilich
Fordham University School of Law, jfreilich1@fordham.edu

Patent Clutter

Janet Freilich[*]

ABSTRACT: Patent claims are supposed to clearly and succinctly describe the patented invention, and only the patented invention. This Article hypothesizes that a substantial amount of language in patent claims is in fact not about the core invention, which may contribute to well-documented problems with patent claims. I analyze the claims of 40,000 patents and applications, and document the proliferation of "clutter"—language in patent claims that is not about the invention. Although claims are supposed to be exclusively about the invention, clutter appears across industries and makes up approximately 25% of claim language. Patent clutter may contribute several major problems in patent law. Extensive clutter makes patent claims harder to search. Excessive language in patent claims may be the result of over-claiming—when patentees describe potential corollaries they do not possess—thereby making the patent so broad in scope as to be invalid. More generally, it strains the comprehensibility of patents and burdens the resources of patent examiners. After arguing that patent clutter may contribute to these various problems, this Article turns to reforms. Rejections based on prolix, lack of enablement, and lack of written description can be crafted to dispose of the worst offenders, and better algorithms and different litigation rules can allow the patent system to adapt (and even benefit from) the remaining uses of excess language. The Article additionally generates important theoretical insights. Claims are often thought of as entirely synonymous with the invention and all elements of the claim are thought to relate equally strongly to the invention. This Article suggests empirically that these assumptions do not hold in practice, and offers a framework for

[*] Associate Professor, Fordham Law School. I thank Yonathan Arbel, Oren Bar-Gill, Colleen Chien, Nestor Davidson, Jesse Frumkin, John Goldberg, John Golden, Jennifer Gordon, Clare Huntington, Joseph Kupferman, Ethan Leib, Joshua Lerner, Oskar Liivak, Alan Marco, Michael Meurer, Kristen Osenga, Mark Patterson, Sarah Pihonak, Nicholson Price, Duane Rudolph, Rachel Sachs, Joshua Sarnoff, Steven Shavell, Holger Spamann, Henry Smith, Haris Tabakovic, Tess Wilkinson-Ryan, Thomas Wollmann, Steven Worthington, and Dennis Yao. This Article has also greatly benefited from feedback from patent examiners and members of the Office of the Chief Economist at the United States Patent and Trademark Office, and from participants at the New England Junior Scholars Workshop, PatCon 6, the 8th Annual Junior Scholars in Intellectual Property Workshop, the 2016 Intellectual Property Scholars Conference, the 2017 Tri-State Intellectual Property Conference, and workshops at St. John's Law School, Harvard Law School, and George Washington Law School.

restructuring conceptions of the relationship between claims and the invention.

I. INTRODUCTION

Perhaps the most famous catchphrase in patent law is "the name of the game is the claim."[1] If claims are the name of the patent game, then patent law has a fundamental problem: "it isn't working."[2] Claims are so "notoriously difficult to understand"[3] that their meaning "is hotly debated in virtually every patent case."[4] Claims are criticized as vague, unreadable, excessively long, impossible to search, and dreadful to interpret.[5] These concerns are longstanding. For example, in 1916, Judge Learned Hand expressively remarked that claims can be "such a waste of abstract verbiage . . . . It takes the scholastic ingenuity of a St. Thomas with the patience of a yogi to decipher their meaning."[6] Claim dysfunctionality has generated a copious amount of literature[7] in addition to policy proposals and changes at the highest level.[8] In recent years, the White House,[9] the Federal Trade Commission,[10] the Patent and Trademark Office,[11] and the Supreme Court[12] have all begun seeking

1. Apple Inc. v. Motorola, Inc., 757 F.3d 1286, 1298 (Fed. Cir. 2014) (citing *In re* Hiniker Co., 150 F.3d 1362, 1369 (Fed. Cir. 1998)).

2. Dan L. Burk & Mark A. Lemley, *Fence Posts or Sign Posts? Rethinking Patent Claim Construction*, 157 U. PA. L. REV. 1743, 1744 (2009).

3. Kristen Osenga, *The Shape of Things to Come: What We Can Learn from Patent Claim Length*, 28 SANTA CLARA COMPUTER & HIGH TECH. L.J. 617, 620 (2011).

4. Mark A. Lemley & Carl Shapiro, *Probabilistic Patents*, 19 J. ECON. PERSP. 75, 85 (2005).

5. *See* JAMES BESSEN & MICHAEL J. MEURER, PATENT FAILURE: HOW JUDGES, BUREAUCRATS, AND LAWYERS PUT INNOVATORS AT RISK 10–11 (2008).

6. Victor Talking Mach. Co. v. Thomas A. Edison, Inc., 229 F. 999, 1001 (2d Cir. 1916).

7. *See infra* Part III.C.

8. *See infra* Part III.C.

9. Press Release, The White House: President Barack Obama, Fact Sheet: White House Task Force on High-Tech Patent Issues (June 4, 2013), https://www.whitehouse.gov/the-press-office/2013/06/04/fact-sheet-white-house-task-force-high-tech-patent-issues.

10. FED. TRADE COMM'N, THE EVOLVING IP MARKETPLACE: THE OPERATION OF IP MARKETS 116–17 (2009) [hereinafter EVOLVING IP MARKETPLACE].

11. *Glossary Initiative*, U.S. PAT. & TRADEMARK OFF., http://www.uspto.gov/patent/initiatives/glossary-initiative#heading-2 (last modified Apr. 3, 2016, 8:59 PM).

12. Nautilus, Inc. v. Biosig Instruments, Inc., 134 S. Ct. 2120, 2129–30 (2014).

improvements to the clarity and quality of patent claims. Yet the problems with claims remain acute.

This Article provides an empirical analysis of the claims of 40,000 patents and applications. The study demonstrates that patent claims are cluttered with vast amounts of language that have little to do with the invention. By law, claims are supposed to be succinct, one-sentence descriptions of the invention,[13] but in practice they are often not. Instead, claims can swell into pages-long (but still one-sentence) descriptions of the invention and a whole host of other topics. Claims are thought to be hard to read because they describe complex technologies,[14] but the problem may also be that claims are performing functions *other* than describing complex technologies. Because no one has documented the cluttered state of patent claims, clutter has not been acknowledged as a contributor to the array of problems associated with claims.

In this study, I measure claim language that is "ancillary"—about something other than the invention. Ancillary language is identified by searching for terms that appear in the claims, but never or rarely in the specification.[15] Surprisingly, although claims are supposed to be exclusively about the invention, approximately 25% of patent claim language is not about the patent's core invention.[16] The practice of including ancillary language in patent claims is roughly analogous to "keyword stuffing"—a technique of optimizing a website's Google ranking by packing the website with popular but irrelevant words to draw in searchers.[17] Ancillary language similarly obscures the invention actually protected by the patent and taxes the ability of judges, examiners, competitors, and the public to read and understand patent claims.

For example, U.S. Patent No. 6,450,942 describes a machine that can create electrical pulses according to particular rhythms, stimulating the human vascular system.[18] The patent then proceeds to claim use of this machine in many different contexts, such as "for cosmic medicine, including

13. U.S. PAT. & TRADEMARK OFF., U.S. DEP'T OF COMMERCE, MANUAL OF PATENT EXAMINING PROCEDURE § 608.01(m) (9th ed. 2015) [hereinafter "MPEP"].

14. For example, it is common to hold "technology tutorials" for courts before they conduct *Markman* hearings to construe the meaning of claim terms. Peter S. Menell et al., *Patent Claim Construction: A Modern Synthesis and Structured Framework*, 25 BERKELEY TECH. L.J. 711, 793 (2010).

15. This proxy is explained further in Part III.A. If a term in the claims is not in the specification, it suggests either that the element is sufficiently well known that it needs no description, and thus that it is not new, or that the inventor is unable to discuss it because they have not yet reduced it to practice and thus, according to patent law, not yet invented it.

16. *See infra* Part III.B.

17. Google disapproves of this practice. *See Irrelevant Keywords*, GOOGLE, https://support.google.com/webmasters/answer/66358?hl=en (last visited Dec. 18, 2017) ("'Keyword stuffing' refers to the practice of loading a webpage with keywords or numbers in an attempt to manipulate a site's ranking in Google search results.").

18. *See* U.S. Patent No. 6,450,942 col. 33 ll. 13–45 (filed Aug. 20, 1999) (issued Sept. 17, 2002).

for the preservation of muscle tone of astronauts."[19] Although this sounds quite exciting, there is no detail in the patent about "cosmic medicine" or astronauts. In reality, treatment of astronauts is far afield from the core invention of the patent.[20] Moreover, nobody (including the patentee) knows how to preserve muscle tone in astronauts.[21] Quite simply, "cosmic medicine" should not be in the patent claim because the patentee has not invented cosmic medicine.

Ancillary language appears across patents in all industries, although it is most common in pharmaceutical patents (almost 30% of all claim language) and least common in computer patents (22%).[22] Patent applications with more ancillary language are more likely to be rejected for lack of enablement or written description, suggesting that claims with ancillary language may not be valid.[23] Also, patents with more ancillary language are less valuable as measured by traditional proxies of value, such as forward citations and payment of maintenance fees.[24]

The empirical analysis in this Article documents *what* is happening with clutter, but it does not fully reveal *why* it occurs. Following the empirical analysis, this Article turns to the theoretical and, using an in-depth review of claim language, creates a typology of incentives for using ancillary language. Patentees may choose to include ancillary language for signaling purposes such as increasing the audience for their patent or advertising the functionality of their core invention, or for decoy purposes such as obscuring the core invention in a flurry of ancillary language.[25] Alternatively, patentees may include ancillary language as a hedge against uncertain future prosecution and litigation by creating options for interpretation or narrower backup claims.[26] Finally, ancillary language may be part of an affirmative strategy to "wear down" the patent examiner.[27]

Understanding this landscape can explain how the ubiquity of ancillary language contributes to many core problems with patent claims. For example, patent searches are known to return too many false positives,[28] which may be because the results are teeming with ancillary terms. In addition,

19. *Id.* at col. 34 ll. 34–35.

20. This is evident from reading the patent, which is directed to "achiev[ing] . . . heart load reduction" in a general sense, but does not discuss treatment of astronauts nor how the patented technology could be used to accomplish this goal. *Id.* at col. 33 ll. 14–21.

21. NASA is currently working on the question. *See Effect of Prolonged Space Flight on Human Skeletal Muscle (Biopsy)*, NASA (Nov. 22, 2016), http://www.nasa.gov/mission_pages/station/research/experiments/245.html.

22. *See infra* Part III.B.2.

23. *See infra* Part III.B.3.

24. *See infra* Part III.B.5.iii.

25. *See infra* Part IV.A.1.

26. *See infra* notes 212–19 and accompanying text.

27. *See infra* Part IV.A.3.

28. *See, e.g.*, THE EVOLVING IP MARKETPLACE, *supra* note 10, at 114.

comprehending patent claims is notoriously difficult.[29] Claims might be clearer if ancillary terms did not overwhelm the true invention and distract the reader. Further, scholars worry that patent examiners have insufficient time to properly review patent applications.[30] Examiners are charged with reviewing "every limitation"[31] of a claim, so a claim with extensive ancillary elements greatly increases examination time and burden. Finally, it is far from clear that these claims are even valid—which is consistent with widespread concerns about patent validity.[32] The Patent Act requires patentees to disclose the claimed invention in detail, to satisfy the "enablement" and "written description" requirements.[33] Copious ancillary language suggests that a claim may not meet these disclosure hurdles.

This Article suggests several policy solutions. First, examiners could use algorithms comparable to the ones developed for this study to flag enablement and written description problems and to cut down on clutter by making better use of the already-available prolix rejection.[34] Second, makers of patent search software could use this study to improve search capabilities.[35] Lastly, we might encourage patent applicants to define all claim terms (cutting down on clutter and confusion) and move excess language from the claims to the specification.[36]

Finally, this Article's empirical study also generates theoretical insights relevant to patent scholarship. It is a "bedrock principle" that the claim and the invention are the same thing, and that "focusing on the invention. . . . is nothing more than a short-hand reference for the claim[s]."[37] This Article presents considerable evidence that claims are neither synonymous nor coextensive with the invention, complicating our understanding of the correspondence between claim and invention. Further, although the ideal for patent claims is to concisely isolate the invention, this ideal may not reflect the realities of claim drafting and the ways people read and understand words. Grounding patent theory in the empirics of patent practice requires rethinking this ideal. The Article concludes with a framework that balances precision, concision, and accuracy as a way to understand the proper shape of patent claims.

---

29. Burk & Lemley, *supra* note 2, at 1744.

30. Michael D. Frakes & Melissa F. Wasserman, *Does the U.S. Patent and Trademark Office Grant Too Many Bad Patents?: Evidence From a Quasi-Experiment*, 67 STAN. L. REV. 613, 619 (2015).

31. MPEP, *supra* note 13, § 2103.

32. *E.g.*, John R. Allison & Lisa Larrimore Ouellette, *How Courts Adjudicate Patent Definiteness and Disclosure*, 65 DUKE L.J. 609, 611 (2016).

33. 35 U.S.C. § 112(a) (2012).

34. *See infra* Part IV.D.1. The prolix rejection allows examiners to reject claims "as prolix" when they contain "long recitations or unimportant details." MPEP, *supra* note 13, § 2173.05(m).

35. *See infra* Part IV.D.2.i.

36. *See infra* Part IV.D.2.ii.

37. Oskar Liivak, *Rescuing the Invention from the Cult of the Claim*, 42 SETON HALL L. REV. 1, 2–3 (2012) (criticizing the "cult" and arguing that the invention is distinct from the claims).

The Article proceeds as follows. Part II provides a background on the theory and goals underlying patent claims. It also surveys literature on practical problems and policy efforts relating to claims, as well as scholarly debates about the relationship between claim and invention. Part III describes the methodology used to identify ancillary language and presents the results of the study. Part IV.A discusses the motivation for using ancillary language in patents. Part IV.B argues that ancillary language may contribute to practical problems in patent law. Part IV.C analyzes theoretical implications, particularly the need to rethink the synonymy of claim and invention, and to reevaluate whether it is actually possible to completely separate the invention from its context. Part IV.D suggests policy approaches intended to return patent claims to their intended format: concise and accurate phrases succinctly outlining the relevant invention.

II. PATENT CLAIMS

A. *UNDERSTANDING CLAIMS*

The first Patent Act of 1790 required inventors to write a "description . . . of the thing or things . . . invented or discovered."[38] This might include "an entire machine, although most parts of it may have been long known and used" and the patent would only cover the improvement to the machine discovered by the patentee.[39] This requirement proved problematic, as the differences between the inventions and the prior art were often vaguely described. The question of what similar devices or processes would infringe was also not specified.[40] Further, inventions were often improvements on already created processes, and patents would describe the improved machine as a whole without specifying the particular improvement.[41] It was common to find courts holding that the patent "is fatally defective.[42] As the matter stands, the nature of the improvement is altogether unintelligible."[43] A frustrated Chief Justice Marshall asked, "how is any man to inform himself what it is that is patented, so that he may avoid the danger of infringement?"[44] An equally frustrated Justice Story complained that "[i]f . . . the description

38. Patent Act of 1790, ch.7, § 2, 1 Stat. 109, 110 (repealed 1793).

39. Thomas P. Jones, *Information to Persons Applying for Patents, or Transacting Other Business at the Patent Office*, 6 FRANKLIN J. & AM. MECHANICS' MAG. 332, 334 (1828).

40. 1 MARK A. LEMLEY, PETER S. MENELL, & ROBERT P. MERGES, INTELLECTUAL PROPERTY IN THE NEW TECHNOLOGICAL AGE: 2016, at III-167 (2016) ("Justice Joseph Story, who would emerge as the leading patent jurist . . . immediately came to see the problems with vague and conclusory descriptions of inventions.").

41. Whittemore v. Cutter, 29 F. Cas. 1123, 1124 (C.C.D. Mass. 1813); *see* Norris Boothe, Note, *Exercising a Duty of Clarity:* Nautilus, Inc. v. Biosig Instruments, Inc., 30 BERKELEY TECH. L.J. 445, 448 (2015).

42. *See, e.g.*, Isaacs v. Cooper, 13 F. Cas. 153, 154 (C.C.D. Pa. 1821).

43. *See id.* at 154.

44. Evans v. Eaton, 20 U.S. 356, 391 (1822).

in the patent mixes up the old and the new, and does not distinctly ascertain for which, in particular, the patent is claimed, . . . it is impossible for the court to say, what, in particular, is covered."[45]

To ensure that patents had clear boundaries, the Patent Act of 1836[46] separated patents into two sections: the specification[47] and the claims.[48] The specification consists of a detailed narrative description of the invention, how it is made and used, preferred embodiments, any experiments or testing performed, and background on the field of the invention.[49] The specification contains context for the invention and may run over 20 or 30 pages.[50] The claims appear at the end of the specification and are a summation of the patent right claimed by the patentee.[51] "[C]laim[s] particularly point[] out and distinctly claim[] the subject matter which the applicant regards as his invention or discovery."[52] Claims are limited to one sentence each and define the boundaries of the patent monopoly.[53]

This change to the patent structure creates two sections with different emphases. The specification is long, descriptive, includes context, and "should ideally serve as a glossary to the claim terms."[54] Claims, on the other hand, value precision and parsimony, deliberately excluding contextual material. Details and context remain important, but are placed in the specification separate from the claim.[55] Thus, a patent claim directed to the invention of a paperclip might read:

"I claim a paperclip."

This claim limits the patent grant to paperclips alone and excludes, for example, staples or binder clips. Claims are often written in language more complex than the paperclip claim above. For example, a claim might describe

45. Lowell v. Lewis, 15 F. Cas. 1018, 1020 (C.C.D. Mass. 1817).

46. RIDSDALE ELLIS, PATENT CLAIMS § 3 (1949). Strictly speaking, claims were not new. Patent applicants had been using claims for years before the statutory amendment as a method of clarifying their application. *Id.*

47. The term "specification" is sometimes used to describe the entire patent—both the narrative portion and the claims. Here it is used to describe only the narrative portion of the patent and excludes the claims.

48. The requirement that patents have both a descriptive specification and a claims section is still in place today. 37 C.F.R. § 1.51 (2016). Patents may also contain drawings. *Id.* § 1.51(b)(3).

49. *Id.* § 1.71; *see* Jeanne C. Fromer, *Patent Disclosure*, 94 IOWA L. REV. 539, 545 (2009).

50. *E.g.*, Dennis Crouch, *Patent Specifications Continue to Rise in Size*, PATENTLY-O (Apr. 23, 2012), https://patentlyo.com/patent/2012/04/patent-specifications-continue-to-rise-in-size.html (In 2010, patent specifications averaged almost 50,000 characters. This is around 30 pages.).

51. MPEP, *supra* note 13, § 2173.

52. 37 C.F.R. § 1.75(a).

53. MPEP, *supra* note 13, § 608.01(m).

54. *Id.* § 2173.03.

55. *Id.* § 608.01(i)(d)(1) ("[C]laim terms must . . . be ascertainable by reference to the description.").

a piece of bread in a sandwich as: "a first bread layer having a first perimeter surface coplanar to a contact surface."[56]

Patents may contain multiple claims so long as no two claims cover the same material,[57] and the claims are directed to the same general invention.[58] The Patent Act provides an organizational format for applications with multiple claims. By statute, claims are divided into two types: independent claims and dependent claims.[59] Independent claims stand alone and do not reference any other claims. Dependent claims must refer back to another claim.[60] This is illustrated by the example below:

> I claim:
>
> (1) A paperclip.
>
> (2) The paperclip of Claim 1 wherein the paperclip is made of metal.
>
> (3) The paperclip of Claim 2 wherein the paperclip is folded into the shape of a spiral.

Dependent claims must be narrower than the independent claims to which they refer.[61] Thus, patents contain multiple claims of successively narrower scope and multiple claims that cover slightly different areas.[62] There is no mandate to use dependent claims, but they are encouraged by the Patent Trademark Office's ("PTO") fee structure.[63]

B. CLAIMS AND PATENT THEORY

Claims are the heart of the patent. As such, there are a broad set of scholarly debates around the ideal shape and purpose of claims. As a preliminary matter, there is a widespread and multi-faceted debate about claim clarity. Some scholars question whether claims are clear. Those who believe that claims are not clear ask why and how lack of clarity arises and

---

56. COLLEEN CHIEN ET AL., REDESIGNING PATENT LAW 2 (forthcoming) (citing U.S. Patent No. 6,004,596 (issued Dec. 21, 1999)).

57. 37 C.F.R. § 1.75(b).

58. 35 U.S.C. § 121 (2012) ("If two or more independent and distinct inventions are claimed in one application, the Director may require the application to be restricted to one of the inventions.").

59. *Id.* § 112(c).

60. *Id.* § 112(d) ("[A] claim in dependent form shall contain a reference to a claim previously set forth and then specify a further limitation of the subject matter claimed. A claim in dependent form shall be construed to incorporate by reference all the limitations of the claim to which it refers.").

61. *See* 37 C.F.R. § 1.75(c).

62. This is a longstanding practice. *See, e.g.*, Parke-Davis & Co. v. H.K. Mulford Co., 189 F. 95, 102–03 (C.C.S.D.N.Y. 1911).

63. As part of the basic filing fee, the PTO allows applicants to file three independent and 17 dependent claims. 37 C.F.R. § 1.16. Applicants can file additional claims, but the Office levies additional per claim fees. *Id.*

explore the consequences of the problem.[64] Other scholars question whether clarity is actually a goal that should be pursued by the patent system.[65] By exploring one source of opacity in claims, this project provides empirical evidence that will generally impact the debate.

In addition, the results of this study relate to debates about the relationship between claim language and the invention. These debates generally revolve around the assumption that claims are exclusively about the invention. Since this Article argues that claims include a great deal of language that is *not* about the invention, the results presented in this Article provide evidence that may shape further thinking in these areas. The debates are sketched below.

Because claims define the boundaries of the patent,[66] patent theory often equates claims with the invention at issue in the patent or uses "claim" and "invention" interchangeably. Christopher Cotropia terms this the "claim-centered" invention view, which regards "the claim itself as the invention for patent law purposes"[67] such that "the claim *is* the invention."[68] Cotropia explains that claim-invention interchangeability has heavily influenced the development of several aspects of patent law.[69]

Oskar Liivak has termed patent theory's equating of patent claims and patented inventions "the cult of the claim."[70] This "cult," Liivak states, argues that the patent claim and the invention are the same thing, and that "focusing on the invention . . . is nothing more than a short-hand reference for the claim[s]."[71] Liivak explains that equating claims and inventions has significant implications for broad areas of patent law such as validity, claim scope, and

64. *See infra* Part II.C.2.

65. *See, e.g.*, Tun-Jen Chiang & Lawrence B. Solum, *The Interpretation-Construction Distinction in Patent Law*, 123 YALE L.J. 530, 533–34 (2013); Jeanne C. Fromer & Mark A. Lemley, *The Audience in Intellectual Property Infringement*, 112 MICH. L. REV. 1251, 1267 (2014); John M. Golden, *Construing Patent Claims According to Their "Interpretive Community": A Call for an Attorney-Plus-Artisan Perspective*, 21 HARV. J.L. & TECH. 321, 368 (2008); Mark A. Lemley, *The Changing Meaning of Patent Claim Terms*, 104 MICH. L. REV. 101, 110 (2005).

66. The Supreme Court has long recognized that "a patent's precise claims mark its monopoly boundaries." Graver Tank & Mfg. Co. v. Linde Air Prods. Co., 339 U.S. 605, 617 (1950). Patent claims are also commonly analogized to the "metes and bounds" system of determining a property's boundaries. *See, e.g.*, Ariad Pharm., Inc. v. Eli Lilly & Co., 598 F.3d 1336, 1347 (Fed. Cir. 2010) ("Claims define the subject matter that, after examination, has been found to meet the statutory requirements for a patent. Their principal function, therefore, is to provide notice of the boundaries of the right to exclude and to define limits . . . ." (citation omitted)).

67. Christopher A. Cotropia, *What is the "Invention"?*, 53 WM. & MARY L. REV. 1855, 1861–62 (2012).

68. *Id.* at 1886.

69. *See id.* at 1887, 1889, 1910 (discussing the influence of claim-invention interchangeability on claim construction, written description, and the doctrine of equivalents).

70. Liivak, *supra* note 37, at 5.

71. *Id.* at 2 (criticizing the "cult" and arguing that the invention is distinct from the claims).

disclosure.[72] Jeffrey Lefstin has also written about the "synonymy of invention and claim,"[73] as have many other scholars,[74] and the leading patent treatise states that "the claim defines the invention for purposes of both patentability and infringement."[75] Many scholars, including those named above, argue that this synonymy causes courts to focus on the claims to the exclusion of the invention—suggesting that they do not believe the two to be entirely synonymous.[76] Despite these rumblings of dissent, for purposes of patent theory and doctrine, the claim is the invention and the invention is the claim.

C. *THE PRACTICE AND POLICY OF CLAIMS*

Debates about claims are not only theoretical; there is a vast literature exploring practical problems with claims, particularly their failure to clearly and "distinctly"[77] point out the boundaries of the invention. This Article argues that claims are often not distinctly pointing out the invention because they are cluttered with ancillary language. Although ancillary language itself is not previously documented in patent literature, many policymakers and scholars report problems that appear to be related at least in part to the prevalence of ancillary language. The discussion below explores what is currently known about these problems.

1. Disclosure

Because claims are a bare-bones way of conveying information, a reader seeking details about the invention must turn to the specification. The specification must therefore provide enough information about the material in the claims to show that the applicant possessed the claimed invention and

72. *Id.* at 9–12.

73. Jeffrey A. Lefstin, *The Formal Structure of Patent Law and the Limits of Enablement*, 23 BERKELEY TECH. L.J. 1141, 1145 n.15 (2008).

74. *See, e.g.*, Bernard Chao, *The Infringement Continuum*, 35 CARDOZO L. REV. 1359, 1366 (2014) (noting the practice of determining the patented "invention by examining claims" but criticizing the practice on the ground that "there may be a surprising disconnect between the specification and the claims"); Amy L. Landers, *Patent Claim Apportionment, Patentee Injury, and Sequential Invention*, 19 GEO. MASON L. REV. 471, 479 (2012) (examining the equivalency and suggesting that it may not be correct "[a]s a practical matter"); Saurabh Vishnubhakat, *Cognitive Economy and the Trespass Fallacy: A Response to Professor Mossoff*, 65 FLA. L. REV. 38, 38 (2014) (discussing claims as proxies for the invention).

75. DONALD S. CHISUM, CHISUM ON PATENTS § 8.01 (2008).

76. *E.g.*, ROBERT PATRICK MERGES & JOHN FITZGERALD DUFFY, PATENT LAW AND POLICY: CASES AND MATERIALS 769 (6th ed. 2013); Burk & Lemley, *supra* note 2, at 1745–46; Chiang & Solum, *supra* note 65, at 540–43; Cotropia, *supra* note 67, at 1897; Liivak, *supra* note 37, at 40.

77. 35 U.S.C. § 112(b) (2012).

to teach someone knowledgeable in the field of the invention[78] how to make and use the claimed invention.[79]

Policymakers are quite concerned that applicants routinely include claim language that is not supported by the specification.[80] In May 2016 the PTO announced a case study to determine if continuation applications claim subject matter that is not sufficiently described, and whether examiners are appropriately enforcing the written description requirement.[81] The PTO has additionally run several new training programs for examiners on the topics of claim clarity, enablement, and written description.[82] Despite reform efforts, policymakers know little about why patentees use claim language in ways incompatible with these requirements. Essentially, the problem is one of mismatch between the material described in the claims and the invention as it is set out in the specification. It may be that this mismatch arises because some portion of the claim language does not actually relate to the invention.

2. Clarity

Patent claims are supposed to clearly delineate the boundaries of the patent monopoly, so that competitors are on notice about the actions they are barred from taking without permission of the patentee.[83] However, claims often "lack clarity" or are "vague" and "otherwise difficult to interpret."[84]

---

78. Formally known as a "person having ordinary skill in the art." This is a hypothetical person familiar with all relevant information in a technical field. *See* Standard Oil Co. v. Am. Cyanamid Co., 774 F.2d 448, 452 (Fed. Cir. 1985). The person having ordinary skill in the art is sometimes analogized to tort law's "reasonable person." *E.g.*, Panduit Corp. v. Dennison Mfg. Co., 810 F.2d 1561, 1566 (Fed. Cir. 1987).

79. This is called the "enablement" requirement. 35 U.S.C. § 112(a). The related written description doctrine requires the specification to contain enough information to demonstrate to a person of ordinary skill in the art that the inventor possessed the claimed invention. Moba, B.V. v. Diamond Automation, Inc., 325 F.3d 1306, 1319 (Fed. Cir. 2003); *see also In re* Wands, 858 F.2d 731, 737 (Fed. Cir. 1988) ("The term 'undue experimentation' does not appear in the statute, but it is well established that enablement requires that the specification teach those in the art to make and use the invention without undue experimentation.").

80. As are courts and scholars. *See* Allison & Ouellette, *supra* note 32, at 611–12.

81. *Selected Topics for Case Studies Pilot*, USPTO, https://www.uspto.gov/patent/initiatives/selected-topics-case-studies-pilot (last modified Jan. 11, 2017, 1:58 PM).

82. *USPTO-led Executive Actions on High Tech Patent Issues*, USPTO, http://www.uspto.gov/patent/initiatives/uspto-led-executive-actions-high-tech-patent-issues (last modified Aug. 1, 2017, 11:11 AM) [hereinafter *Actions on High Tech Patent Issues*].

83. *E.g.*, Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co., 535 U.S. 722, 732 (2002) ("If competitors cannot be certain about a patent's extent, they may be deterred from engaging in legitimate manufactures outside its limits, or they may invest by mistake in competing products that the patent secures. In addition the uncertainty may lead to wasteful litigation between competitors . . . .").

84. FED. TRADE COMM'N, THE EVOLVING IP MARKETPLACE: ALIGNING PATENT NOTICE AND REMEDIES WITH COMPETITION 81 (2011), https://www.ftc.gov/sites/default/files/documents/reports/evolving-ip-marketplace-aligning-patent-notice-and-remedies-competition-report-federal-trade/110307patentreport.pdf [hereinafter ALIGNING PATENT NOTICE AND REMEDIES WITH COMPETITION].

Compounding the challenge of reading and understanding patent claims, patentees may choose (sometimes deliberately, sometimes unavoidably) to use terms that are imprecise.[85]

In response to these concerns, the White House called for patents "with clearly defined boundaries to provide adequate notice to help others avoid costly and needless litigation down the road."[86] The PTO recently completed a "Glossary Pilot Program" for software-related applications that encouraged patent applicants to include glossaries in the specification "to define terms used in the patent claims."[87] A recent Government Accountability Office report on patent quality criticized the PTO because it "does not specifically require patent applicants to clearly define the terms used in their applications . . . ."[88] This problem arises in part because terms in the claims are not defined in the specification. As described in more detail in Section III, *infra*, ancillary language is often not defined in the specification. It may therefore contribute to this problem.

3. Searchability

An additional way in which claims can fail to provide notice of their boundaries arises in the context of patent searches. Before launching a product, it is good practice to search for patents that might potentially be infringed.[89] Patents are often sought through keyword searches of patent claims,[90] but, in areas where the number of patents is high and patent language is abstract and non-standardized (software in particular), searches can be extremely costly and sometimes almost impossible.[91] A substantial

85. Nautilus, Inc. v. Biosig Instruments, Inc., 134 S. Ct. 2120, 2124 (2014).

86. *Actions on High Tech Patent Issues*, *supra* note 82.

87. *USPTO Launches New Glossary Pilot Program to Promote Patent Claim Clarity*, USPTO, http://www.uspto.gov/about-us/news-updates/uspto-launches-new-glossary-pilot-program-promote-patent-claim-clarity (last modified Dec. 11, 2014, 5:09 PM). The Pilot ran from 2014 to 2016 and the PTO has not yet fully analyzed the results, but a survey indicated that a majority of examiners believed that a glossary requirement would improve patent quality and claim clarity. U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-16-490, INTELLECTUAL PROPERTY: PATENT OFFICE SHOULD DEFINE QUALITY, REASSESS INCENTIVES, AND IMPROVE CLARITY 33–34 (2016), http://www.gao.gov/assets/680/678113.pdf.

88. U.S. GOV'T ACCOUNTABILITY OFFICE, *supra* note 87, at 36. The report went on to warn: "Without making use of tools to improve the clarity of patent applications . . . the agency is at risk of issuing unclear patents that may not comply with statutory requirements." *Id.*

89. *See* ALIGNING PATENT NOTICE AND REMEDIES WITH COMPETITION, *supra* note 84, at 3.

90. B.P. Nagori & Vipin Mathur, *Basics of Writing Patent Non-Infringement and Freedom-to-Operate Opinions*, 14 J. INTELL. PROP. RTS. 7, 8–9 (2009).

91. ALIGNING PATENT NOTICE AND REMEDIES WITH COMPETITION, *supra* note 84, at 16 ("Identifying and reviewing the patents and applications that might conceivably apply to a new product often present daunting challenges in IT industries."); Christina Mulligan & Timothy B. Lee, *Scaling the Patent System*, 68 N.Y.U. ANN. SURV. AM. L. 289, 317 (2012) ("[P]atent clearance is practically impossible. In software, for example, patent clearance would require the services of many more patent attorneys than exist in the United States."). In the pharmaceutical area, where searches are routinely performed, these searches can cost from four figures to six figures. *See* BESSEN & MEURER, *supra* note 5, at 48, 55.

portion of the cost of clearance searches is caused by false positive results swept in by keyword searches that must then be manually removed by a patent attorney or other searcher.[92] This is because, unlike a Google search, where a searcher can stop as soon as relevant results are found,[93] patent searchers cannot stop looking until all results have been reviewed,[94] as failing to find a relevant patent can result in millions of dollars in damages.[95]

The complexity of this process means that clearance searches are difficult, expensive, and sometimes ineffective.[96] An FTC report describing the challenges of sifting through a large number of patents recommends that "the PTO . . . convene a government/industry task force or hold a workshop to explore ways of" clarifying patent language and improving the public's ability to search for patents.[97]

The nature of patent searching means that the drafter's choice of words included in the claims has a great deal of influence on whether that patent will appear in later searches. Ideally, patent claims would include sufficient descriptive terminology so that competitors could find the patent, but would not include additional less relevant terminology. Because searchers are generally only interested in the invention covered by the claims, ancillary language may be a significant source of false positives.

4. Examinability

The PTO is under attack for granting poor quality patents.[98] Much of the criticism focuses on the small amount of time allocated to examining each

92. For example, a 2003 survey of industry respondents found that "there may be a large number of patents to consider initially—sometimes in the hundreds, and that this number is surely larger than in the past" but that, even in a complicated case, the number of relevant patents was about six to twelve. John P. Walsh et al., *Effects of Research Tool Patents and Licensing on Biomedical Innovation*, *in* PATENTS IN THE KNOWLEDGE-BASED ECONOMY 285, 294 (Wesley M. Cohen & Stephen A. Merrill eds., 2003). This number is larger in the software industry. BESSEN & MEURER, *supra* note 5, at 213 ("Checking thousands of patents is clearly infeasible for almost any software product.").

93. Google searchers rarely go past the first page of results. *2nd Page Rankings: You're the #1 Loser*, GRAVITATE ONLINE (Apr. 12, 2011, 9:29 PM), http://www.gravitateonline.com/google-search/2nd-place-1st-place-loser-seriously (finding that 94% of people do not click on a second page result).

94. Kristine H. Atkinson, *Toward a More Rational Patent Search Paradigm*, *in* PROCEEDINGS OF THE 1ST ACM WORKSHOP ON PATENT INFORMATION RETRIEVAL 37, 38 (2008) ("Any searcher knows that the first thing to do when you get a thousand hits is to tighten the strategy—but . . . . [a] more stringent strategy necessarily poses the risk that keywords that captured a patent may get trumped by excluding a word that is innocuous in context . . . .").

95. *See* Michael J. Mazzeo et al., *Explaining the "Unpredictable": An Empirical Analysis of U.S. Patent Infringement Awards*, 35 INT'L REV. L. & ECON. 58, 63 tbl.1 (2013).

96. EVOLVING IP MARKETPLACE, *supra* note 10, at 90.

97. *Id.* at 111.

98. Robert P. Merges, *As Many as Six Impossible Patents Before Breakfast: Property Rights for Business Concepts and Patent System Reform*, 14 BERKELEY TECH. L.J. 577, 589–91 (1999); R. Polk Wagner, *Understanding Patent-Quality Mechanisms*, 157 U. PA. L. REV. 2135, 2137 (2009).

patent—an average of 18 hours per patent.[99] Part of the strain on patent examiners is the length and clarity of claims. Patent examiners are charged with evaluating every element of a patent claim. The Manual of Patent Examining Procedure instructs examiners that, "when evaluating the scope of a claim, *every* limitation in the claim must be considered."[100] It further requires that, in order to make a rejection of a patent, "personnel must articulate . . . a finding that the prior art included *each element* claimed . . . ."[101] As more claims are added to each patent,[102] examining the claims surely becomes more time intensive. Further, if claims are complex, long, vague, and difficult to interpret, it stands to reason that evaluation time would also increase. If ancillary language increases both the intricacy and length of claims, it may increase examination time.

* * *

In sum: there are a wide range of claim issues where policy makers are presently seeking greater understanding of the problem and suggestions for improvement. Given the widespread concerns about the efficacy of claims, a better understanding of both claims and problems with claims is vital.

III. EMPIRICAL STUDY

This study is a large scale textual analysis of the claims of 40,000 patents and applications. It arose from the observation that many patent claims appeared to describe concepts that were not elaborated upon in the specification, and did not seem closely related to the core invention of the patent. In light of patent law's requirement that claim elements be described in and enabled by the specification, and the theoretical understanding that claims are "synonymous"[103] with the invention, this observation was surprising. This project sets out to understand the magnitude of this phenomenon and obtain information about its characteristics and motivations.

A. METHODOLOGY

This Article uses the terminology "central" to refer to claim language that describes the patented invention and "ancillary" to refer to all additional claim language. "Ancillary" may refer to two different types of language. First, it may describe concepts that are simply not part of a new product or process, and thus not invented by the patentee. Second, it may describe concepts that may have been newly conceived by the patentee but are not anything that the

99. Mark A. Lemley, *Rational Ignorance at the Patent Office*, 95 NW. U. L. REV. 1495, 1500 (2001).

100. MPEP, *supra* note 13, § 2103(I)(C) (emphasis added).

101. *Id.* § 2143(I)(A)(1) (emphasis added).

102. The number of claims per patent is increasing. *See* John R. Allison & Mark A. Lemley, *The Growing Complexity of the United States Patent System*, 82 B.U. L. REV. 77, 103–04 (2002); Jean O. Lanjouw & Mark Schankerman, *Characteristics of Patent Litigation: A Window on Competition*, 32 RAND J. ECON. 129, 140 (2001).

103. Lefstin, *supra* note 73, at 1145.

patentee actually appears to have made or to be capable of making (in patent parlance, these claims would be said to be lacking proper "written description" and/or not "enable[d]").[104]

In order to undertake a large-scale analysis of whether claim language is central or ancillary, I sought characteristics of claim language that could be categorized by a computer program. The key proxy for whether claim language is central is the number of times the language appears in the specification of a patent. This proxy is explained below.

As discussed in more detail in Part II, *supra*, patents consist of two primary sections: the specification[105] and the claims.[106] The specification is a detailed narrative description of the invention, how it is made and used, preferred embodiments, any experiments or testing performed, and background on the field of the invention.[107] It is followed by the claims, which define the boundaries of the patent.

All elements of a claim must be enabled by and described in the specification.[108] This means that discussion of the element in the specification must be sufficient to demonstrate to someone knowledgeable in the field of the invention that the inventor possessed the claimed invention.[109] The specification must also be sufficient to teach a skilled artisan how to make and use the claimed invention without "undue experimentation."[110] The only exception to the requirement that an element of a claim must be described in the specification is for elements that are "well known in the art."[111] For example, a patent on a yeast microorganism wherein a particular gene is inactivated would not have to describe how to deactivate the gene if methods of doing so were well known to those working in the field.[112]

---

104. *See* 35 U.S.C. § 112(a) (2012).

105. The term "specification" is sometimes used to describe the entire patent—both the narrative portion and the claims. Here it is used to describe only the narrative portion of the patent and excludes the claims.

106. 37 C.F.R. § 1.51(b)(1) (2016). Patents may also contain drawings. *Id.* § 1.51(b)(3).

107. *See id.* § 1.71.

108. 35 U.S.C. § 112.

109. Moba, B.V. v. Diamond Automation, Inc., 325 F.3d 1306, 1320 (Fed. Cir. 2003) (citing Enzo Biochem, Inc. v. Gen-Probe Inc., 323 F.3d 956, 966 (Fed. Cir. 2002)).

110. *In re* Wands, 858 F.2d 731, 733 (Fed. Cir. 1988). Precisely what level of experimentation is "undue" varies in different circumstances. *Id.* at 737 ("The determination of what constitutes undue experimentation in a given case requires the application of a standard of reasonableness, having due regard for the nature of the invention and the state of the art.").

111. *In re* Buchner, 929 F.2d 660, 661 (Fed. Cir. 1991); MPEP, *supra* note 13, § 2164.01 ("A patent need not teach, and preferably omits, what is well known in the art.").

112. Butamax(TM) Advanced Biofuels LLC v. Gevo, Inc., 746 F.3d 1302, 1314–15 (Fed. Cir. 2014) (reversing the district court's holding of invalidity due to lack of enablement; finding a question of material fact as to whether methods of deactivating the pathway of the gene were known by those of ordinary skill in the art).

Patentable subject matter is, by definition, not "well known in the art." To be patentable, an invention must be novel and nonobvious.[113] Any invention that is novel and nonobvious cannot be "well known." It must, therefore, be described in some detail in the specification.[114] For example, a patent application on a novel process for weakening tropical storms by releasing super coolants into the storm was rejected for lack of enablement because the specification did not describe the amount of super coolant needed or the optimal timing of the release.[115]

A novel and nonobvious invention may contain elements that are not themselves novel and nonobvious, which can be necessary parts of a patent claim, but that have aspects of novelty and nonobviousness, requiring them to be described in the specification. For example, a patent may combine two elements that are individually well known in the art in a manner that is novel and nonobvious. However, the combination is not well known in the art, and would have to be described in detail in the specification. Claim elements that are not themselves novel or nonobvious might also be added to satisfy other requirements of patentability, for example to narrow a claim to ensure that the claim is useful, or that it is enabled by the specification. In each of these cases, there would likely need to be a discussion of that element in the specification because it interacts in vital ways with the novel and nonobvious elements of the claim.

For example, U.S. Patent No. 6,505,572 claims "[i]n combination, a water vessel and a chair."[116] The claim further requires that the chair be positioned at the back of the vessel such that it creates a larger wake behind the vessel.[117] The specification explains that this makes the vessel more suitable for pulling a wake boarder or water skier.[118] The two elements of this invention, water vessels and chairs, are certainly well known in the art. However, the combination of the two for the purpose of enhancing wake requires more explanation in the specification. In particular, the chair must be positioned in a specific manner, should be made of an impermeable material so that it does not get soggy, and will work better if it is inflatable and filled with fluid.[119] Enabling the combination of well-known elements requires some explanation in the specification. In this example, both "vessel" and

113. 35 U.S.C. §§ 102–03.

114. *E.g.*, Auto. Techs. Int'l, Inc. v. BMW of N. Am., Inc., 501 F.3d 1274, 1283 (Fed. Cir. 2007) ("Although the knowledge of one skilled in the art is indeed relevant, the novel aspect of an invention must be enabled in the patent."); Genentech, Inc. v. Novo Nordisk A/S, 108 F.3d 1361, 1366 (Fed. Cir. 1997) ("It is the specification, not the knowledge of one skilled in the art, that must supply the novel aspects of an invention in order to constitute adequate enablement.").

115. *In re* Hoffman, 558 F. App'x 985, 986 (Fed. Cir. 2014).

116. U.S. Patent No. 6,505,572 col. 3 ll. 64–67 to col. 4 ll. 1–7 (filed Oct. 27, 2000) (issued Jan. 14, 2003).

117. *Id.*

118. *Id.* at col. 1 ll. 46–51.

119. *Id.* at cols. 1–3.

"chair" are instances of central language, as they are mentioned frequently in the specification.

Thus, if an element in a claim is mentioned very little in the specification, it is likely for one of two reasons: (1) it is an old element that does not contribute novelty to the invention; or (2) it is attempting to claim something that the patentee has not enabled or described sufficiently in the specification—meaning that the element may not be something the patentee invented. This study uses the frequency with which a claim element is mentioned in the specification as a proxy for whether the element is central or ancillary.

1. Measuring Element Frequency

The sample for this study is a randomly generated[120] list of 25,000 patents and 15,000 applications filed before 2015[121] available on the USPTO's website.[122] For each patent in this list, an algorithm was used to identify the text of the claims.[123] The claims were then broken down into individual words using the Natural Language Toolkit ("NLTK"), a platform interpreting textual data using Python.[124] The NLTK algorithm[125] was used to identify singular nouns within the patent claims. The study is restricted to singular nouns in order to minimize false positives caused by changes in grammar, such as verb tense, and from prepositions, articles, and legal language (for example, "comprising," "whereby," "said," etc.), which are not relevant to this study.[126] After claim words were identified, duplicate words were removed. Some patents did not have text available on the USPTO's website and were excluded, leaving a total of 24,116 patents.

For each identified word in a patent's claims, the algorithm determined how many times the word appeared in the patent's specification. Reasonable minds may differ on how little a word must appear in the specification before we can be confident that the claim element is ancillary. However, it seems unlikely that an element that is part of a novel and nonobvious invention can be adequately described and enabled if it is never mentioned, or mentioned

120. Patent numbers were generated using a random number generator.

121. The USPTO database of full-text granted patents begins with patents granted in 1976 and applications in 2001. Note that because the number of patents granted each year has been increasing, the sample has a greater number of patents for more recent years.

122. USPTO PATENT FULL-TEXT AND IMAGE DATABASE, USPTO, http://patft.uspto.gov/netahtml/PTO/search-bool.html (last visited Dec. 18, 2017).

123. All algorithms mentioned in this Part are on file with the author.

124. *Natural Language Toolkit*, NLTK, http://www.nltk.org (last visited Dec. 18, 2017).

125. Specifically, the "pos_tag" function.

126. I specifically excluded these words, and all other words that are "stop words" in the USPTO's patent search. "Stopwords are terms that appear so frequently in patent text that they lose their usefulness as search terms" and include words such as "invention," "claim," "means," "embodiment," and others. *Stopwords*, USPTO PAT. FULL-TEXT & IMAGE DATABASE USPTO, http://patft.uspto.gov/netahtml/PTO/help/stopword.htm (last visited Dec. 18, 2017).

only once or twice. Thus, for purposes of this study, "ancillary" language is defined as language in a patent claim that appears in the specification two or fewer times.[127] Note that the choice of "two or fewer" is somewhat arbitrary—it is also unlikely that an element in a novel or nonobvious invention can be adequately described if it is mentioned only three or four times in the specification. Because this is the first study of its type, I chose to use a low number to provide a conservative definition. Thus, I can confidently say that, since the amount of ancillary language is large using my definition, the true amount of ancillary language is likely *larger*.

To illustrate the methodology, consider U.S. Patent No. 5,650,185. The patent claims a "non-aerosol product delivery system for use in food preparation," producing a "uniform, widely dispersed spray pattern," which can be used to spray flavors on foods.[128] Previous spray systems were unable to deliver this uniform spray pattern, instead producing a "ragged, uneven spray."[129] The claims specify that the system may be used to deliver a variety of flavors as diverse as "rosemary" and "liver."[130] The algorithm identified the words "uniform," "spray," "rosemary," and "liver" (among others) and then determined how many times each word appeared in the specification. The patent's key advance over the prior art is the ability to produce a uniform spray, thus, the concept is discussed at length in the specification, and the word "uniform" appears 34 times and "spray" 92 times. These terms are therefore classified as central language. By contrast, the ability to incorporate rosemary and liver flavors into a spray are apparently somewhat incidental to the main invention and may have been known in the prior art, and the words "rosemary" and "liver" appear only once in the specification. These terms are therefore classified as ancillary language.

2. Additional Data

Additional data were also collected on other characteristics of the sampled patents. Specifically, data were obtained on the number of claims, specification length, number of forward and backward citations,[131] filing date,

127. To validate the choice to define ancillary language as appearing two or fewer times in the specification, several of the analyses in Part III were also conducted using a definition of ancillary language of either "appearing once in the specification" or "appearing three or fewer times in the specification." The directionality and significance of the results did not change.

128. U.S. Patent No. 5,650,185 claim 1 (issued July 22, 1997).

129. *Id.* at col. 1 l. 39.

130. *Id.* at claims 11, 13. Specifically, the claims read "wherein said oil-based liquid food flavor concentrate is further defined as comprising one or more selected from the group consisting of . . . liver." *Id.* at claim 13.

131. Obtained from Google's patent files. GOOGLE, https://patents.google.com (last visited Nov. 7, 2017). Forward citation counts include citations through May 4, 2015. Forward citations are widely used as a proxy for patent value, under the theory that more important patents will be cited more often. *E.g.*, Dietmar Harhoff et al., *Citation Frequency and the Value of Patented Inventions*, 81 REV. ECON. & STAT. 511, 512 (1999). Forward citations are not considered a precise proxy.

grant date,[132] prosecution length,[133] whether the application was a continuation or continuation-in-part of an older application, NBER industry classification,[134] and maintenance data.[135]

Further data were obtained on the amount of ancillary language in the original application and the change in ancillary language between the original application and the granted patent.[136] The full text of patent applications is only available for applications filed in 2001 or later, thus only 8,974 of the patents in this sample had a corresponding application available.

An additional 15,000 applications were randomly selected.[137] This sample includes applications that both were and were not ultimately granted. I used data provided by Michael Frakes and Melissa Wasserman[138] to determine whether each application had been rejected during prosecution on any of the following sections of title 35 of the United States Code: 101,[139] 102,[140] 103,[141] 112 paragraph 1,[142] 112 paragraph 2.[143]

C. Gay & C. Le Bas, *Uses Without Too Many Abuses of Patent Citations or the Simple Economics of Patent Citations as a Measure of Value and Flows of Knowledge*, 14 ECON. INNOVATION & NEW TECH. 333, 335 (2005). Backward citations may correlate with patent breadth, likelihood of validity, and patent value, although these correlations are speculative. *See* Dietmar Harhoff et al., *Citations, Family Size, Opposition and the Value of Patent Rights*, 1596 RES. POL'Y 1, 8 (2003).

132. All dates obtained from Google's patent files. GOOGLE, https://patents.google.com (last visited Dec. 18, 2017).

133. Prosecution length was calculated by subtracting the filing date from the grant date (note that this does not account for continuations).

134. NBER's technology classification was obtained from the USPTO's PatentsView search tool. *Data Download Tables*, USPTO PATENTSVIEW, http://www.patentsview.org/download (last updated Aug. 8, 2017).

135. Maintenance data was obtained from a USPTO bulk download file hosted by Reed Tech. *Patent Maintenance Fee Events (September 1981–Present)*, REED TECH, http://patents.reedtech.com/maintfee.php (last visited Dec. 18, 2017).

136. This was done by comparing text of applications to the text of the granted patent; both were obtained from the USPTO website. *Search for Patents*, USPTO, https://www.uspto.gov/patents-application-process/search-patents#heading-2 (last visited Dec. 18, 2017). Full code on file with the author.

137. The random sample was generated by taking a random selection of application numbers from the replication data file for Michael D. Frakes & Melissa F. Wasserman, *Replication Data for: "Is the Time Allocated to Review Patent Applications Inducing Examiners to Grant Invalid Patents?: Evidence from Micro-Level Application Data,"* DATAVERSE (Dec. 11, 2015) https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/ABE7VS.

138. *Id.*

139. 35 U.S.C. § 101 (2012) (allowing rejection for lack of utility or lack of patentable subject matter).

140. *Id.* § 102 (rejection for anticipation).

141. *Id.* § 103 (rejection for obviousness).

142. *Id.* § 112(a) (rejection for lack of written description, enablement, or best mode).

143. *Id.*.§ 112(b) (rejection for indefiniteness).

3. Synonyms

Data used for the initial analysis captures only verbatim repetition of a claim word in the specification, not use of synonyms. Verbatim repetition was chosen for the primary level of analysis because the use of synonyms is discouraged in patent drafting.[144] Thus, counting frequency with which an element recurs in the specification is unlikely to significantly understate the amount of information provided by a patent. However, in order to verify this assumption, the methodology was repeated with an algorithm designed to capture synonyms of claim words.[145]

The synonym-capturing algorithm likely overestimates the number of synonyms, and thus underestimates the amount of ancillary language. This occurs because the algorithm seeks synonyms across all senses of a word, many of which may not be true synonyms of the claim word.[146] Thus the true difference between the results when accounting for synonyms and when not accounting for synonyms is likely overstated due to methodological limitations.

Thus, the default algorithm does *not* include synonyms; unless otherwise specified, all figures and calculations do not include synonyms. Note that although including synonyms reduces the amount of ancillary language (see Figure 3), it does not change the directionality of any results.

B. RESULTS

Each of the 25,000 patents in the sample was analyzed to determine how many times words appearing in the claims appeared in the specification. Ancillary claim language is enormously prevalent in many patents: approximately one quarter of all claim language is ancillary. This Subpart begins with an example to illustrate the use of ancillary claim language and then proceeds to a quantitative analysis of the data. Summary statistics and a regression analysis of all data collected are available in Appendices A and B.

U.S. Patent No. 7,776,844 is an example of a patent with a significant amount of ancillary claim language. The patent claims a composition (related

144. PRACTISING LAW INSTITUTE, ADVANCED PATENT PROSECUTION WORKSHOP 2013: CLAIM DRAFTING & AMENDMENT WRITING, 2013 (Jay P. Lesser et al. ed., 2013). *See also* Phillips v. AWH Corp., 415 F.3d 1303, 1314 (Fed. Cir. 2005) ("[C]laim terms are normally used consistently throughout the patent . . . ."); Markman v. Westview Instruments, Inc., 52 F.3d 967, 979 (Fed. Cir. 1995) ("For claim construction purposes, the description [in the specification] may act as a sort of dictionary, which explains the invention and may define terms used in the claims.").

145. To identify synonyms, the algorithm uses WordNet, a lexical database developed by Princeton University. *WordNet: A Lexical Database for English*, PRINCETON UNIV., https://wordnet.princeton.edu (last visited Dec. 18, 2017).

146. For example, if a claim directed to methods of organizing a law school curriculum included the word "course," the algorithm would search the specification for "course" and synonyms including "class," "line," "trend," "path," and "track." While "class" is a true synonym of "course" as it is used in the context of the claim, "line" is not.

to the herbicide Roundup®)[147] "for treatment of a cosmetic condition or dermatological disorder."[148] The patent further claims the composition used in combination with an agent

> selected from the group consisting of: abacavir, abciximab, acamprosate, acarbose, acebutolol, acetaminophen, acetaminosalol, acetazolamide, acetic acid, acetohydroxamic acid, N-acetylcysteine and its esters, N-acetylglutathione and its esters, acitretin, aclometasone dipropionate, acrivastine, acthrel, actidose, actigall, acyclovir, adalimumab, adapalene, adefovir dipivoxil, adenosine, agalsidase, albendazole, albumin, albuterol, aldesleukin, alefacept, alemtuzumab, alendronate, alfuzosin, alitretinoin, allantoin, allium, allopurinol, alloxanthine, almotriptan, alosetron, alpha tocopheral, alpha$_1$-proteinase, alprazolam, alprenolol, alprostadil, alteplase, altretamine, aluminum acetate, aluminum chloride, aluminum chlorohydroxide, aluminum hydroxide, amantadine, amifostine, amiloride, aminacrine, amino acid, aminobenzoate, p-aminobenzoic acid, aminocaproic acid, aminohippurate, aminolevulinic acid, aminosalicylic acid, amiodarone, amitriptyline, amlodipine, amocarzine, amodiaquin, amorolfine, amoxapine, amoxicillin, amphetamine, amphotericin, ampicillin, amprenavir, anagrelide, anakinra, anastrozole, anisindione, anthralin, antihemophilic, antithrombin, anti-thymocyte, antivenin, apomorphine, aprepitant, aprotinin, arbutin, argatroban, aripiprazole, arnica, ascorbic acid and its esters, ascorbyl palmitate, aspirin, atazanavir, atenolol, atomoxetine, atorvastatin, atovaquone, atropine, azathioprine, azelaic acid, azelastine, azithromycin, baclofen, bacitracin, balsalazide, balsam, basiliximab, beclomethasone dipropionate, bemegride, benazepril, bendroflumethiazide, benzocaine, benzonatate, benzophenone, benzoyl peroxide, benztropine, bepridil, beta carotene, betamethasone dipropionate, betamethasone valerate, betaxolol, bethanechol, bevacizumab, bexarotene, bicalutamide, bimatoprost, bioflavonoids, biotin, biperiden, bisacodyl, bisoprolol, bivalirudin, bortezomib, bosentan, botulinum, brimonidine, brinzolamide, bromocriptine, brompheniramine, budesonide, bumetanide, bupivacaine, buprenorphine, bupropion, burimamide, buspirone, busulfan, butabarbital, butalbital, butenafine, butoconazole, butorphanol, butyl aminobenzoate, cabergoline, caffeic acid, caffeine, calcipotriene, calcitonin-salmon, calcitriol, calfactant, camellia sinensis, camphor, candesartan cilexetil, capecitabine, capreomycin, capsaicin, captopril, carbamazepine, carbamide peroxide, carbidopa, carbinoxamine,

147. *Glyphosate and Roundup Brand Herbicides*, MONSANTO (May 16, 2017), http://www.monsanto.com/glyphosate/pages/default.aspx.

148. U.S. Patent No. 7,776,844 col. 31 l. 49–50 (filed Apr. 23, 2009) (issued Aug. 17, 2010).

cefditoren pivoxil, cefepime, cefpodoxime proxetil, celecoxib, cetirizine, cevimeline, chitosan, chlordiazepoxide, chlorhexidine, chloroquine, chlorothiazide, chloroxylenol, chlorpheniramine, chlorpromazine, chlorpropamide, ciclopirox, cilostazol, cimetidine, cinacalcet, ciprofloxacin, citalopram, citric acid, cladribine, clarithromycin, clemastine, clindamycin, clioquinol, clobetasol propionate, clocortolone pivalate, clomiphene, clonidine, clopidogrel, clotrimazole, clozapine, coal tar, coal tar extracts (LCD), codeine, cromolyn, crotamiton, cyclizine, cyclobenzaprine, cycloserine, cytarabine, dacarbazine, dalfopristin, dapsone, daptomycin, daunorubicin, deferoxamine, dehydroepiandrosterone, delavirdine, desipramine, desloratadine, desmopressin, desoximetasone, dexamethasone, dexmedetomidine, dexmethylphenidate, dexrazoxane, dextroamphetamine, diazepam, diclofenac, dicyclomine, didanosine, dihydrocodeine, dihydromorphine, diltiazem, 6,8-dimercaptooctanoic acid (dihydrolipoic acid), diphenhydramine, diphenoxylate, dipyridamole, disopyramide, dobutamine, dofetilide, dolasetron, donepezil, dopa esters, dopamide, dopamine, dorzolamide, doxepin, doxorubicin, doxycycline, doxylamine, doxepin, duloxetine, dyclonine, econazole, efalizumab, eflomithine, eletriptan, emtricitabine, enalapril, ephedrine, epinephrine, epinine, epirubicin, eptifibatide, ergotamine, erythromycin, escitalopram, esmolol, esomeprazole, estazolam, estradiol, etanercept, ethacrynic acid, ethinyl estradiol, etidocaine, etomidate, famciclovir, famotidine, felodipine, fentanyl, ferulic acid, fexofenadine, flecamide, fluconazole, flucytosine, fluocinolone acetonide, fluocinonide, 5-fluorouracil, fluoxetine, fluphenazine, flurazepam, fluticasone propionate, fluvoxamine, formoterol, furosemide, galactarolactone, galactonic acid, galactonolactone, galantamine, gatifloxacin, gefitinib, gemcitabine, gemifloxacin, glucarolactone, gluconic acid, gluconolactone, glucuronic acid, glucuronolactone, glycolic acid, griseofulvin, guaifenesin, guanethidine, N-guanylhistamine, haloperidol, haloprogin, hexylresorcinol, homatropine, homosalate, hydralazine, hydrochlorothiazide, hydrocortisone, hydrocortisone 21-acetate, hydrocortisone 17-butyrate, hydrocortisone 17-valerate, hydrogen peroxide, hydromorphone, hydroquinone, hydroquinone monoether, hydroxyzine, hyoscyamine, hypoxanthine, ibuprofen, ichthammol, idarubicin, imatinib, imipramine, imiquimod, indinavir, indomethacin, infliximab, irbesartan, irinotecan, isoetharine, isoproterenol, itraconazole . . . [the claim continues for an additional 360 words].[149]

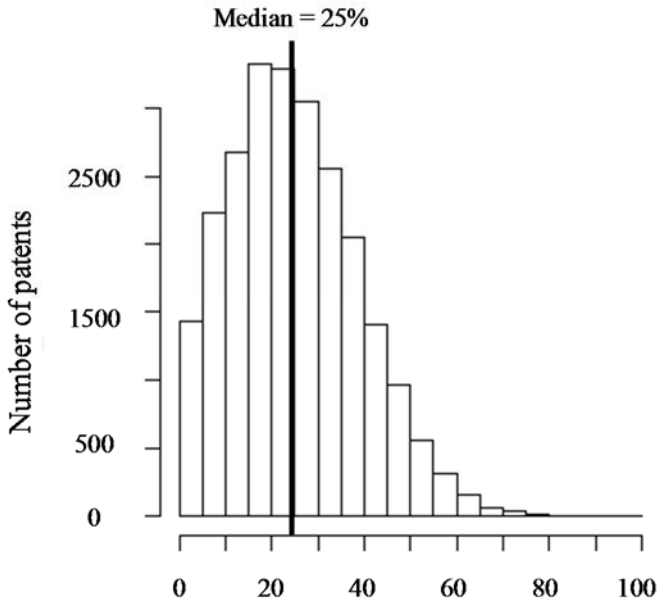
149. *Id.* at col. 35 l. 21 to col. 38 l. 2.

Of the 662 drugs listed in this claim, none are described in examples in the specification[150] and all are examples of ancillary language. To be clear, this is not necessarily a bad patent—the technology described by the patent may be very useful and beneficial to society. The problem with this patent is that the exceedingly large number of pharmaceuticals recited in the claims are not part of the core invention, were not tested with the core invention,[151] and may not actually work as described with the core invention.

1. Prevalence

Figure 1, below, is a histogram showing the prevalence of ancillary language across the sample, as a percentage of all analyzed words in the claims. A median of 25% of claim language is ancillary. Given that patent claims are supposed to be exclusively about the invention and entirely supported in the specification, this is unexpectedly high.

Figure 1: Histogram showing percentage of ancillary claim words relative to total number of words in claims, per patent



Next, for each patent, I used the algorithm described above to determine the percentage of claim words that appeared a given number of times in the

150. *See id.* at col. 19 l. 11 to col. 31 l. 42.

151. Or at least, no testing is recorded in the patent.

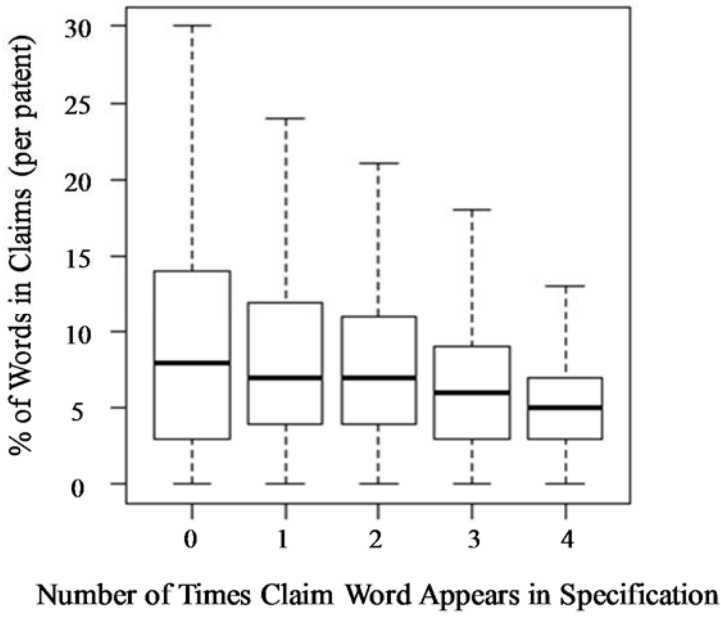
specification of that patent. Figure 2 shows the data for claim words appearing between zero to four times in the specification of the corresponding patent. This breaks down the 'ancillary' measure into its constituent bins. A median of 8% of claim words simply did not appear in the specification at all, as represented by the red bar in the histogram below.[152] This is a surprisingly large number, given that patentees are legally required to describe the claimed invention in the specification with sufficient detail to show that the inventor was in fact in possession of the invention.[153] It is also surprising because many patent drafting guides recommend that prosecutors copy and paste the words of the claims verbatim into the specification in order to ensure that all claims are mentioned in the specification.[154] A further 7% of claim terms appear only once in the specification, and the same number appear only twice. This means that many words in patent claims are described little, if at all, in the specification.

152. The interquartile range is 3–15%.

153. LizardTech, Inc. v. Earth Res. Mapping, Inc., 424 F.3d 1336, 1345 (Fed. Cir. 2005).

154. *E.g.*, ADRIANA L. BURGY, FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER, LLP, DRAFTING THE PATENT SPECIFICATION 7-3 (2009), http://www.aipla.org/learningcenter/library/papers/bootcamps/2009patentbootcamp/Documents/Burgy_Paper.pdf ("The easiest way to [write the specification] is to block copy the claims with your word processing program. Then paste them into your specification document twice.").

Figure 2:[155] Percentage of claim words appearing zero to four times in specification, relative to all words in claims of the corresponding patent



Because textual analysis can be imprecise and this methodology underestimates ancillary language, the figures should not be taken as an exact quantification of the phenomenon. Rather, the figures and numbers below indicate that ancillary language is pervasive and that patent claims do much more than merely describe the invention. This Article defines ancillary language as words appearing two times or fewer in the specification. This definition is somewhat arbitrary but, as can be seen from Figure 2,[156] the main finding (that there is a large amount of ancillary language in patent claims) would be the same even if ancillary language were defined as words appearing one time or fewer (or three times or fewer).

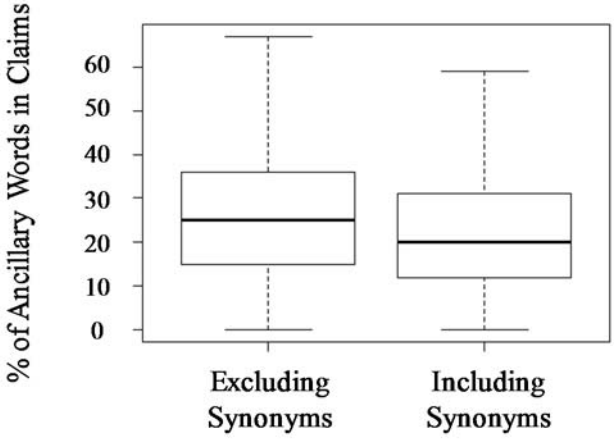
Figure 3 shows the percentage of ancillary claim words relative to the total number of words in claims in the sample under two alternate algorithms:

155. All boxplots in this Article were made using the default settings in R. The thick black line indicates the median value, and the upper and lower edges of the box indicate the 75% and 25% percentile values, respectively. The whiskers extend to 1.5 times the length of the box. Outliers are omitted using the function "outline=FALSE". Martin Mächler, *Box Plots*, ETH ZURICH, https://stat.ethz.ch/R-manual/R-devel/library/graphics/html/boxplot.html (last visited Dec. 18, 2017).

156. All boxplots in this Article were made using the default settings in R. The thick black line indicates the median value, and the upper and lower edges of the box indicate the 75% and 25% percentile values, respectively. The whiskers extend to 1.5 times the length of the box. Outliers are omitted using the function "outline=FALSE." *Id.*

excluding and including synonyms. The box on the left shows the frequency of ancillary language in the specification when synonyms are not counted in the analysis. Excluding synonyms is the default for the analysis and is used for all graphs below. The box on the right shows the frequency of ancillary language in the specification when the algorithm searches for both verbatim matches of a claim word in the specification and synonyms of the claim word. Thus, 20% of words in patent claims are not present more than twice in the specification either verbatim or as synonyms. This compares to 25% of words that are not in the specification verbatim more than twice.

Figure 3: Ancillary Language in Specification, excluding and including synonyms



2. Industry

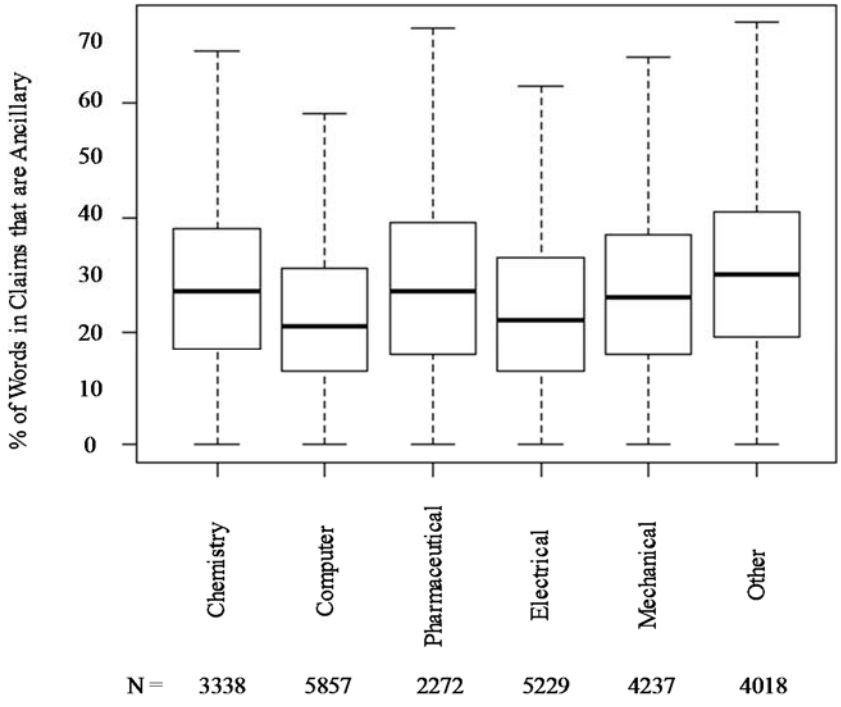
Industry matters for patents. Although patent law is facially industry-neutral, in practice doctrine is applied in industry-specific ways.[157] As a result, patent drafting conventions differ across industries.[158] Thus, ancillary language may be used in different ways and have different consequences depending on the industry involved in the patented technology. The figures below show that different industries have somewhat varying rates of ancillary language. Section IV discusses how there are varying concerns associated with ancillary language depending on the industry.

157. *See* Burk & Lemley, *supra* note 2, at 1796.

158. *See, e.g.*, ROBERT C. FABER, FABER ON MECHANICS OF PATENT CLAIM DRAFTING § 4 (6th ed. 2014) (including separate subsections for method claims directed to chemical processes, electrical methods, business methods, and software).

Figure 4 shows the percentage of words in a patent's claims that are ancillary, divided by industry. Ancillary language is somewhat higher in the chemical, pharmaceutical, mechanical, and "other" industries,[159] as compared to the computer and electrical industries, though the magnitude of the difference is not large. Differences between all industries are significant at $p<0.05$ except the pair chemistry/pharmaceutical, where the difference in percent of ancillary language is not significant. Figure 5 further divides industries into sub-categories and shows the percent of words in a patent's claim that are ancillary, divided by sub-industry (in the figure, sub-industries are grouped by industry, beginning with chemical, followed by computer, pharmaceutical, electrical, mechanical, and other).[160]

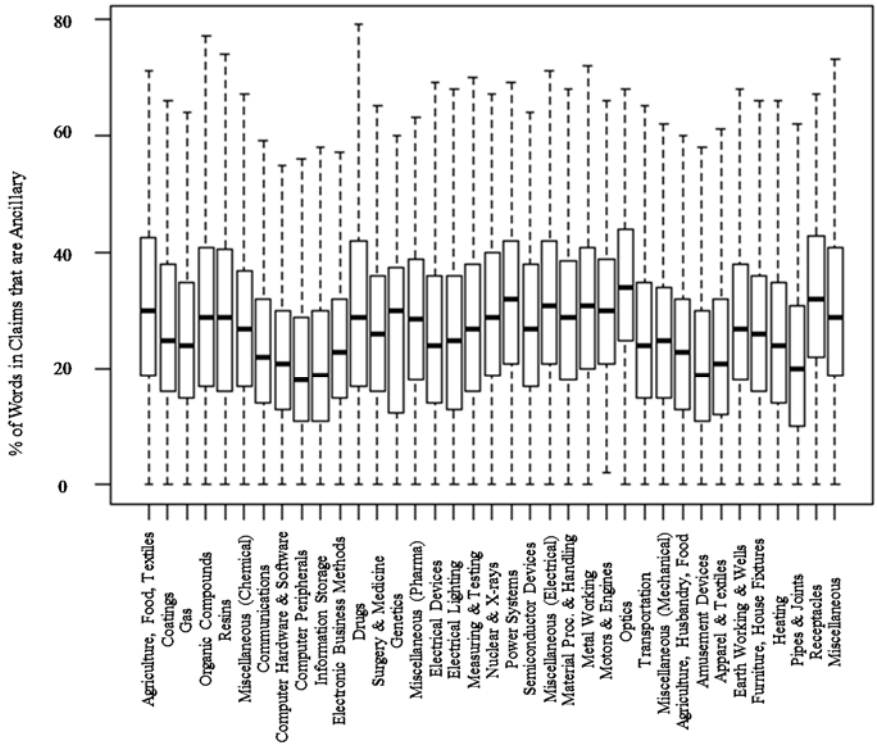
Figure 4: Ancillary Language by Industry



159. "Other" includes the subcategories "Agriculture, Husbandry, Food," "Amusement Devices," "Apparel & Textile[s]," "Earth Working & Wells," "Furniture, House Fixtures," "Heating," "Pipes & Joints," "Receptacles," and "Miscellaneous." *Data Download Tables*, PATENTSVIEW, http://www.patentsview.org/download (last updated Aug. 8, 2017) (listing categories from the NBER industry and subindustry classification file).

160. *Id.*

Figure 5: Ancillary Language by Sub-Industry



3. Application Characteristics

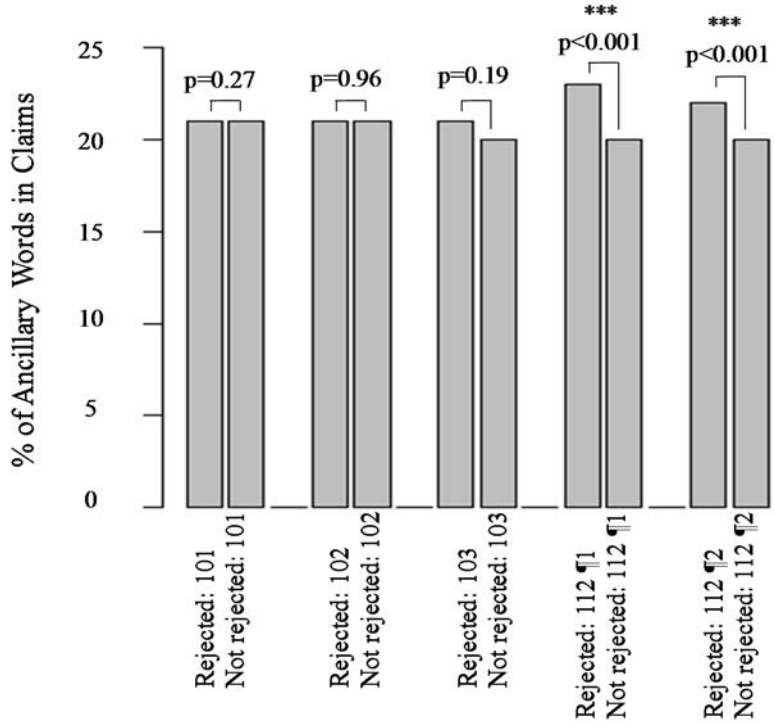
Not all patent applications become granted patents. This section examines ancillary language in a sample of 15,000 randomly selected applications and includes applications that were never granted.[161]

Applications with more ancillary language fare differently during examination. Applications rejected under 35 U.S.C. § 112 have more ancillary language than applications not rejected under § 112. Section 112 ¶ 1 rejections are for lack of enablement or written description (where language in the claims is not adequately supported by the specification) and § 112 ¶ 2 rejections are for indefiniteness (where claim language is incomprehensible). Although the applications may differ in ways other than the amount of ancillary language, and these other differences may contribute to the divergent rejections, ancillary language is a plausible cause of the rejections.

161. For a discussion of how the sample was obtained, see *supra* Part III.A.

It stands to reason that patents with a great deal of ancillary language would be more likely to be rejected for these reasons, as the claim language is neither supported nor defined in the specification. Indeed, the examiners' stated reasons for the rejections support this. For example, one examiner complained that a claim "recites the term 'unique electronic fingerprint' . . . what is a unique electronic fingerprint?" and then rejected the application under § 112 ¶ 2.[162] The term "unique electronic fingerprint" in this situation is ancillary language, as it is not mentioned anywhere in the specification.[163] Interestingly, there is no difference in ancillary language between applications rejected on non-112 grounds and their non-rejected counterparts. This is likely because these grounds for rejection have less to do with the language in the claim.

Figure 6: Rejections During Prosecution



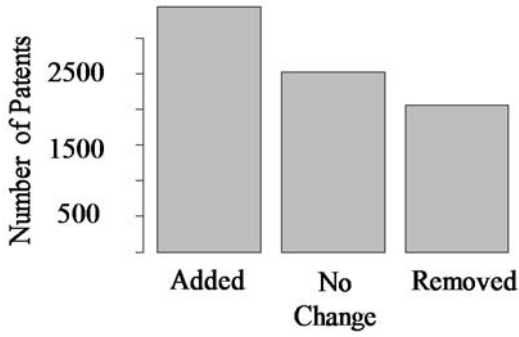
162. U.S Patent Application No. 2007/0258894, Non-Final Rejection 3 (Feb. 26, 2010).

163. U.S. Patent Application 10/588,749, Publication No. 2007/0258894 (published Nov. 8, 2007) (Richard J. Melker et al., applicant).

4. Examination Characteristics

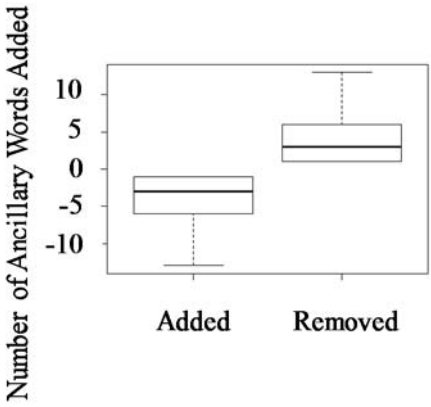
In many patents, the amount of ancillary language in the claims changes during the prosecution process—the result of claim amendments over the course of examination before the Patent and Trademark Office. The data described in this section are based on the applications that eventually became the patents used in this study, and thus includes only granted patents.[164] The average change in the number or percent of ancillary terms is quite small (-0.2 words and 1.46%, respectively), but, because many applications added ancillary words while many others removed ancillary words during examination, the average belies the true scope of the effect. Figure 7 shows that a plurality of patents (42%) increased the amount of ancillary language during prosecution while many (28%) experienced no change or a decrease in ancillary language (30%). Figure 8 shows the magnitude and variability of the change in ancillary language during prosecution for those patents where the amount of ancillary language increased or decreased.

Figure 7: Directionality of Change in Ancillary Language During Prosecution



164. Because the full text of patent applications is available electronically only for applications filed after 2001, a total of 8,014 of the sampled patents had applications available for comparison.

Figure 8: Magnitude of Addition/Removal of Ancillary Language During Prosecution



When prosecution increased the number of ancillary terms, it was often in response to a rejection on the grounds that the claim was indefinite or overly broad. In response, the applicants added ancillary elements to narrow or define a claim. For example, Application No. 10/097,113, claiming a labeling compound that can be attached to nucleic acids to monitor gene expression,[165] was filed with an independent claim directed to a novel compound linked to a "detectable moiety."[166] The examiner repeatedly rejected this claim on the ground that it was indefinite.[167] The examiner complained, "detectable how? . . . Is NMR included? How about IR? . . . what detection limit is required to make the group detectable?"[168] The examiner suggested either amending the specification or adding specific compounds to claim 1.[169] The applicants chose the latter, adding a list of potential compounds including "a ra[d]iolabel, a magnetic particle, colloidal gold, fluorescein, texas red, rhodamine" and almost 200 other compounds (most instances of ancillary language, as the specification does not discuss the compounds).[170] The examiner then granted the patent.[171]

165. U.S. Patent Application No. 10/097,113, Publication No. 20020182625 ¶ [0012] (published Dec. 5, 2002) (Glenn McGall & Anthony Barone, applicants).

166. *Id.* at claim 1.

167. *E.g.*, U.S Patent Application No. 2007/097,113, Non-Final Rejection 7 (Jan. 7, 2003).

168. *Id.*

169. *Id.*

170. U.S. Patent No. 7,468,243 claim 1 (filed Mar. 12, 2002) (issued Dec. 23, 2008).

171. *Id.* In a similar instance, an examiner suggested adding examples of "surface stabilizers" to a claim previously directed to the category in general in order to overcome a written

In many instances, ancillary language was removed during prosecution, either by amending a claim to remove the language or by cutting the claim out entirely. Take, for example, Application No. 10/156,744, directed to a method of illuminating objects to improve digital capture and analysis of said objects.[172] As filed, the application contained many claims directed to uses for this method, with little description in the specification of how these applications would actually be implemented. For instance, claim 262 describes a "method of and apparatus for securing an airport" including steps such as "detecting suspicious conditions revealed by x-ray images of baggage" and "running intelligent information processing algorithms [on] each passenger and baggage attribute record . . . in order to detect any suspicious conditions which may given [sic] concern or alarm" and, based upon these steps, "determining if a breach of security appears to have occurred."[173] The specification contains little detail about how these steps would be implemented, despite their complex nature, with the inventors merely noting that the algorithms needed for the steps "are within the knowledge of those skilled in the art."[174] The claim is, therefore, full of ancillary language. During the examination process, the applicants removed this claim as well as many other claims containing ancillary language, resulting in a significant reduction of ancillary language during prosecution, and a granted patent with claims more focused on the core invention.[175]

5. Patent and Claim Characteristics

i. Specification Length

Ancillary language correlates with certain aspects of how the patent is drafted, in particular the number of unique words in the claims and length of specification. Patents with more ancillary language have wordier claims, which is not surprising, because more words provide more opportunities for ancillary language. Similarly, patents with more ancillary language have shorter specifications, which is again not surprising because patents with shorter specifications will not have the opportunity to describe as many features.

---

description rejection. U.S Patent Application No. 10/444,066, Applicant Arguments/Remarks Made in an Amendment 1 (Dec. 21, 2005). The applicant did so, adding a list of several hundred (ancillary) examples of surface stabilizers. U.S. Patent No. 7,276,249 col. 53–58 (filed May 23, 2003) (issued Oct. 2, 2007).

172. U.S. Patent Application No. 10/156,744, Publication No. US 7,104,455 B2, at [1] (published Sept. 12, 2006) (Constantine J. Tsikos et al., applicants).

173. *Id.* at claim 262(k)–(m).

174. *Id.* at [1525].

175. *See* U.S. Patent No. 7,104,455 (issued Sept. 12, 2006).

ii. Independent and Dependent Claims

Overall, slightly more than half of ancillary language comes from independent claims (a median of 57% across the entire data set). This varies across industries: from a low of 50% in pharmaceutical patents to a high of 60% of ancillary language from independent claims in mechanical patents. Only ancillary language in independent claims narrows the overall breadth of the patent,[176] so ancillary language affects overall scope in more than half of all patents. Note that ancillary words in independent claims may also appear in dependent claims.

iii. Value Indicators

Because patentees include ancillary language in the text of patent claims, they presumably perceive some benefit from this increase. Thus, ancillary language increases the value of patents in some way or at least increases the perceived value of patents. It is not obvious from the empirical analysis why this is. Patents with the most ancillary language perform poorly on proxies for value such as number of forward citations[177] and payment of maintenance fees (data in Appendices A and B).[178] Because patents with the most ancillary language have fewer forward citations (controlled for year of grant) and lower payment of maintenance fees, these patents might be thought to be of lower value than patents with less ancillary language.

However, the analysis in this study cannot account for differences in value between a patent without ancillary language and the same patent after adding ancillary language. It may be that patents with ancillary language have some other feature that renders them less valuable, and that ancillary language recoups some of the lost value. Additionally, one explanation consistent with lower citation counts is that patents with more ancillary language are blocking downstream development in the area covered by the patent, meaning that other innovators are not creating new technologies that would cite the patent.[179]

176. *See supra* Part I.B.

177. The number of forward citations received by a patent is thought to be a function of the value and importance of the invention to its field. *See, e.g.*, Bronwyn H. Hall et al., *Market Value and Patent Citations*, 36 RAND J. ECON. 16, 17 (2005).

178. The USPTO requires patentees to pay periodic maintenance fees to prevent their patents from expiring. *Maintain Your Patent*, USPTO, http://www.uspto.gov/patents-maintaining-patent/maintain-your-patent (last updated July 27, 2017, 12:02 PM). Because these fees are not negligible, they are a proxy for the value of the patent to its owner. *See id.* Only patents with a value greater than the maintenance fee will be maintained. *See id.*

179. Jonathan H. Ashtor, Does Patented Information Promote Progress? 9–11 (June 22, 2017) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2857697.

iv. Time

The amount of ancillary language has been decreasing over time.[180] It is not clear why this is the case. It may be the result of increasingly longer specifications that presumably use more unique words.[181] Note that this does not necessarily mean that the problem posed by ancillary language is diminished, because the reduction may simply be due to the practice of copy-and-pasting claim language into the specification more times or defensively including words in the specification without a clear purpose.[182] Alternatively, it may be in response to changing written description jurisprudence.[183]

C. VALIDATING THE METHODOLOGY

This Article uses a novel methodology that has not been employed elsewhere. In order to support the results presented above, I validated the methodology through (1) manual review and (2) automated review by a different method. Both tests support the validity of the methodology.

I first manually reviewed the ancillary terms generated by the algorithm for 200 randomly selected patents. 173 terms were indeed ancillary. Of the remaining terms, 11 were in fact in the specification but were not found by the algorithm because the claim term contained a typo. The remaining 16 terms were not language that should have been included in the study either because the terms were legal terminology (for example, "thereon") or because they were formatted slightly differently in the specification and claims (for example "time-constant" vs. "time constant"). Note that my manual review accounted only for language erroneously identified as ancillary because it appeared verbatim in the specification. The review did not check for either language erroneously identified as central or for language identified as ancillary that was not in the specification verbatim but was in the drawings or was described in a comparable sense in the specification.

Second, I validated the methodology using an external algorithm. Patent prosecutors are increasingly using drafting software with functionalities that check whether the claims are supported by the specification. These functionalities differ among software programs, but are generally intended to help prosecutors avoid rejections for lack of enablement or written description. I ran a random sample of 1000 patents through Microsystem's Patent Companion software.[184] I used the "Unsupported Terms" functionality

180. *See infra* Appendix B.

181. Dennis Crouch, *The Rising Size and Complexity of the Patent Document* (Univ. of Mo. Sch. of Law, Legal Studies Research Paper No. 2008-04, 2008), http://ssrn.com/abstract=1095810.

182. *See* Kristen Osenga, *The Shape of Things to Come: What We Can Learn from Patent Claim Length*, 28 SANTA CLARA COMPUTER & HIGH TECH. L.J. 617, 623 (2012).

183. *See* Ariad Pharm., Inc. v. Eli Lilly & Co., 598 F.3d 1336, 1344 (Fed. Cir. 2010).

184. *Patent Companion*, MICROSYSTEMS, https://www.microsystems.com/products/legal/patent-companion (last visited Dec. 18, 2017).

to flag "terms and phrases in the [c]laims section . . . . that are not supported by the [s]pecification[]."[185] The precise methodology used by Patent Companion is proprietary, but it is apparent that the software uses an algorithm different from that of this Article. Specifically, Patent Companion divides the claims into terms (rather than words) and does not restrict itself to nouns. The average number of ancillary words by my methodology was 14.6 as compared to 11.4 terms for Patent Companion.

Because the software uses a different algorithm, there is no reason to expect that the results will be identical to the results of this Article. However, as the results are within the same order of magnitude, it strengthens both the results herein and the legal implications presented below. For the former, a correspondence between the measures suggests that ancillary language is indeed highly prevalent.[186] For the latter, that Patent Companion flags a great deal of language in the claims unsupported by the specification suggests that we ought to be concerned about the contents of the claims.

IV. DISCUSSION AND IMPLICATIONS

Claims were created to distinctly point out the invention, resulting in a belief that claims are synonymous with the invention and that structuring a patent with separate specification and claim sections would isolate the old from the new. The results from this study suggest that the relationship between claim and invention and claim and specification is far more complex than previously thought. This Section asks how these results contribute to our understanding of the practical, theoretical, and policy issues surrounding patent claims.

Section A explores why clutter exists. It is not clear from the empirical data that it increases patent value, yet it must benefit patentees in some way. This section closely analyzes selected examples of ancillary language and lays out a typology of incentives for ancillary language that are consistent with the empirical results.

Section B returns to the problems with patent claims summarized in Part II.B. Although claims are well known to be dysfunctional and reform attempts have grown common, the problems persist. This section suggests that many of the common problems with claims can be partially explained by the prevalence of ancillary language and provides examples of how ancillary language clutters claims in challenging ways.

This deeper understanding of the causes of problems with claims can aid reform efforts. Section C begins with theoretical reforms and questions how theorists should understand claims, since the relationship between claims and

185. *Patent Companion – How it Works: Unsupported Terms*, MICROSYSTEMS (Feb. 17, 2016, 9:00 AM), http://blog.microsystems.com/blog/patent-companion-how-it-works-unsupported-terms-0.

186. Though the Patent Companion is not explicitly measuring ancillary language, but rather something somewhat similar.

the invention is not straightforward. It proposes a framework for understanding when ancillary language is desirable and when it is problematic. Section D then introduces policy proposals informed by the evidence of ancillary language and new theoretical framework.

A. *UNDERSTANDING CLUTTER*

As an initial matter, it is important to consider why patentees add ancillary language to claims and how this adds value to the patent. Oddly, the value proxies measured for this study suggest that patents with more ancillary language are *less* valuable than patents with less ancillary language.

However, informal discussions with patent attorneys and patent examiners revealed a host of other factors that are hard to quantify but may be driving the use of ancillary language. The discussion below explores why ancillary language may be valuable to patentees. In particular, patentees may choose to include ancillary language for signaling purposes, such as increasing the audience for their patent or advertising the functionality of their core invention, or for decoy purposes, such as obscuring the core invention in a flurry of ancillary language. Alternatively, patentees may include ancillary language in order to "wear out" the patent examiner or as a "back-up" to hedge their bets against uncertain future prosecution, litigation, and commercial uncertainty more generally. For any of these uses, ancillary language allows patentees to increase the amount or types of information conveyed by the patent, thereby increasing its value.

1. Signaling

Ancillary language may increase patent value by providing an informational benefit that is separate from any substantive legal effect of the language. The concept that, "[u]nder some circumstances, the informational function of patents may be more valuable to the rights holder than the substance of the rights," has been termed "patent signaling."[187] Patent signaling has been explored extensively in both legal and management literature,[188] but the focus has always been on the entirety of the patent, rather

187. Clarisa Long, *Patent Signals*, 69 U. CHI. L. REV. 625, 625 (2002); *see also id.* at 643–55 (describing the informational function of patents).

188. *See, e.g.*, *id.*; Michael Abramowicz, *The Danger of Underdeveloped Patent Prospects*, 92 CORNELL L. REV. 1065, 1074 (2007) (positing that a reason why some patents go unlicensed is because they serve a signaling function); Jeanne C. Fromer, *A Psychology of Intellectual Property*, 104 NW. U. L. REV. 1441, 1447 (2010) (discussing patent signaling theories); David H. Hsu & Rosemarie H. Ziedonis, *Patents as Quality Signals for Entrepreneurial Ventures*, 2006 ACAD. MGMT. BEST PAPER PROC. 1, 1; Gregory N. Mandel, *Proxy Signals: Capturing Private Information for Public Benefit*, 90 WASH. U. L. REV. 1, 3–8 (2012). *See generally* David B. Audretsch et al., *Financial Signaling by Innovative Nascent Ventures: The Relevance of Patents and Prototypes*, 41 RES. POL'Y 1407 (2012) (describing the use of patent signaling as a method of attracting investors); Timothy R. Holbrook, *The Expressive Impact of Patents*, 84 WASH. U. L. REV. 573 (2006) (discussing patents in

than language in the claims. Nevertheless, theories of signaling apply well to ancillary language in patent claims.

Including ancillary information in the claims, rather than in the specification, increases the likelihood that others will find the patent.[189] Some types of patent searches search only the claims, not the specification.[190] Another signaling benefit is communication with potential infringers. A survey of patent attorneys found that including additional elements in a patent claim "is also useful as a way for potential licensees to know that the proposed product is literally and explicitly covered."[191] Although ancillary language does not increase the scope of a patent claim, and therefore is not itself responsible for creating the claim's coverage of the proposed product, it puts potential infringers on notice that the patentee is on the lookout for products with that particular combination.

Language in claims can also signal information to users, researchers, or investors. This function has been examined in the context of the specification.[192] Claims can similarly be used to highlight potential uses of the core invention or complementary products, a sort of advertising function.[193] For example, a patent on fire-extinguishing microcapsules that can be incorporated into a coating for various surfaces includes in the claims a recitation that it can be used on articles such as "textile fabric," "pillow[s]," "pen[s]," "furniture," "packaging," "printer component[s]," "fuel pump[s]," "disc drive[s]," "vehicle console component[s]," and "bellows."[194]

Alternatively, ancillary language in patent claims may serve a "decoy" function to mislead competitors.[195] In the context of patents as a whole, companies commonly patent numerous inventions or mechanisms in order to hide the "one good one in a flood of bad inventions."[196] In the context of

---

the context of social signaling); Sean B. Seymore, *The Teaching Function of Patents*, 85 NOTRE DAME L. REV. 621 (2010) (discussing how to use patents as sources of technical information).

189. Note that being found by searches is not considered advantageous by all patentees, and indeed, many patentees go to some lengths to hide their patents from searches. *See, e.g.*, David Russo et al., *Functional-Based Search for Patent Technology Transfer*, *in* 2 ASME 2012 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference 529, 530 (2012) ("[S]ometimes patentees write patents using very general terms for hiding their contents or extending their validity.").

190. For example, a clearance search.

191. Dennis Crouch, *Theory of Dependent Claims: Survey Results*, PATENTLY-O (May 22, 2008), https://patentlyo.com/patent/2008/05/theory-of-depen.html.

192. Long, *supra* note 189, at 647.

193. *Id.* ("Even if patents conferred no protection, firms might find it desirable to obtain them as a means of credibly advertising their inventions.").

194. U.S. Patent No. 8,465,833 B2 col. 15 l. 14–26 (filed June 18, 2013).

195. There is evidence that patents as a whole are used this way. *See, e.g.*, Corinne Langinier, *Using Patents to Mislead Rivals*, 38 CANADIAN J. ECON. 520, 522 (2005) ("There is considerable evidence that firms use 'decoy patents' to direct competitors into unprofitable fields of research.").

196. *Id.*

ancillary claim language, companies might include claims directed to many different possible embodiments in order to hide their true preference.

As compared to signaling in the specification, signaling in the claims is harder to achieve due to the strict constraints on claim format.[197] However, there are several advantages to including information in the claims, rather than (or in addition to) the specification of a patent. One of the advantages of using claims to signal information is that "[t]he information contained in a patent is at least minimally credible"[198] both because it is malpractice to misrepresent information before the PTO[199] and because the patent has been reviewed by a patent examiner (a subject area expert) who is charged with rejecting inventions that would not be considered credible by a person skilled in the art.[200] While these benefits apply to all sections of a patent, they apply more strongly to the claims than to the specification. This is because the patent examiner assesses the credibility of "the claimed invention"[201]—the invention described in the claims, not in the specification—and makes rejections based on the contents of the claims.

2. Insurance

Other uses for ancillary language include insurance functions to hedge against some of the uncertainty of the patent application and litigation processes. It is well established that some language found in dependent claims[202] serves functions such as clarifying the language of other claims through the principle of claim differentiation,[203] or highlighting a key embodiment.[204] Because it is difficult at the drafting stage to predict what

197. *See* MPEP, *supra* note 13, § 608.01(m). As seen from the example above, although claims are restricted to one sentence, this sentence can be very long. *See supra* text accompanying notes 150–51.

198. Long, *supra* note 189, at 649.

199. *See* 37 C.F.R. § 1.56(a) (2016) ("Each individual associated with the filing and prosecution of a patent application has a duty of candor and good faith in dealing with the Office . . . .").

200. *See* MPEP, *supra* note 13, § 2107(II)(C)(2). In practice, patent examiners are very unlikely to make this sort of rejection. *See* John R. Thomas, *Collusion and Collective Action in the Patent System: A Proposal for Patent Bounties*, 2001 U. ILL. L. REV. 305, 311 ("For most inventions, the utility requirement now appears to be satisfied by only a minimal showing that the invention is operable and provides a tangible benefit."). Moreover, patent examiners are instructed to accept the claims of the patent application unless there is clear evidence that they are incorrect. MPEP, *supra* note 14, § 2107(II).

201. MPEP, *supra* note 13, § 2107(II)(C)(2).

202. Patents may contain two types of claims, independent and dependent claims. Independent claims do not refer back to any other claims. *See* 37 C.F.R. § 1.75(c). Dependent claims refer back to and limit another claim in the same patent. *Id.* Dependent claims are required by statute to be narrower in scope than the independent claims from which they depend because the dependent claims must both "incorporate by reference all the limitations of the [independent] claim" and "specify a further limitation of the subject matter claimed." 35 U.S.C. § 112(d) (2012).

203. *See* 35 U.S.C. § 112(d).

204. Jeanne C. Fromer, *Claiming Intellectual Property*, 76 U. CHI. L. REV. 719, 740 (2009).

arguments might need to be made during litigation, adding a wide range of elements to claims allows for a commensurately wide range of potential arguments during litigation.

A similar "insurance" policy against findings of anticipation and non-obviousness might motivate patentees to include ancillary elements in patent claims.[205] A patent is valid only if it is novel (not anticipated) and non-obvious, meaning that the invention is not disclosed in or rendered obvious by the "prior art" (information made public prior to the patent's filing).[206] The anticipation analysis requires that the party alleging patent invalidity prove that *each element* of a claim is present in one single disclosure (generally a publication).[207] Statistically, the greater the number of elements in a claim, the more difficult this will be to accomplish. Such claims are more difficult for an examiner to reject on grounds of anticipation, because they include so many elements that it is difficult to find a reference including *all* elements.[208] In this manner, patent drafters can add ancillary elements to create claims that are more likely to be patentable.

3. Wearing Out the Examiner

Examiners have a small amount of time allocated to examining each patent—an average of 18 hours per patent.[209] Examiners earn bonuses and promotions by disposing of a greater number of cases.[210] If an examiner rejects a patent, the patentee can amend the claims and ask the examiner to reconsider, consuming more of the examiner's time for little credit.[211] Thus,

---

205. A survey by Dennis Crouch found that "backup" in validity challenges was one of the most common purposes of dependent claims. Dennis Crouch, *Theory of Dependent Claims: Survey Results*, PATENTLY-O (May 22, 2008), https://patentlyo.com/patent/2008/05/theory-of-depen.html.

206. 35 U.S.C. §§ 102–03.

207. *See* Merck & Co. v. Teva Pharm. USA, Inc., 347 F.3d 1367, 1372 (Fed. Cir. 2003) ("An 'anticipating' reference must describe all of the elements and limitations of the claim in a single reference . . . .").

208. It will also be difficult to reject on the grounds of obviousness (35 U.S.C. § 103) because, although references can be combined, there must be some "reason that would have prompted a person . . . to combine the elements in the way the claimed new invention does." Takeda Chem. Indus., Ltd. v. Alphapharm Pty., Ltd., 492 F.3d 1350, 1356–57 (Fed. Cir. 2007) (quoting KSR Int'l Co. v. Teleflex Inc., 550 U.S. 398, 418 (2007)).

209. Lemley, *supra* note 99, at 1496 n.3 ("Examiners have astonishingly little time to spend on each application—on average, a total of eighteen hours, including the time spent reading the application, reading the submitted prior art, searching for and reading prior art in databases accessible to the PTO, comparing that prior art to the application, writing an office action, reading and responding to the response to office action, iterating the last two steps at least one and often more times, conducting an interview with the applicant, and ensuring that the diagrams and claims are in form for allowance.").

210. The system is slightly more complex than described here—examiners also get some recognition for first rejections and other actions. *See* ADAM B. JAFFE & JOSH LERNER, INNOVATION AND ITS DISCONTENTS: HOW OUR BROKEN PATENT SYSTEM IS ENDANGERING INNOVATION AND PROGRESS, AND WHAT TO DO ABOUT IT 97 (2004).

211. *Id.*

scholars argue that there is an incentive for examiners to grant applications rather than go through repeated rounds of rejections.[212]

Ancillary language may be part of an affirmative strategy by patent applicants to "wear down the examiner"[213] by capitalizing on overworked examiners seeking to meet their productivity quotas. For example, U.S. Patent No. 7,354,888 claims an antibacterial solution to clean various food products.[214] The patent further claims use of the cleaner on products such as "spinach, kale, chard . . . capon, game hen, pigeon . . . tuna, swordfish, [or] shark" and many other foods.[215] If an examiner wishes to reject a claim containing such a list of alternates for obviousness or anticipation, the examiner generally need only produce prior art relating to one of the objects in the list.[216] However, the applicant can then simply amend the claim to remove that object from the list, requiring the examiner to find a different reference with a different object. For example, if the examiner were to find a reference teaching use of the claimed cleaner on spinach, the applicant could amend the claim to include only kale and chard, creating additional work for the examiner.

B. THE PROBLEM WITH CLUTTER

If ancillary language is good for patentees, it may also be good for the public. Simplistically, if patentees benefit from ancillary language, its presence increases the value of the patent, which ought to increase incentives for patentees to innovate.[217] However, innovation is an iterative process, and a boon to patentees may hamper downstream innovators. The discussion below demonstrates that ancillary language may be a substantial contributor to the multitude of problems associated with patent claims' clarity, readability, searchability, examinability, and compliance with disclosure requirements. By deepening our understanding of the underlying shape of patent claims, we can better understand why claims perform poorly and improve reform efforts.

---

212. *Id.*

213. This strategy is well-established in other contexts. *See, e.g.*, Joseph Farrell & Robert P. Merges, *Incentives to Challenge and Defend Patents: Why Litigation Won't Reliably Fix Patent Office Errors and Why Administrative Patent Review Might Help*, 19 BERKELEY TECH. L.J. 943, 944 (2004) ("Biased procedures . . . permit[] a strategy of 'wearing down the examiner' to obtain a patent . . . ."); Mark A. Lemley & Kimberly A. Moore, *Ending Abuse of Patent Continuations*, 84 B.U. L. REV. 63, 75 (2004) ("[C]ontinuation applications permit the applicant to wear down the examiner [and] obtain[] a patent that the PTO would otherwise refuse to grant . . . .").

214. U.S. Patent No. 7,354,888 col. 12 l. 50–67, col. 13 l. 1–35 (filed Nov. 10, 2004) (issued Apr. 8, 2008).

215. *Id.* at col. 16 l. 11–23.

216. MPEP, *supra* note 13, § 803.

217. *See* Robert P. Merges & Richard R. Nelson, *On the Complex Economics of Patent Scope*, 90 COLUM. L. REV. 839, 868–84 (1990) (explaining how broader patents incentivize innovation, but also explaining how broader patents can impede secondary innovation).

1. Disclosure

Because ancillary language often describes concepts that are far afield from the core invention, there is a risk that the language will reflect a patent claim that goes beyond describing the invention to describing potential corollaries that the patentee does not actually possess and cannot explain how to make or use. Because patents must disclose certain information about the claimed matter, ancillary language that describes concepts not disclosed in detail in the specification may sometimes reflect claims that are not actually valid because the claims encompass scope greater than the invention.

Granted patents are presumptively valid[218] so, because this Article studies granted patents, the studied claim language is presumptively enabled and described. However, in practice, many granted patents are in fact invalid. [219] Ancillary language might signal patents that are invalid for non-enablement of the concepts claimed by that language, or insufficient written description. This is especially true in the unpredictable arts such as chemistry or biology. For example, given the complexity of determining if a compound can be used to treat a disease in humans, a claim directed to using a novel compound to treat various diseases, where no testing had been done, might in fact not be enabled. Though it is certainly possible for ancillary language to be enabled and adequately described,[220] its widespread presence is at least a red flag with regards to enablement and written description, particularly in the unpredictable arts. This is supported by the finding that applications with more ancillary language are more likely to be rejected for lack of enablement or written description (Figure 6).

A practical challenge with patents that claim greater breadth than they have enabled or described is a chilling effect on future research. A company choosing an area in which to conduct research often begins with an overview of the patent landscape. If the patent landscape appears to be crowded, the company may choose to focus its research efforts elsewhere. The company is unlikely to spend tens of thousands of dollars sifting through patents before beginning its research program, so the presence of ancillary claim language, which creates the appearance of a crowded landscape, is significant.

Ancillary claim language has a further chilling effect on future research. Ancillary language in some instances discloses an untested idea:[221] the combination of the ancillary element and the central element claimed by the patent or the use of the central element for an ancillary method. In this sense,

---

218. Doug Lichtman & Mark A. Lemley, *Rethinking Patent Law's Presumption of Validity*, 60 STAN. L. REV. 45, 46–47 (2007).

219. *See, e.g.*, Frakes & Wasserman, *supra* note 30, at 619 (discussing the "PTO's perceived penchant for issuing invalid patents").

220. The specification does not need to describe claim elements "*in haec verba*." Vasudevan Software, Inc. v. MicroStrategy, Inc., 782 F.3d 671, 682 (Fed. Cir. 2015).

221. Or, if the invention has been tested, the results are not reported in the patent.

ancillary language allows patentees to precisely, but inaccurately, claim something that they have not invented. In the pharmaceutical industry, where ancillary language is very prevalent,[222] these disclosed combinations can have surprising synergistic (or toxic) effects, or may turn out to be capable of treating an unexpected condition.[223] Thus, there is a great deal of downstream research that is needed before such a combination can be commercialized.

For example, a patent might claim a novel molecule and further claim use of that novel molecule to treat "respiratory diseases; inflammatory diseases . . . allergies; ophthalmic diseases; cutaneous diseases . . . gastrointestinal (GI) disorders . . . renal disorders . . . Alzheimer's disease, Down's syndrome, Huntington's" and many others.[224] However, the patent does not describe any testing of the compound for treatment of this highly-varied set of conditions (thus, it is ancillary language), and, even if the compound were effective for all of the listed conditions, significant additional experimentation and many clinical trials would be needed to determine the proper dosage form, amount, and regimen. Those experiments may not be done, because it will be difficult for companies other than the patentee to get a patent on use of the compound to treat any of the listed conditions.[225]

There is some evidence that patentees deliberately disclose more information than necessary in order to create this chilling effect. Firms leading a patent race strategically disclose information in order to "lessen the expected value of the patent[s]" that might be granted to others and to "signal the leader's relative position vis-à-vis the laggards."[226]

---

222. *See supra* Part III.B.2.

223. G.W.A. Milne, *Very Broad Markush Claims; A Solution or a Problem? Proceedings of a Round-Table Discussion Held on August 29, 1990*, 31 J. CHEMICAL INFO. & COMPUTER SCI. 9, 29 (1991) (saying, in the context of broad chemical Markush groups: "it inhibits somebody actually making it and it may well be that that compound has an entirely different activity which could be useful"); *see also* Arti K. Rai & Grant Rice, *Use Patents Can Be Useful: The Case of Rescued Drugs*, 6 SCI. TRANSLATIONAL MEDICINE 1, 1–2 (2014) (discussing drugs that failed clinical trials for the predicted use, but were later found to have other therapeutic uses).

224. U.S. Patent No. 6,277,862 claim 11 (filed May 22, 1998) (issued Aug. 21, 2001).

225. The tenuousness of later patents on prior-disclosed combinations is illustrated in *Merck & Co., Inc. v. Biocraft Labs., Inc.*, 874 F.2d 804 (Fed. Cir. 1989). In *Merck*, the plaintiff had patented a diuretic drug consisting of a combination of amiloride hydrochloride and hydrochlorothiazide and obtained FDA approval for the product. *Id.* at 805–06. An earlier patent had disclosed "various . . . guanidines, one of which is amiloride," and that these compounds "are useful in combination with other classes of diuretic agents," one of which is hydrochlorothiazide. *Id.* at 806. Although the plaintiff argued that its combination was "medically synergistic" and therefore not obvious in light of the prior disclosed combination, the court disagreed and held the plaintiff's patent invalid. *Id.* at 808–09. Fear of this outcome is likely to inhibit testing on and commercialization of previously disclosed combinations.

226. Douglas Lichtman et al., *Strategic Disclosure in the Patent System*, 53 VAND. L. REV. 2175, 2179 (2000). *But see* Oren Bar-Gill & Gideon Parchomovsky, *The Value of Giving Away Secrets*, 89 VA. L. REV. 1857, 1873–74 (2003) (suggesting that firms strategically disclose potentially patentable information without patenting it). A survey of patent attorneys reports that "disclosure

2. Clarity

Patent claims are hard to read.[227] One prevalent criticism is the length of patent claims. Because claims must be restricted to one sentence, elongating that sentence is likely to reduce its clarity,[228] making it close to incomprehensible.[229] Ancillary elements may number in the dozens or even hundreds. Given the already existing concerns about claim length and comprehensibility, adding large numbers of words to claims has the potential to seriously obstruct the reader's ability to grasp the meaning of patent claims.

Ancillary language impedes readability in another way: it is not defined in the specification. Patent claims are intended to be read "in light of the specification" [230] so when a word in a claim is ambiguous or lacks clarity, the first place for the reader to turn is the specification. Ancillary claim language is, by definition, not discussed in the specification. The reader is therefore unlikely to be able to turn to the specification for clarification if an ancillary term should prove to be ambiguous or unclear. Though many ancillary terms may be quite clear without reference to the specification, others are not. For example, U.S. Patent No. 4,415,459, directed to systems and methods for disposing of industrial and municipal waste,[231] includes terms in the claims such as "waste-resist[a]nt"[232] (used to describe a quality needed in an internal lining for a waste-disposal container), "vehicular access means . . . facilitating vehicular transport,"[233] and "space-conservative" (referring to the method of storing waste matter).[234] None of these terms are mentioned in the specification, and yet these are precisely the types of terms that litigators fight over in claim construction proceedings. Many questions can arise as to the scope and meaning of the terms. What is a "waste-resistant" lining? What types of waste need it resist? Does it include nuclear waste?[235] Does it need to be completely waste-resistant or merely practically waste-resistant?

---

can serve a defensive purpose: disclosure makes it more difficult for rivals to patent inventions related to the disclosed information." Scott Baker & Claudio Mezzetti, *Disclosure as a Strategy in the Patent Race*, 48 J.L. & ECON. 173, 176 (2005). Note that these comments are made in the context of disclosures in the specification, but disclosures made in the claims have the same effect.

227. *See* Osenga, *supra* note 3, at 620.

228. Kelly Casey Mullally, *Patent Hermeneutics: Form and Substance in Claim Construction*, 59 FLA. L. REV. 333, 349 (2007) ("Intended to be a succinct statement of the invention, claims must be written as a single sentence, often requiring extreme contortions of language.").

229. Osenga, *supra* note 3, at 620 ("Patent claims are notoriously difficult to understand. . . . One factor that affects comprehension of language is the word length of the passage to be understood.").

230. *Id.* at 643.

231. U.S. Patent No. 4,415,459 col. 1 ll. 6–15 (filed June 8, 1981) (issued Nov. 15, 1983).

232. *Id.* at claim 10.

233. *Id.* at col 13 l. 2–9.

234. *Id.* at col. 13 l. 15. "Space-conservative" is part of the preamble, so it may not be a limiting part of the claim. MPEP *supra* note 13, § 2111.02.

235. Possibly. A different claim mentions "material of relatively low nuclear radiation transmissibility." '459 Patent, *supra* note 233, at col. 11 l. 53–54.

Even very prosaic inventions can benefit from clarification of claim terms. U.S. Patent No. 5,256,432 is directed to a method of fusing pizza-topping ingredients together to create a "toppings disc."[236] Several claims include the step of fusing the ingredients by softening the cheese and then "resolidifying said quantity of cheese."[237] "Resolidifying" is not defined in the specification. Attorneys could argue all day about whether soft cheese is solid. The finding (presented in Figure 6) that applications with more ancillary language are more likely to be rejected for indefiniteness[238] supports the contention that ancillary language diminishes claim clarity in this manner.

Of course, not all definitional problems can be solved by discussion in the specification. Language is inherently ambiguous,[239] and thus patent claims cannot be fully free of ambiguity. Yet for claim terms discussed in the specification, those seeking to understand the meaning of a claim term at least have a place to start. Not so for ancillary language.

However, in some instances ancillary language may actually improve the clarity of patent claims. One challenge to claim clarity is that drafting attorneys are taught to use general, rather than specific terms.[240] However, general terms can be challenging to understand. Linguistic and cognitive scientists have found that humans are best at understanding general categories when they are given several central cases or models as "exemplars."[241] Use of such exemplars can reduce notice costs of patent claims.[242]

Thus, lists of ancillary language that may not accurately reflect what the patentee has actually invented might still add precision in a way that aids comprehensibility of patent claims. To illustrate, consider U.S. Patent 8,910,876, claiming a method for displaying programming material on a computer or related device, where the programming material is guided by a user profile.[243] Though Claim 1, an independent claim, broadly encompasses this method, Claim 5, a dependent claim, narrows the user profile to "information selected from the group consisting of . . . travel preferences, product preferences, dating preferences" and several others.[244] Claim 14, also dependent, further narrows "dating preferences" to information "selected

236. U.S. Patent No. 5,256,432 col. 7 l. 7–17 (filed Sept. 30, 1992) (issued Oct. 26, 1993).

237. *Id.* at col. 7 l. 26–29, col 15 l. 17–20, 47–49.

238. A claim is rejected for indefiniteness if it, "read in light of the specification . . . fail[s] to inform, with reasonable certainty, those skilled in the art about the scope of the invention." Nautilus, Inc. v. Biosig Instruments, Inc., 134 S. Ct. 2120, 2124 (2014).

239. And some ambiguity may be desirable, particularly for new inventions. *See* Kristen Osenga, *Cooperative Patent Prosecution: Viewing Patents Through a Pragmatics Lens*, 85 ST. JOHN'S L. REV. 115, 148 (2011).

240. Particularly in independent claims.

241. Fromer, *supra* note 206, at 765.

242. *Id.* at 761–66.

243. U.S. Patent No. 8,910,876 col. 14 ll. 21–37 (issued Dec. 16, 2014).

244. *Id.* at col. 14 ll. 49–56.

from the group consisting of . . . user's or prospective dating partner's . . . occupation, educational level, religion, family background, interests, hobbies, likes and dislikes, political orientation" and several others.[245] Though most of this additional information is ancillary and somewhat far afield from the core of the invention, it does clarify what the inventor meant by broader terms such as "a user profile." However, such information would be equally elucidatory if it were in the patent's specification, rather than the claims, so it is still not clear that ancillary language has a beneficial role when it occurs in the claims.

3. Searchability

Section II.C.3 explained how patent searches, generally keyword searches of claim language, are conducted by casting a wide net using many different keywords and then narrowing possibly relevant patents through manual review. This strategy is thorough but expensive—even if it takes only a few minutes to scan and discard a false positive, doing so for thousands of patents becomes time consuming and inefficient. The need to whittle down patents is a practical problem, with a survey of industry respondents finding "that there may be a large number of patents to consider initially" but only a handful that were relevant.[246] The need for complete thoroughness in these searches makes false positives inevitable and expensive.

Ancillary language has the potential to greatly increase the number of false positives in clearance searches, exacerbating the problem of poor notice caused by the presence of an overwhelming number of patents.[247] For example, a search for granted patents claiming "movie" will return more than 3,000 patents,[248] including U.S. Patent No. 8,750,468. The '468 patent is directed to a method of personalizing the music or message heard while on hold on a telephone call.[249] The patent then claims methods of personalizing the call based on, among other things, "at least one movie rented" or "at least one product purchased."[250] In this patent, "movie" is ancillary language, as the specification does not discuss movies and methods for using movies to personalize advertising for unfortunate listeners placed on hold. It is highly unlikely that this result will be relevant to a searcher seeking patents covering movies. However, this patent and others like it will be caught in the search net

245. *Id.* at col. 15 ll. 38–46.

246. John P. Walsh et al., *Effects of Research Tool Patents and Licensing on Biomedical Innovation*, *in* PATENTS IN THE KNOWLEDGE-BASED ECONOMY 285, 294 (Wesley M. Cohen & Stephen A. Merrill eds., 2003).

247. *See supra* Part II.C.3.

248. Results of Search in US Patent Collection db for: ACLM/movie, USPTO PATENT FULL-TEXT AND IMAGE DATABASE, http://patft.uspto.gov/netahtml/PTO/search-bool.html (input "movie" in Term 1 field and select "Claim(s)" from drop-down menu in Field 1 field and select "search") (providing over 3,000 results for "movie" patent claims as of December 2017).

249. U.S. Patent No. 8,750,468, col 2, ll. 11–34 (issued June 10, 2014).

250. *Id.* at claim 3.

and will have to be manually removed. By including language in patent claims that is not part of the core invention, drafters of ancillary language drive up the cost and complexity of searches.

4. Examinability

When patents contain large amounts of ancillary language, or when the language is so unclear or unsupported that examiners require patent applicants to remove it, the presence of ancillary language may increase the time and cost of examination. It stands to reason that patent claims with a larger number of elements will take more time to evaluate. Thus, the addition of ancillary claim elements increases the burden on examiners. Examiners have complained that "extremely long claim sets only for appearance's sake increase examination time and do not generally provide any added benefit."[251] Moreover, to the extent that ancillary language is part of an affirmative strategy by patent applicants to "wear down the examiner" by capitalizing on overworked examiners seeking to meet their productivity quotas,[252] it reduces the examinability of patent claims.

C. CLUTTER AND PATENT THEORY

1. A New Framework for Claim and Invention

Claims were created to distinctly point out the invention and separate the new from the old. The presence of ancillary language indicates that this separation is far from clear cut. Plainly, claims include additional information beyond the essence of the invention and thus the relationship between claims and the invention is complex and multifaceted. Claims do not entirely separate the new and the old, instead, they appear to be a combination of the new, the old, and the too new (novel ideas that the patentee has not actually figured out how to practice). The view that the claim is the invention and the invention is the claim is too simplistic; the presence of ancillary language indicates that claims are far more intricate than previously thought.

As an initial matter, it is important to consider whether or not we can in fact accomplish the goal of precisely pointing out the invention without incorporating other information. This aspiration simply may not reflect the realities of claim drafting and the ways in which people read and understand words. Patent drafters are faced with the task of describing an invention that is entirely new without reference to the old in a way that is still understandable. Can something completely new be described with no reference to the old, to context, or to aspirational uses? It seems unlikely that this could be accomplished without a significant loss of clarity, and thus this

251. Lauren Anderson & Ryan Cagle, *An Examiner's Tips for Speedier Patent Prosecution*, IPWATCHDOG (Dec. 19, 2016), http://www.ipwatchdog.com/2016/12/19/examiners-tips-speedier-patent-prosecution.

252. *See supra* note 211 and accompanying text.

Article takes the position that it will not be fully possible—nor is it desirable—to entirely eliminate ancillary language.

However, ancillary language has a cost. Part B, above, explained how ancillary language contributes to many practical problems in patent law. The original purpose of patent claims was to isolate the invention in order to avoid the confusion caused by mixing the novel aspect of the patent with other descriptive information. Congress sought to define the invention in the claims and restrict the other information to the specification. If claims contain too much ancillary language, it negates the purpose of this separation and brings the patent system full circle back to its pre-claim condition where Justice Story complained of patents that "mix[] up the old and the new," making "it [] impossible for the court to say, what, in particular, is covered as a new invention."[253]

Given that the historical goal of complete separation of old and new may not be possible, but that ancillary language comes at a cost to the patent system, we must consider the role that ancillary language plays in achieving or hindering the goals of patent claims and how to calibrate it accordingly. Because patent claims are intended to convey information, the use and impact of ancillary language can be assessed through an information-cost framework. Providing information can improve the clarity of legal boundaries up to a point, after which excess information muddles and confuses the audience, leading to sub-optimal results.[254] There is a tradeoff between concision, precision, and accuracy. Ancillary language may improve the precision of patent claims, but at a cost to their concision and accuracy.

Ancillary language in many cases increases the precision of the information conveyed by patent claims because it provides additional detail. However, increased precision comes at the cost of concision and accuracy. Concision is lost when ancillary elements lengthen claim terms. Accuracy is lost when ancillary elements masquerade as part of the invention or describe material that the patentee has not enabled or described in the specification.

Accuracy may be the simplest element of the framework to evaluate. We should always aim for accuracy. Ancillary language that imagines how the invention could be used but does not reflect what the inventor has actually created has no place in patent claims. Similarly, ancillary language that misleads readers has no place in patent claims.

Balancing the goals of concision and precision is more difficult, because both are helpful but achieving one often comes at the cost of the other. A guiding general principle is that there is a sliding scale between the two. Some amount of language is always necessary to communicate, so complete concision is not functional. However, complete precision is also not

253. Lowell v. Lewis, 15 F. Cas. 1018, 1020 (C.C.D. Mass. 1817) (No. 8568).

254. *See* Thomas W. Merrill & Henry E. Smith, *Optimal Standardization in the Law of Property: The* Numerus Clausus *Principle*, 110 YALE L.J. 1, 8 (2000).

functional because the level of detail required would be impossible both to write and to read. In the middle is an area where adding some language enhances communication but where too much language confuses and hinders information transmission and increases transaction costs. An optimal patent claim would include information up to the point where the benefits of this information are outweighed by its costs.

These principles can guide our thinking about the proper purpose and content of claims, and when ancillary language is appropriate. Ancillary language that creates a sentence spanning four pages or reflects concepts that are purely the patent drafter's imagination has too much precision and too little accuracy, respectively. Conversely, ancillary language that clarifies a general concept may aid comprehension without excessively cluttering.

Our view of ancillary language should also reflect how patents are actually used. For example, in the era before computers, the possibility that ancillary language would create false positives in searches was not a problem, because keyword-based computer searches did not exist. It is presently a problem, but if computer scientists can create technology to clear the clutter, it may again cease to be a problem. Similarly, should we make broad advances in the pharmaceutical sciences such that experimentation becomes less unpredictable, issues of whether ancillary language is enabled may be less pressing. This is to say that the appropriate contours of language in patent claims are not static. Patents evolve, technology evolves, and our use of patents evolves. The question of how claims should be crafted should be continually revisited in light of these changes.

D. Reforming Clutter

Ancillary language in patent claims has the potential to harm by hampering searches, confusing readers, hobbling examiners, and chilling future research. Yet it should not be eliminated outright because it also has the potential to help patentees navigate commercial and legal uncertainty, clarify genus claims, and possibly increase the reward to the patentee. Moreover, it may not even be possible to write claims without any ancillary language. Thus, policy changes should be calibrated to retain as much benefit as possible while reducing harms. This Article takes a two-pronged approach to addressing ancillary language. First, examiners should remove particularly egregious uses of ancillary language. This includes ancillary language that is extremely long, such as claims that stretch on for pages, and ancillary terms that describe embodiments that the patent does not enable or describe. Subpart 1, below, explores some mechanism for examiners to do this. Second, for ancillary language that is not obviously harmful, Subpart 2 proposes a combination of improved search technology and better legal structures to aid readers in identifying and processing ancillary language.

1. Strategies for Removing Ancillary Language

i. Prolix

Some uses of ancillary language will be particularly outrageous because they are excessively long or simply confusing. For these extreme cases, patent examiners should have tools to reject claims and require applicants to redraft.[255]

Patent examiners may reject a patent claim "as prolix" when it contains "such long recitations or unimportant details that the scope of the claimed invention is rendered indefinite thereby."[256] While examiners do reject patents for prolix, the rejection appears to be used relatively infrequently.[257] Prolix patents hide the key elements of the invention among a large amount of less relevant verbiage, making it difficult for readers to pick out the precise invention disclosed in the patent or useful details about how to make or use the invention. Frustrated courts have described patents as "a long, prolix combination with a liberal sprinkling of adverbs and almost no punctuation."[258] "All claims were prolix as they contained long recitations of unimportant details . . . ."[259] Patent claims with lengthy ancillary sections appear to be a perfect target for a prolix rejection, although the requirement that prolixity be so extensive that the claim is "rendered indefinite thereby" may mean that prolix rejections will be viable only in cases of severe patent clutter. Nonetheless, increased use of the prolix rejection could be a tool to tackle ancillary language in the worst cases.

ii. Enablement and Written Description Red Flags

Some uses of ancillary language will be particularly outrageous because they describe embodiments that are not enabled or described by the patent. The PTO is instituting several new training modules and pilot programs to address lack of enablement and written description more generally.[260] The

255. It may be best to limit examiner policing of ancillary language to extreme cases because at some point the cost of having examiners decide whether or not to permit such language may exceed the cost to society of having the language in granted patents. *See* Andres Sawicki, *Better Mistakes in Patent Law*, 39 FLA. ST. U. L. REV. 735, 777 (2012).

256. MPEP, *supra* note 13, § 2173.05(m); *see also* Mowry v. Whitney, 81 U.S. 620, 644 (1871) ("The law requires every inventor, before he can receive a patent, to furnish a specification . . . avoiding unnecessary prolixity . . . ."); *In re* Wood, 155 F.2d 547, 551 (C.C.P.A. 1946) ("[W]e think it wholesome to discourage this tendency towards undue prolixity . . . .").

257. Dennis Crouch, *The Requirement that the Written Description be Concise*, PATENTLY-O (Dec. 9, 2009), http://patentlyo.com/patent/2009/12/the-requirement-that-the-written-description-be-concise.html ("The USPTO does reject individual claims that it deems overly wordy or 'prolix' . . . . [But a] search[of] the BPAI database of decisions (1997–2009) . . . found only one case reviewing a rejection where the Examiner found claims invalid as prolix . . . .").

258. Thermo King Corp. v. White's Trucking Serv., Inc., 292 F.2d 668, 675 (5th Cir. 1961).

259. Minn. Mining & Mfg. Co. v. Norton Co., 280 F.Supp. 674, 677 (N.D. Ohio 1967).

260. *Examination Guidance and Training Materials*, USPTO (Sept. 20, 2017), https://www.uspto.gov/patent/laws-and-regulations/examination-policy/examination-guidance-and-training-materials.

concept of ancillary language as described by this Article can be used to guide application of these programs in ways that will result in less ancillary language and better adherence to the enablement and written description requirements.

For example, at present the PTO has two different standards for assessing written description, depending on whether the claim was part of the original application or was added as an amendment. For the former, the examiner is instructed to presume that there is adequate written description, and the applicant does not need to provide any specific argument that the requirement is met.[261] For the latter, there is no such presumption and the applicant is instructed to "show support in the original disclosure for the new or amended claims."[262] To reduce levels of ancillary language that are reflective of lack of written description, the requirement that the applicant specifically point to support for amended claims could be extended to claims in the original application. This might take the form of a claim chart, where each row in the first column is a claim element and each row in the second column is support in the specification. Applicants and examiners are already used to seeing this sort of support as it is currently common for amended claims.[263]

Alternatively, the presence of ancillary language might be used as an internal metric for the examiner to quickly find claims that might not be adequately described or enabled. The examiner could use an algorithm comparable to the one developed for this Article to provide a list of claim terms that appear little in the specification, and could flag those terms for closer review. Of course, this approach might cause patent applicants to attempt to game the algorithm by repeating claim terms in the specification without any actual substantive discussion, which would not solve the problem of ancillary language and would simply make the specification harder to read. However, at least for currently pending applications it may be a helpful strategy for the examiner. An algorithmic and automatable helper may be particularly useful because time-crunched examiners admit to often only skimming the specification.[264]

---

261. MPEP, *supra* note 13, § 2163.03; *see also In re* Wertheim, 541 F.2d 257, 262 (C.C.P.A. 1976) (detailing the way in which the court will evaluate the patent application's written description).

262. MPEP, *supra* note 13, § 2163.

263. For example, a patent prosecution textbook recommends that when claims are amended,

> [t]he basic approach applicants should use when attempting to add additional information in the application is to provide a detailed explanation of why the introduced subject matter is supported by the originally filed application. Accordingly, applicants should refer to specific portions of the originally filed application that provide support to obtain the most favorable decision from the Examiner.

1 IRAH H. DONNER, PATENT PROSECUTION: LAW PRACTICE AND PROCEDURE, 2-483 (10th ed. 2017).

264. Lauren Anderson & Ryan Cagle, *An Examiner's Tips for Speedier Patent Prosecution*, IPWATCHDOG (Dec. 19, 2016), http://www.ipwatchdog.com/2016/12/19/examiners-tips-speedier-

The PTO recently concluded a Glossary Pilot Project that required participants to establish the scope of a term "by presenting a positive statement of what the term means," with a stipulation that the definition "cannot consist solely of a list of examples, synonyms, and/or exclusions."[265] The Project also required participants to "include definitions that w[ould] assist in clarifying the claimed invention."[266] This definitional clarification could be useful in technologies where terms are vague, such as software. In other areas, such as chemistry and pharmaceuticals, terms are not generally thought to be vague.[267] Instead, the problem of ancillary language may be more one of excessive text that may not be properly enabled or described.

A Glossary-type requirement could ask patentees to create a glossary with an explanation of how the ancillary element could be used with the core invention. For example, a patentee who discovered a molecule to treat diabetes but additionally claimed use of the molecule in combination with many other drugs such as "cholesterol absorption inhibitors," "fibrates," "antioxidants," and "sulfonylureas," would be asked to specify how those drugs could be used with the novel molecule.[268] This requirement would highlight claim terms that have no connection at all to the novel aspect of the patent as well as claim terms where the patentee is unable to provide a detailed or substantial explanation. However, adding additional claims would require additional attorney costs, potentially dissuading patentees from including the terms if they are of little importance. Care must be taken in adding prosecution expenses for patentees, as the patenting process is already prohibitively expensive for some inventors.[269]

2. Strategies for Adapting to Ancillary Language

i. Better Searching

Certain problems relating to ancillary language, such as increased search costs, can be at least partially addressed through improvements in technology. Search algorithms could use the frequency with which a word appears in the specification—the proxy used by this Article—as a factor in ranking results by relevance (though again, this would be open to gaming by patentees). Alternatively, search services could give users the option not to return results

---

patent-prosecution (citing an interview with an examiner where the examiner stated "that one view is that the drawings and claims are most important during review of a patent application and the specification is mostly skimmed").

265. USPTO, GLOSSARY PILOT PROGRAM: FREQUENTLY ASKED QUESTIONS 4, http://www.uspto.gov/sites/default/files/patents/init_events/faq_glossaryapplicant_07012014.pdf.

266. *Id.*

267. Janet Freilich & Jay P. Kesan, *Towards Patent Standardization*, 30 HARV. J.L. & TECH. 233, 243–44 (2017).

268. U.S. Patent No. 7,759,366 (filed Feb. 5, 2009) (issued July 20, 2010).

269. Tamara Monosoff, *Don't File for That Patent Yet*, ENTREPRENEUR (Feb. 8, 2010), https://www.entrepreneur.com/article/204918.

if the results included ancillary language. Search algorithms could use the methodology developed by this Article as a springboard to implement far more complex methods of sorting results including, for example, determining whether ancillary language appeared as part of a set of possibilities, a list of exemplars, or an enumeration of limitations. Each category would have different value to searchers depending on the particular goal of the search, and sophisticated searchers would greatly benefit from the ability to identify and exclude those that were less relevant. These more complex methods would presumably be secret and proprietary and might therefore also be harder for patentees to game.

A further possibility is to use legal approaches to facilitate better searching. For example, Peter Menell and Michael Meurer have suggested in other contexts that applications be required to indicate structures and materials corresponding to claim terms by hypertext.[270] Menell and Meurer's suggestion is in the context of means-plus-function claims, but it could be expanded to other types of claims. If applicants had to highlight portions of the specification corresponding to a particular claim term, it would be very clear when claims were ancillary because no language or little language would be highlighted. This would not remove ancillary language, but it would emphasize its presence, allowing readers to easily skip over it, or perhaps allowing search engines to easily de-emphasize it in search algorithms.

Note that even if the solution is technological, there is value to discussing questions regarding ancillary language in legal articles. That search engines have not yet developed algorithms to avoid ancillary language suggests that they may not be aware of it. Studies by lawyers deeply familiar with patents play a role in enhancing our understanding of the characteristics of and problems with patent text. This enhanced awareness allows computer scientists to craft technological solutions where appropriate.

ii. Removal to the Specification

Many of the problems with ancillary language come down to distractions—to searching, examining, or reading—posed by the language. These distractions would be minimized if ancillary claim language were moved to the specification. Indeed, to the extent that ancillary language is used to provide specific examples of a more general category or to speculate on possible uses for a technology, it belongs in the specification. The purpose of the specification is to contain such examples and descriptions of such methods.

The downside of moving ancillary language from the claims to the specification is that patentees would no longer be able to use these claims as backup during litigation. However, the effect of this loss is mitigated because

270. Peter S. Menell & Michael J. Meurer, *Notice Failure and Notice Externalities*, 5 J. LEGAL ANALYSIS 1, 33 (2013).

patentees could still use ancillary language (in the specification) as backup during prosecution, by drafting new narrower claims based on the language in the specification, if the broader claim was rejected.[271] After being granted a patent, it is also possible to narrow claims in reissue proceedings.[272] Moreover, since ancillary language describes concepts that the patentee has never made, used, or elaborated upon in the specification, it seems equitable to reduce the patentee's ability to use such language as backup in patent claims, particularly when the language is harmful to third parties. A further problem with moving ancillary language to the specification is that it may require patent claim readers to cross-reference claims with the specification to fully understand the claim. This is already a significant problem (most notably in means-plus-function claims).

If patent examiners reject more claims for prolix or lack of disclosure, applicants should start to naturally move ancillary language from the claim to the specification (where it would not face rejection). Applicants could be further encouraged to do this by training examiners to recommend such a course of action. In several applications cited above,[273] examiners suggested that patentees claiming a broad genus either give examples in the specification or replace the broad genus with examples in the claims. Patentees responded by adding dozens (or hundreds) of ancillary elements to the claims.[274] Though it is beneficial for examiners to reject overly broad claims, examiners should be encouraged to do so by asking for elaboration in the specification[275] or by suggesting that the claim be amended to include a small number of narrower categories, rather than an open-ended suggestion to add examples to the claims. Note that the suggestion to ask for examiner action should be used minimally and cautiously, as examiners are already overworked. Here, it would only replace rejections that examiners are already writing, and so would not add to their workload.

E. ADDRESSING POTENTIAL LIMITATIONS

This Article presents data suggesting that claim language is not entirely focused on the invention, and presents several practical and theoretical reformulations that spring from those results. The methodology used herein has several limitations. I do not believe that these change the key conclusion

271. There would be no new matter rejection because the language would already be in the specification. MPEP, *supra* note 13, § 2163.06.

272. *See, e.g.*, 35 U.S.C. § 251 (2012).

273. *See supra* Part III.B.3.

274. *See supra* Part III.B.4.

275. Patent applicants are often concerned about amending the specification because it may trigger a change in the priority date of the application. However, at least as a theoretical matter, if a claim could be amended to include additional terms, the terms are not new matter, and therefore applicants should be able to add the terms to the specification without altering the priority date. *See* MPEP, *supra* note 13, § 2163.07.

of this study, namely that the relationship between claim and invention is more complex than previously thought. However, the limitations should be taken into account when considering the implications of this project.

First, this Article employs the number of times a word appears in the specification as a measure of the centrality of the word. It is, of course, a proxy, and therefore not a perfect measure of centrality. For one, the frequency measure is only effective to distinguish between central and ancillary language, and does not measure the level of centrality across patents once the central threshold has been met. For example, a claim element that is mentioned only once in the specification is unlikely to be central, but there is not necessarily a difference in centrality between an element that is mentioned 30 times in one patent and an element that is mentioned 60 times in another patent.

A second note is that counting the number of times a claim element is mentioned in the specification likely *underestimates* the amount of ancillary language in the claims. A claim element that is not mentioned in the specification is likely ancillary, but the reverse is not true: an element that is mentioned several times in the specification is not necessarily central. In addition, the methodology divides claims into individual words, rather than terms. In some cases, the words may be prevalent in the specification while the term is not, producing a false negative by this methodology. For example, the Federal Circuit affirmed a finding of invalidity for lack of written description when the term "selling computer" was "not present anywhere in the specification or original provisional application."[276] The patent at issue in that case did not contain the term "selling computer" in the specification but did contain the individual words "selling" and "computer."[277] Consequently, the true prevalence of ancillary language in patents is likely even higher than found in this study.

An additional set of limitations results from the computer-based nature of the study. First, the analysis is limited to nouns for accuracy, though this likely means that the study fails to account for some ancillary language. Second, the analysis does not account for typos or other clerical errors. Of potential significance, the analysis does not capture situations where a species is claimed, but the genus (and not the species) is described in the specification. This scenario may be particularly common in Markush claims, as these claims describe a genus.[278] However, if a species is not described in the specification, that species is arguably not part of the core invention, even if other species in the genus might be. Finally, this analysis does not search drawings. Thus, it may overestimate ancillary language for patents that rely heavily on drawings. This may have a particularly large effect on mechanical

276. Driessen v. Sony Music Entm't, 640 Fed. App'x 892, 896 (Fed. Cir. 2016)

277. U.S. Patent No. 7,003,500 col. 1 (filed Aug. 1, 2000) (issued Fed. 21, 2006).

278. MPEP, *supra* note 13, § 803.02.

patents, where drawings are common, but less of an effect on, for example, chemical patents, where drawings are not common.[279]

Finally, this study cannot differentiate between claim language that is not in the specification because it describes an old and well-known concept, and language that describes a "too new" concept that exists only in the inventor's imagination. It is not clear from this study how many ancillary words belong to each group. Based on my informal review of the results, it appears to be a mix of both, with the unpredictable arts such as chemistry and the life sciences having more non-enabled language and the predictable arts such as the mechanical and computer industries having more old and well-known ancillary language.

These limitations might change the magnitude of the effect reported in this paper, but they are unlikely to change the *existence* of the phenomenon. This Article should not be taken as a precise quantification of ancillary language—the methodology is not adequately accurate to pinpoint such a number—but rather as a demonstration that ancillary language occurs in sufficient amounts that we should account for it in theory, doctrine, and policy.

V. CONCLUSION

Patent claims have been in serious trouble for a long time and efforts to fix them have not been successful. In this Article, I propose that one problem with patent claims is "clutter"—ancillary language that pads claims and extends them beyond easy readability. I analyze 40,000 patents and applications and additionally provide closer case studies of a small portion of this sample in order to uncover and understand ancillary language. I find that approximately 25% of claim language is ancillary, and that it appears across industries. Although claims with more ancillary language are not more valuable by conventional value proxies, I suggest other ways, including signaling and strategic purpose, that ancillary language increases patent value.

The presence and prevalence of ancillary language is troubling as a matter of policy because it confuses the boundaries of the patent claim. Ancillary language likely causes harm to the public in the form of increased search costs, potentially invalid patents, chilling future research, rendering patents more difficult to read and understand, and increasing the Patent Office's cost of examination.

Ancillary language in patent claims is also troubling as a matter of patent theory. Patent claims are thought to be synonymous with the invention. But if claims are synonymous with the invention, why do claims contain so much language that is irrelevant to the invention? Large volumes of ancillary claim

279. Bernadette Marshall, *Better Drawings Make a Better Patent*, WORLD INTELL. PROP. ORG. (Apr. 2010), http://www.wipo.int/wipo_magazine/en/2010/02/article_0008.html.

language suggest that claims are not entirely the same as the invention, but serve some alternative or additional function.

Further, close analysis of ancillary language highlights the disparity between claim and invention even more closely. For example, a patent on a novel molecule that can be used "in medicine" claims use of that molecule to treat a wide range of diseases such as "osteoarthritis," "anxiety," "Alzheimer's," "alcoholism," "asthma," and many others, though the patent describes no research on these diseases and provides no detail on how the compound might treat these diseases.[280] It strains credulity that the same molecule might be able to treat this diverse array of conditions, especially when the patentee has provided no tests or other evidence. The patent claims treatment of these diseases but such treatments are plainly not part of the invention.

Empirical studies of patent law are growing in popularity. However, they are predominantly about patent litigation. Studies of litigation ignore the 99.8% of patents that are never litigated, and also often focus on only the opinion and litigation-related documents, rather than the underlying patent. This Article is an attempt to provide insight into the text of the patent document because the text is the foundation of the patent right and the underlying cause of the oft-studied litigation. By providing one of the first empirical windows into the actual language of the patent document, it is my hope that this Article initiates further studies of and discussion on this topic.

* * *

280. U.S. Patent No. 6,277,862 col. 47 ll. 5–48 (filed Nov. 22, 1996) (issued Aug. 21, 2001).

Appendix A: Summary Statistics by Quartile

	1st Quartile	2nd Quartile	3rd Quartile	4th Quartile
Mean % of Ancillary Language	8.96	20.48	30.72	45.88
Mean % of Ancillary Language in Independent Claims	54.02	54.82	57.04	61.59
Mean Priority Date	2001	1998	1994	1991
Mean Filing Date	2003	1999	1995	1992
Mean Grant Year	2006	2002	1998	1994
Mean Prosecution Length (Years)	2.81	2.71	2.51	2.30
Mean Number of Ancillary Words Added/Removed During Prosecution (Absolute Value[281])	3.08	3.19	4.43	6.62
Mean Forward Citations/Year	1.44	1.51	1.37	1.01
Mean Backward Citations	31.89	26.13	20.59	15.01
Mean Number of Claims	15.09	16.43	15.88	13.09
Mean Length of Specification (Words)	10,741	6,776	4,692	3,319
% of Patents That Are Continuations	11.19	12.38	12.81	12.30
% of Patents That Are Continuations-in-Part	7.06	7.53	7.98	6.19
% Having Year 4 Maintenance Fee Paid	82.72	83.81	83.31	78.49
Industry (NBER Categories) (% of Total Sample in Quartile)				
Chemical	10.88	12.54	14.49	15.56
Computer	30.48	28.19	21.31	12.45
Pharmaceutical	8.30	7.98	9.05	10.17
Electrical	25.72	21.22	19.54	17.21
Mechanical	14.69	16.18	17.77	20.03
Other	9.93	13.90	17.84	24.58

281. Absolute value is shown here because applications both added and removed ancillary language during prosecution, thus, absolute value is a better representation of the magnitude of the change. Change in language during prosecution is discussed in more detail in *supra* Part IV.D.

Appendix B: OLS Regression: Granted Patents

Dependent variable: log(percent ancillary language + 1)[282]

Results shown in the table are exponentiated β coefficients – 1 * 100 (which represent the percentage change in the outcome variable caused by a one unit change in the independent variable)[283]

*p<0.05; **p<0.01; ***p<0.001

Variable	(1)	(2)	(3)	(4)
Number of Claims	0.51***	0.59**		0.80***
Specification Word Count (in thousands)	-2.75***	-2.72***		-2.42***
Priority Date			-2.45***	-2.56***
Prosecution Length				-0.93***
Forward Citations/Year				-0.11
Year 4 Maintenance Fee Paid[284]				-5.25***
Backward Citations				0.01
Industry[285] (chemical)		24.73***	5.26**	11.96***
Industry (computer)		1.14	2.78	4.41**
Industry (pharmaceutical)		35.74***	16.19***	26.68***
Industry (mechanical)		17.15***	8.48***	9.91***
Industry (other)		35.32***	25.02***	25.58***
R^2	0.09	0.12	0.12	0.17
N	24,116	23,944	23,944	19,153

282. The dependent variable was log transformed to improve normality. CORNELL STATISTICAL CONSULTING UNIT, INTERPRETING COEFFICIENTS IN REGRESSION WITH LOG-TRANSFORMED VARIABLES 3 (2012), https://www.cscu.cornell.edu/news/statnews/stnews83.pdf.

283. *Id.*

284. Dummy variable set at 1 if the fee was paid and 0 if not. For this variable, patents granted in 2012 or later were excluded because this Article's analysis was done before the maintenance fee was due.

285. The electrical industry was used as the reference sector.