HERE-, THERE-, AND EVERY WHERE-: EXPLORING THE ROLE OF PRONOMINAL ADVERBS IN LEGAL LANGUAGE

David Chandler

Follow this and additional works at: https://scholarsarchive.byu.edu/studentpub_uht

BYU ScholarsArchive Citation
Chandler, David, "HERE-, THERE-, AND EVERY WHERE-: EXPLORING THE ROLE OF PRONOMINAL ADVERBS IN LEGAL LANGUAGE" (2024). Undergraduate Honors Theses. 392.
https://scholarsarchive.byu.edu/studentpub_uht/392

This Honors Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Undergraduate Honors Theses by an authorized administrator of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.
HERE-, THERE-, AND EVERY WHERE-: EXPLORING THE ROLE OF
PRONOMINAL ADVERBS IN LEGAL LANGUAGE

by
David Chandler

Submitted to Brigham Young University in partial fulfillment
of graduation requirements for University Honors

Linguistics Department
Brigham Young University
April 2024

Advisor: Brett Hashimoto
Honors Coordinator: Don Chapman
ABSTRACT

HERE-, THERE-, AND EVERY WHERE-: EXPLORING THE ROLE OF PRONOMINAL ADVERBS IN LEGAL LANGUAGE

David W. Chandler
Linguistics Department
Bachelor of Arts

Legal language is distinctive and sometimes difficult to understand. One type of word that appears to be part of so-called ‘legalese’ is the pronominal adverb (PA). The purpose of this thesis is to examine the prevalence of PAs in legal registers compared to more general registers of contemporary American English to determine the extent to which these words are distinctly legal. The study will also explore why different types of PAs may be (in)frequent in specific legal registers to better understand their use. The frequency of PAs was extracted from corpora that are designed to represent six registers of English (3 legal—contracts, statutes, and court opinions; 3 non-legal—academic articles, news articles, and TV/movie transcripts). The rate of occurrence of PAs per text were then compared across registers using a non-parametric analysis of variance including the Kruskal-Wallis test and the Dunn post-hoc test. Results indicate that there is a significant difference between the frequency of PAs in non-legal and legal registers of English. The post-hoc tests also reveal that there is also a significant difference in the frequency of PAs in each legal register. A closer examination of the PAs considered both individually as well as grouped by locative adverb (i.e., here-, there-, and where-).
indicates that some PAs are also more distinctive to certain legal registers. This study opens the discussion as to the utility and necessity of PAs in legal language and provides suggestions for legal writers on how to remove these words without inhibiting clarity or effectiveness.
ACKNOWLEDGMENTS

A special thanks to my mentor Brett Hashimoto for his contribution to the research, writing, and editing process. I would also like to thank Ute Römer and Jesse Egbert for being wonderful guest editors, Don Chapman and Dane Thorley for helpful input at early stages of this project, and Earl Brown for help with data visualization.
# TABLE OF CONTENTS

Title ...................................................................................................................................................... i
Abstract ................................................................................................................................................ iii
Acknowledgments ................................................................................................................................. vi
Table of Contents ................................................................................................................................. viii
List of Tables and Figures ................................................................................................................... X

I. Introduction ...................................................................................................................................... 1
II. Literature Review ........................................................................................................................... 1
III. Methodology ............................................................................................................................... 4
IV. Results and Discussion .................................................................................................................. 7
V. Conclusion ....................................................................................................................................... 21
References .......................................................................................................................................... 26
Appendix 1 .......................................................................................................................................... 29
Appendix 2 .......................................................................................................................................... 31
Appendix 3 .......................................................................................................................................... 33
# LIST OF TABLES AND FIGURES

TABLE 1: Legal and non-legal register names and relevant information about them......5

TABLE 2: A comprehensive list of PAs separated into *here*-, *there*-, and *where* types....7

TABLE 3: Pairwise comparisons of register for all PAs................................................8

TABLE 4: Pairwise comparisons of register for *here-* PAs..............................................13

TABLE 5: Pairwise comparisons of register for *there-* PAs...........................................16

TABLE 6: Pairwise comparisons of register for *where-* PAs............................................19

TABLE 7: Differences in frequency of PA types across legal and non-legal registers... 20

TABLE 8: Summary of functions of PA types................................................................. 22

FIGURE 1: The total PA distribution per million words by register.................................... 8

FIGURE 2: The *here-* PA distribution per million words by register................................13

FIGURE 3: The *there-* PA distribution per million words by register.............................. 16

FIGURE 4: The *where-* PA distribution per million words by register............................. 19
I. Introduction
Both legal scholars and linguists have characterized legal language as technical, dense, archaic, vague, opaque, or complex (see, e.g., Bednarek, 2014; Geng, 2018; Wydick, 2005; Tiersma, 1999; Schriver, 2017; Waltl & Matthes, 2014). Because of these characteristics, legalese or “traditional legal prose” has been demonstrated to be difficult to comprehend by the lay audience to the extent that it can feel like another dialect or language, “traditional legal prose” is oftentimes somewhat pejoratively referred to as “legalese” (Benson & Keller, 1986, p. 301). Researchers have observed that legal language seems difficult to ordinary citizens, and this has created criticism of legal writing practices over the last couple of decades (e.g., Geng, 2018; Goźdz-Roszkowski, 2011; Wydick, 2005; Tiersma, 1999). The general push to simplify and clarify highly difficult language is called the Plain English Movement (Wydick, 2005; Williams, 2007; Adler, 2012; Schriver, 2017).

Although legalese has been systematically used in most legal contexts for centuries, some scholars have recognized the disadvantage that lay citizens have in understanding the laws they must abide by and the language in which professional agreements are made. The U.S. Supreme Court has also held that “laws regulating person or entities must give fair notice of what conduct is required or proscribed” (FCC v Fox Television Stations, Inc., 2012). Previous empirical evidence noted by Benson and Kessler (1986) indicates that the average person has difficulty understanding jury instructions (Benson, 1984; Elwork et. al, 1982; and Charrow & Charrow, 1979), consumer contracts (Davis, 1977), and statutes (Benson, 1984). In addition, in quoting some of Benson’s previous work, Benson and Kessler (1986) asserted that readability formulas predict that “a person of average education would not be able to understand voters’ pamphlets, ballot propositions, statutes, standard contract forms, trust clauses, government notices, or other typical legal documents for consumers” (p. 302).

The fact remains that considerably more empirical research is needed into the linguistic description of legal language in order to determine precisely why it is so different, why it is so hard, and if certain aspects of legal language are unnecessary or needlessly complex and should be changed. The present study addresses this need by examining an aspect of English legal language that has been described as potentially problematic by previous research: pronominal adverbs (Osminkin, 2020). By adding to the emerging research on specific aspects of legal language, this research hopes to understand specific complexities contained in the use of pronominal adverbs in legal language. By so doing, I hope to aid in the determination of what changes to legalese could be made to satisfy proponents and critics of the Plain English Movement.

II. Literature Review
   Legalese effects on the reader
   Being unable to understand legal language takes a particular toll on people who speak English as a second language. There are already resources available that are attempting to bridge this gap not just for L2 (second language) English speakers generally, but for L2
lawyers, businessmen, paralegals, secretaries and researchers as well (Breeze, 2015; Brown and Rice, 2009). Similarly, translators are tasked with presenting difficult legal constructs in a variety of languages. Some researchers in translation studies have focused on understanding and translating legalese for this purpose (e.g., Osminkin, 2020; Geng, 2018).

Another significant issue is that most people are surrounded by legal language, sometimes without realizing it. Hartley (2000) points out that in a given week, the average person is probably going to come in contact with some of the following texts that contain legal language:

“All lottery tickets, food labels, bus/train/air tickets, regulations concerning the use of libraries/sports facilities, telephone/gas/electricity/water bills, text on medicine bottles, patient information leaflets, informed consent forms, income tax forms, hire purchase schemes/loans/overdrafts, guaranties and warranties, financial information, the acquisition of bonds, shares, unit trusts etc., conditions of service, copyright agreements, legal contracts, and wills and testaments” (p. 4).

Although the language complexities in this list may vary, the point is that people are more exposed to legalese than they may realize.

While plain English supporters call for the simplification of legal language, there are also critics of the Plain English Movement’s desire to rewrite legal discourse. These critics claim that legal language has historically been used for specific purposes that have been discussed and defined in court for centuries, and simplifying the language will potentially create more issues rather than solving them (see Stephenson, 2017 for a summary). Additionally, some argue that plain language recommendations given thus far have failed to eliminate confusion or reduce complexity and are not sufficiently based on empirical investigation or linguistic descriptions (e.g., Ződi, 2019; Barnes, 2010). As a result of these conflicting views, more work is needed to investigate pronominal adverbs and their role in legal language.

**Pronominal adverbs**

A pronominal adverb (PA), such as hereby, thereupon, or whereof, consists of a locative adverb (here, there, where) and a preposition (e.g., by, to, for, in, unto, with, etc.). These words are compound adverbs as well as “adverbs of place” meaning they “can show position, direction, or distance” (Biber et al., 1999, p.552). In a multi-register corpus-based keyness analysis, Goźdź-Roszkowski (2011) found that some registers of legal language made frequent use of PAs, especially legal contracts. However, the study does not speak specifically to their relationship to legal language as a whole, especially in comparison to non-legal registers.

More recently, Osminkin (2020) provided a detailed explanation of the history and grammar of PAs within British and American English business contracts. This article concludes that “‘here’ and ‘this’ are semantically associated with the area in which the speaker is situated whereas ‘there’ and ‘that’ refer to some other place situated outside
the speaker’s area.” (p.65) With this in mind, the reader is encouraged to view a PA like *herein* as “in *this* (name of the deictic center or current document) and to view a PA like *therein* as “in *that* (name of document reference outside the deictic center)”. Osminkin (2020) continues on to explain that “where” is interrelated to “which” and that “the where- connectors relate the clause to its antecedent in a more abstract and general manner than those based on here- and there-.” (p.76) Similar to the *here*- and *there*-construction, a PA such as *wherein* could be equally understood to mean “in which”. As Osminkin (2020) mentioned, the *where*-PAs perform the function of connecting a clause to an antecedent and do not behave deictically like the *here*- and *there*-PAs. Finally, the article identifies a number of PAs such as *therefore, thereupon,* and the *where*-PAs as behaving more like conjunctions especially in certain idiomatic expressions such as “Now, therefore” and “In witness whereof”.

The difficulty of understanding is made apparent by a number of court cases in which the meaning of certain PAs has been at issue. For instance, in RCJV Holdings, Inc. v. Collado Ryerson (2014), RCJV Holding Inc. (“RCJV”) had an agreement with Collado Ryerson, S.A. de C.V. (“Collado”) that Collado would pay $2,655,000 to RCJV by virtue of a promissory note by the day April 15th, 2011, which Collado failed to pay. However, Collado claimed that the subordination provision of the note stated that paying RCJV before the Natixis Credit Facility would cause a default which would in turn violate the third condition of the provision, all of which must be met to satisfy the provision. RCJV argued that the default would not counter the condition due to a “parenthetical exception” that states “(it being understood that this clause (iii) shall not apply to any covenant added to the Natixis Credit Facility after the date *hereof* expressly prohibiting payment of this Note).” The court found *hereof* ambiguous as to whether it referred to the date of the note or the date of the credit facility agreement. The first interpretation favors Collado by not satisfying the parenthetical exception while the second interpretation would favor RCJV by satisfying it. (RCJV Holdings, Inc. v. Collado Ryerson, S.A. de C.V., 2014). Ultimately, the ambiguity of the scope of the PA played a significant factor in the settlement of this case which might have been avoided with different verbiage.

Likewise, Mellinkoff (2004) provides a similar example with the PA *herein* in discussing a case that came before the US Supreme Court. As Mellinkoff (2004) puts it,

> In a typical situation, interpreting ‘Except as herein expressly provided,’ a trial court said that *herein* meant the whole statute. An intermediate panel of three appellate judges thought it ‘manifest’ that ‘herein means in this section.’ Four supreme court judges who held, ‘here it is clear that the word refers to the entire act’ could not convince the colleagues that it was so.” (p. 315).

Additional court cases revolving around the interpretation of PAs include Atlantic Fish Spotters Ass’n. v. Evans., 2003, Nichols v. Stat Radiology Medical Corporation, 2022, and Lisle Corp. v. A.J. Manufacturing Co., 2005. These court examples demonstrate how PAs can be confusing to the lawyers and judges as well as the layperson and all would benefit from a better understanding of how PAs function.
The present study
This purpose of this study was to empirically determine the distribution of PAs across legal and non-legal registers of American English in order to understand how and why they are being used. This was accomplished through a corpus-based analysis of the use of PAs across three legal registers and three non-legal registers of American English. Although some have claimed that PAs are regarded as almost exclusively legal English (e.g., Mellinkoff, 2004; Tiersma, 1999), I know of no study that empirically demonstrates that PAs are more frequent in legal registers.

This project adds to Osminkin’s (2020) study of PAs in that more legal registers are used, a more detailed grammatical analysis is employed, more representative data sets are examined, and the analysis focuses on identifying possible issues that PAs can cause and offers suggestions on how to avoid those issues. In short, this study attempts to answer the question:

Research Question: To what extent are PAs distinctive to three legal registers as compared to three general registers of American English?

III. Methodology
Corpora
In the present study, the frequency of different types of PAs was extracted from six registers of English (3 legal; 3 non-legal). The non-legal registers of TV/movie scripts, news articles, and academic articles come from the Corpus of Contemporary American English (COCA) (Davies, 2009). One justification for these registers is that they are some of the most studied registers using corpus methods, and are therefore, quite well documented. News and academic articles appear all the way back in the Brown Corpus, in the British National corpus, in the Longman Corpus, and COCA, and are frequently used as a point of reference for describing variation across situational contexts. Conversational speech is also in all of these corpora, except in the COCA, the spoken register is more formal because it is broadcast and many of the texts are in an interview format. Some research has demonstrated that TV/movie dialogue has many of the characteristics of authentic speech and is also a widely studied register in its own right (see, e.g., Van Heuven et al., 2014; New et al., 2007; Paetzold and Specia, 2016; Quaglio, 2009; Bednarek, 2018). Therefore, I use this register because it is spoken-like, interactive, and a less formal register--and perhaps therefore, useful for understanding everyday language use. The legal registers are legal contracts from the Corpus of English Business Contracts (COEBC), legal statutes from the Corpus of Current US Code (COCUSC), and Supreme Court opinions from the Corpus of US Supreme Court Opinions (COSCO-US). These legal registers are important to this study because they are perhaps more important for ordinary people to understand. The COCUSC contains the entirety of statutory law in the US, and the COSCO-US contains virtually all case law by which all US citizens are expected to abide. Similarly, the COEBC contains a variety of agreements that citizens may have to comprehend. The design of the COEBC underwent iterative stages of domain analysis, stratified sampling, and resulted in a large and representative corpus of contracts (see, Author, forthcoming). The purpose of using these
corpora is to have a variety of legal and non-legal English registers from which to compare PA frequencies. A description of each register is provided in Table 1.

Table 1. Legal and non-legal register names and relevant information about them.

<table>
<thead>
<tr>
<th>Register</th>
<th>Time period</th>
<th>Corpus</th>
<th>Population description</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>TV/Movies</td>
<td>1990-2019</td>
<td>COCA</td>
<td>American English transcripts from scripted television shows and movies</td>
<td>A stratified sample of subtitles from OpenSubtitles.org</td>
</tr>
<tr>
<td>Talk-news</td>
<td>1990-2019</td>
<td>COCA</td>
<td>American English unscripted talk-news programs</td>
<td>Sampled from unscripted TV and radio talk-news programs across the US</td>
</tr>
<tr>
<td>News articles</td>
<td>1990-2019</td>
<td>COCA</td>
<td>American English newspaper articles</td>
<td>Sampled from newspapers across the US with a good mix between different sections of the newspaper</td>
</tr>
<tr>
<td>Academic articles</td>
<td>1990-2019</td>
<td>COCA</td>
<td>American English peer-reviewed academic journal articles</td>
<td>Stratified sample to cover all academic domains as outlined in the Library of Congress classification system; sampled from more than 750 American publisher peer-reviewed academic journals</td>
</tr>
<tr>
<td>Contracts</td>
<td>2000-2022</td>
<td>COEBC</td>
<td>US business contracts that are used by publicly traded companies</td>
<td>Stratified sample by 54 contract types from Security and Exchange Commission filings</td>
</tr>
<tr>
<td><strong>Statutes</strong></td>
<td>Update between the years 2015-2020</td>
<td>COCUS</td>
<td>US federal statutes</td>
<td>Contains the entire population of federal statutes from the US Code as published by the Office of Law Revision Counsel</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>58 texts; 50+ million words</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Opinions</strong></th>
<th>COSCO-US</th>
<th>Opinions written by the Supreme Court, which are also case/common law</th>
<th>Includes the entire population of opinions published by the Supreme Court through the 2017 term</th>
</tr>
</thead>
<tbody>
<tr>
<td>61,809 texts; 98+ million words</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1754-2021</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The Corpus of English Business Contracts contained sub registers categorized by contract purpose. A detailed list of COEBC sub registers can be found in Appendix A.*

**Corpus querying**

The research question was approached by performing a frequency-based analysis focused heavily on register (Biber et al., 1999). Custom Python scripts were used to identify PAs within the text of each register and calculate the normalized frequency for each word per text. Regular expressions were used to find every instance of PAs in the data. Further explanation about the queries can be found in Appendix B. Afterwards, each set of results returned were manually checked for accuracy to ensure a precision of 100%, and an accuracy analysis was conducted of 30 texts for recall resulting in a recall of 100%. A custom Python script also calculated normalized frequencies for each text for the following:

1. the total number of PAs
2. the *here*-*, *there*-, and *where*- PAs respectively
3. for each type of PA.

After the normalized frequency of either the total, the *here*-*, *there*-, and *where*- PAs, or each individual PA was calculated per text, the normalized frequency calculation for each text was then plotted on a boxplot grouped by register.

**Analysis**

Statistical tests for this study were all performed using SPSS (29.0) (IBM Corp, 2022). The normalized frequencies for each PA in each register were plotted into boxplots to visualize their use within and across register.

Because the data did not meet the assumptions of normality and homogeneity of variance, and due to the high number of outliers, Kruskal-Wallis tests were employed with Dunn post-hoc tests for register-wise comparisons. The outliers were values far beyond the average dispersion, and most of the outliers in this study were values much higher than the average dispersion. For each test, the independent variable was the register type while the dependent variable was the normalized frequency of the PA or PAs in question. The results were then further explored using a concordance analysis to determine why PAs
are more or less frequent between registers. These concordance lines allowed us to analyze why certain PAs were more or less common in legal registers.

IV. Results and Discussion
Comparison of all PAs across register
A list of PA types found in the data is contained in Table 2. There are 22 here- connector PA types, 25 there- connector PAs, and 17 where- connector PAs. Together, that amounts to 64 PAs.

Table 2. A comprehensive list of PAs separated into here-, there-, and where- types.

<table>
<thead>
<tr>
<th>Here (22 total)</th>
<th>There (25 total)</th>
<th>Where (17 total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hereabove</td>
<td>Thereabout</td>
<td>Whereabout</td>
</tr>
<tr>
<td>Hereafter</td>
<td>Thereabouts</td>
<td>Whereafter</td>
</tr>
<tr>
<td>Herebefore</td>
<td>Thereafter</td>
<td>Whereafter</td>
</tr>
<tr>
<td>Hereby</td>
<td>Thereagain</td>
<td>Whereby</td>
</tr>
<tr>
<td>Herefor</td>
<td>Therealong</td>
<td>Wherefor</td>
</tr>
<tr>
<td>Herefore</td>
<td>Thereat</td>
<td>Wherefore</td>
</tr>
<tr>
<td>Herefrom</td>
<td>Thereby</td>
<td>Wherefrom</td>
</tr>
<tr>
<td>Herein</td>
<td>Therefor</td>
<td>Wherein</td>
</tr>
<tr>
<td>Hereinabove</td>
<td>Therefore</td>
<td>Whereof</td>
</tr>
<tr>
<td>Hereinafter</td>
<td>Therefrom</td>
<td>Whereto</td>
</tr>
<tr>
<td>Hereinbefore</td>
<td>Therein</td>
<td>Whereupon</td>
</tr>
<tr>
<td>Herewith</td>
<td>Thereinbefore</td>
<td>Wherewithal</td>
</tr>
</tbody>
</table>

Figure 1 contains a box and violin plot that shows the rate of occurrence of all PAs within each of the six legal and non-legal registers per million words. Because the distributions are non-normal, it is perhaps most useful to compare the median rate of occurrence across groups as an indicator of central tendency. Other descriptive statistics are provided in Appendix C.

The median for each register was as follows: TV/movies = .00; news = .00; academic = 224.20; statutes = 961.40; opinions = 1800.18; contracts = 9527.51. The result of the non-parametric ANOVA for all PAs was F(5, 171,407) = 105160.06, p < .001, eta² = 0.61, which indicates a large effect size. Post-hoc results indicate that there was a significant difference between all registers except for between opinions and statutes.
The three legal registers had much higher medians than the three non-legal registers. The lowest median in the non-legal register (statutes) was four times more than the highest median of the non-legal register (academic). The contracts register had the highest median which was five times more than the median of the next highest register.
(opinions). The median for the TV/movies and news registers were both zero which indicates that PAs are not used frequently in those registers.

**Functions of PAs**

One explanation for why PAs are used more often in legal language is that legal language implements a particular use of deixis in relationship to the immediate part of the text. (i.e., textual reference). For example, in the TV and movies register of contemporary American English, reference to the text is highly unusual, and therefore, there is little need to say, “the protagonist will appear in the next scene hereof”. The purpose of this register is to engage the audience and bring a sense of reality to a series of events. PAs acknowledge the text as the deictic center which, if used in this the TV/movies register would remind viewers that they are not experiencing reality rather a representation of it. This is less appealing to viewers and would therefore be less useful for script writers. The news and academic registers may have reasons to reference themselves as the content is meant for reading. There can be utility in informing readers on the location or time of certain details with the text as the deictic center (i.e., “The results are provided on the next page hereof.” or “A clear description of the suspect has not heretofore been acquired.”) However, written non-legal registers use other linguistic means, such as conjunctive adverbs, (e.g., next; firstly) or direct reference to other parts of the prose. (e.g., in the previous section; further explanation is provided in Section 3.1.; see; Geng, 2018, p.113) that fulfill the same purpose and are generally more frequent (Biber, 2006; Biber et al., 1999)

Another reason why PAs appear so frequently in legal registers is because of their ability to not just make deictic reference intratextually (within the text), but also intertextually (outside the text), as in Example 1. This is particularly true of the there- and where- PAs. Here and elsewhere in examples, PAs are underlined and other important words to highlight discussed further in the prose are bolded.

1. this Agreement is a supplement to and in furtherance of the indemnification provided in the Bylaws and any resolutions adopted pursuant thereto, and shall not be deemed a substitute therefor nor to diminish or abrogate any rights of Indemnitee thereunder (Contracts, Indemnity Agreement)

In example 1, all three PAs refer to “the Bylaws”, a separate document. Since multiple references are being made to the same document, legal language may include PAs because they are able to make more references in fewer words. These legal language registers make frequent intertextual references because of the purpose of many legal registers. When legal agreements are drafted and statutes are created, these legally binding texts have the potential to add to, take away from, or contradict one another. It becomes necessary for that information to be communicated in the current text. Thus, intertextual references are made, and PAs assist in making those references especially when there is more than one reference to a given legal document within the same sentence or group of sentences (Osminkin, 2020; Mohammad et al., 2010).
PAs are also used to reference nominal elements, particularly reoccurring or lengthy nominal elements. PAs appear to be used in this way to avoid redundancy as shown in Example 1 or when the reference is a list or other large nominal construction such as a multiword phrase or clause as in Example 2.

2. That until the property acquired by the United States pursuant to this section has been cleared of the aforesaid reservations, the Secretary in his discretion shall furnish such protection thereover as may be necessary... (US Statues, United States Code Section 16)

In Example 2, thereover references “the property” acquired by the United States pursuant to this section.” Thus, a PA can potentially help to make a text less wordy. However, the scope of the reference thereover may not be readily apparent to the average reader who may mistakenly think that it refers to one of the several other nominal elements in the sentence. Osminkin (2020) notes that the one way to paraphrase some PAs is to switch the order of the pronoun and the preposition. In this example thereover could be changed to over there, that, or them using almost the same number of characters. Either way, though, there remains a pronominal element, which is inherently more likely to be ambiguous or vague than repeating the noun. Alternatively, the phrasing could be edited as “the Secretary in his discretion shall furnish such protection over the/such property.” There are useful alternatives, especially because plain language detractors commonly argue that it is not possible to be equally clear and concise by using non-legalalese (see, e.g., Stephenson, 2017). These alternatives are similar in terms of length and further testing could determine whether other criteria, such as increased reader comprehensibility or writer implementability could help determine which alternate version would be preferable. Overall, PAs can help avoid repetition especially when referents are lists or longer phrases, and additional exploration should consider the extent to which different types of nominal elements that PAs are commonly used in place of are more or less easily replaced by simple methods of rephrasing.

However, not all PAs are used for referential purposes. They are also used as performative markers (see Example 3) and as discourse connectors (see Example 4).

3. The Presidential Service Certificate (“Certificate”) is hereby continued, the design of which accompanies and is hereby made a part of this order. (United States Code Section 10)

4. IN WITNESS WHEREOF, the parties hereto have duly caused this Purchase Agreement to be executed as of the day and year first above written. (Contracts, Purchase and Sale of Business Agreement)

Note that these words do not have a specific referent as many other PAs do. Although the word hereby in Example 3 does refer to the document, the function of hereby is primarily to indicate that official action is being taken (see, Kurzon, 1986 for an extensive discussion). In other words, hereby indicates that the verb that it attaches to is explicitly performative as in Example 3, which is particularly useful in cases where the verb itself is
not an explicit performative verb (e.g., authorize, permit) (Kurzon, 1986). The performative marker function is almost exclusively noted with hereby, although others are possible (e.g., herefor, heretofore).

In Example 4, the PA in the phrase “in witness whereof” acts as a commonly occurring transition between the content of the legal document and the signature (as it occurs in 85.48% of the contracts). Again, in this case, the primary function is not to refer to a nominal element, but to organize parts of the document as a discourse connector. A common discourse marker in contemporary English is therefore, which means “for that reason”. However, in legal registers, other PAs also act as discourse markers, such as wherefore, thereupon, and whereupon. The word therefore in the context of “NOW, THEREFORE” (these words together tend to be capitalized) found in many of the examples in the contracts register (as it occurs in 66.96% of contracts) acts as a marker signifying a transition between the initial setting of terms, background, and definitions in the contract and the actual agreement itself. Wherefore has traditionally appeared in Supreme Court opinions when summarizing the petitioner’s “prayer for relief” or part of the complaint where a plaintiff states the remedies it is seeking from the court. In all of these cases, PAs functioning as discourse connectors are almost exclusively used to connect the same parts of legal documents, which make them genre features because they are not pervasive, only occurring in specific parts of a text, often at the beginning or end and appear to be as conventional as they are functional (see, Biber & Conrad, 2019, p. 7). Because they are only genre features, it should be possible to substitute them for other conventionalized phrases.

In non-legal registers, the data indicates that PAs do exist but are far, far less frequent. In the non-legal registers, it appears that these words are almost exclusively used in referencing legal or business interactions as in Example 5 from the TV and movies register.

5. Jury is instructed to disregard any evidence heretofore presented regarding those matters. (COCA, TV/Movies, Gosnell: The Trial of America's Biggest Serial Killer)

The news register often quotes legal agreements and statutes. Example 6 is particularly enlightening because not only does it report the government’s attempt to promote plain English, but it also uses elements of legal language to poke fun at legal language as a whole. Notice that the author uses three PAs to get that point across. The author caricaturizes legalese, specifically using identified PAs as a specific feature with which legal language could be caricaturized.

6. The federal government is rolling out a new official language of sorts: plain English. # That's right: Pursuant to regulations promulgated thereunder and commencing in accordance with a statute signed herein by President Barack Obama, the government shall be precluded from writing the pompous gibberish heretofore evidenced, to the extent practicable. (COCA, News, Associated Press)
Another observation made in the TV and movies register was that PAs were also used to represent historical forms of English as in Example 7, which comes from an adaptation of Shakespeare’s *Hamlet*. Some of these examples are found authentically in remakes of Shakespeare plays while some were inserted into the script to merely give an impression of the time period.

7. Tomorrow I shall beg your leave to your kingly eyes, where I shall, first asking your pardon *thereunto*, recount the occasion of my sudden and more strange return. (COCA, TV/Movies, Hamlet)

In the academic register, the majority of examples found were in law reviews and in journals that discuss legal issues and politics, although examples were occasionally used in reference to older versions of English as well.

These examples show that when PAs appear in contemporary American English, they are often a result of quoting and imitating legal registers and older versions of English. The lack of PA usages outside of legal and historical context strongly indicate that outside of these contexts, PAs are seldom, if ever, used in contemporary American English. This further exemplifies the distinct nature of PAs in legal language and provides evidence that they are a distinctive, and likely problematic, feature of legalese.

Now that PAs have been analyzed generally, I will continue to investigate these words within their adverbial connector categories: *here-*-, *there-*-, and *where-*-

### 3.1.2 Comparison of *here-* connector PAs across register

The following boxplot contains the rate of occurrence per million words of the *here-* connector PAs within each of the six legal and non-legal registers. The results of the non-parametric ANOVA were $F(5, 171,407) = 64908.80, p < .001, \eta^2 = 0.38$, which indicates a large effect size. The median for each register was: TV/movies = .00; news = .00; academic = .00; statues = 224.56; opinions = .00, contracts = 5625.33. Post-hoc results indicate that there was a significant difference between all registers.
Figure 2. The *here* PA distribution per million words by register.

Table 4. Pairwise comparisons of register for *here* PAs.

<table>
<thead>
<tr>
<th>Sample 1-Sample 2</th>
<th>Test Statistic</th>
<th>SE</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>News-TV/Movies</td>
<td>2420.40</td>
<td>179.06</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>News-Academic</td>
<td>4591.81</td>
<td>173.06</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>News-Opinions</td>
<td>29724.31</td>
<td>166.42</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>News- Statutes</td>
<td>80128.31</td>
<td>3232.23</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>News- Contracts</td>
<td>91026.08</td>
<td>480.65</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>TV/Movies-Academic</td>
<td>2171.41</td>
<td>220.26</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>TV/Movies- Opinions</td>
<td>27303.91</td>
<td>215.09</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>TV/Movies- Statutes</td>
<td>77707.91</td>
<td>3235.10</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>TV/Movies- Contracts</td>
<td>88605.68</td>
<td>499.59</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Academic- Opinions</td>
<td>25132.50</td>
<td>210.12</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Academic- Statutes</td>
<td>75536.50</td>
<td>3234.78</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Academic- Contracts</td>
<td>86434.27</td>
<td>497.47</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Opinions - Statutes</td>
<td>50404.00</td>
<td>3234.42</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Opinions - Contracts</td>
<td>61301.77</td>
<td>495.20</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Statutes - Contracts</td>
<td>10897.77</td>
<td>3265.70</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Note. Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same. Significance values have been adjusted by the Bonferroni correction for multiple tests.

The highest medians in the *here* set of PAs were in the statutes register and the contracts register. The remaining registers had a median of 0 including the opinions register. The
median of the contracts register was the highest median being 25 times greater than the median of the next highest register (statutes). The median of all 3 non-legal registers was 0 which shows that here- connector PAs are not very frequent in those registers.

here- connector PAs apply deixis where the speaker’s location is the legal document itself. Essentially, Here- connector PAs create self-referential deixis. Examples (8) shows how references can be made to the content of the document, example (9) shows how references can be made to the time before or after the document is put into effect, and example (10) shows how a reference can be made to the physical document itself.

8. For all services to be rendered, payments to be made and costs to be assumed by you as provided in sections 2, 3, 4 and 5 hereof the Trust on behalf of the Fund shall pay you in United States Dollars on the last day of each month… (Contracts, Investment Management Agreement)
9. This Agreement represents the entire agreement between the Parties regarding the subject matter hereof and shall supersede all previous communications, representations, understandings, acknowledgements and agreements... whether heretofore or hereafter disclosed between the Parties. (Contracts, Nondisclosure Agreement)
10. the state of Wyoming... has caused these presents to be executed by its superintendent and the seal thereof to be hereunto affixed. (Opinions, State Highway Commission of Wyoming v. Utah Const. Co)

Contracts are often very long documents with a variety of sections and clauses, and these sections and clauses do not exist independently of one another. Rather, certain words, phrases, and conditions in one section could be discussed in multiple other sections that may be several pages away. Therefore, Here- connector PAs are likely more frequent in this register because of this need to frequently reference other sections in the text. In addition, contracts are written with specific explanations because there is no tolerance for ambiguity, and there may not be an opportunity for further explanation later on in the text. The result is these PAs that connect all these references together. However, these PAs may prove more problematic than helpful. The literature review of this study established that here- connector PAs also have the potential to be ambiguous. The word can reference the document itself or the section or paragraph the word is located in, and if the reader cannot easily locate the reference, they will likely not benefit from having been given the reference in the first place. Not being able to connect the information in the contract may be detrimental to the reader if he or she misses an important connection that puts the agreement in the other party’s favor. If the location of the referenced word, condition, or section was identified in addition to or in place of a PA, this may help the reader better navigate the references made in contracts.

Despite the disadvantages that PAs may cause to the reader, PAs persist as a common feature for self-referential deixis in contracts. Although archaic words like these are often justified as the product of the legal discipline’s hesitancy to change, perhaps there is more going on here. Contracts often begin as templates where the crucial information is copied and pasted from one contract to one with a similar purpose. This has ultimately made the
process faster, cheaper and more efficient (Adams 2008). With this in mind, if contracts are made in part by copying and pasting necessary elements, it is likely that the location of the reference will be different between the new contract and the template. PAs seem to provide enough vagueness that they can account for these location changes. Therefore, perhaps PAs have been advantageous to lawyers because they can still enhance efficiency without the hassle of specifically identifying the location of each reference.

However, a contract drafting manual also noted that this process does not exist without pitfalls, “Obviously, lawyers draft contracts by copying contract models … Uncritical copying is expedient, but it invariably results in you including in any new contract all sorts of deficient usages.” (Adams, 2008). PAs may help but cannot guarantee that each referent has been effectively referenced. Relying on that assumption may still cause errors or ambiguity. Though there is no certain solution to this dilemma, it is important for those drafting contracts to consider a balance between effective vagueness and the ability to help the reader comprehend the text.

As for the other registers, the non-legal registers do not really use self-references to this extent. While the academic register uses many references to other texts, there are less references to the text itself. This is probably due to the shorter length of academic articles compared to contracts, and sections are usually labeled by topic which creates a sufficient guide for references to be made. Like the non-legal registers, the median in the opinions register was 0. This likely reflects the fact that there is less of a need for self-referential deixis. Opinions have fewer sections to reference, and the names, laws, and other significant details that lead to the verdict are often restated as opposed to being referenced.

3.1.3 Comparison of there- connector PAs across register

The following boxplot contains the rate of occurrence per million words of the there-connector PAs within each of the six legal and non-legal registers. The results of the non-parametric ANOVA were F(5, 171,407) = 102107.22, p < .001, eta^2 = 0.60. The median for each register was: TV/movies – .00; news – .00; academic – 170.33; US statues – 799.31; opinions – 1464.13, contracts – 3112.74.
Figure 3. The *there* PA distribution per million words by register.

Table 5. Pairwise comparisons of register for *there* PAs

<table>
<thead>
<tr>
<th>Sample 1-Sample 2</th>
<th>Test Statistic</th>
<th>SE</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>News-TV/Movies</td>
<td>1800.13</td>
<td>283.92</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>News-Academic</td>
<td>37423.96</td>
<td>274.40</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>News-Opinions</td>
<td>75274.10</td>
<td>263.88</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>News-Statutes</td>
<td>76727.07</td>
<td>5124.99</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>News-Contracts</td>
<td>95060.39</td>
<td>762.12</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>TV/Movies-Academic</td>
<td>35623.83</td>
<td>349.25</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>TV/Movies-Opinions</td>
<td>73473.98</td>
<td>341.04</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>TV/Movies-Statutes</td>
<td>74926.95</td>
<td>5129.54</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>TV/Movies-Contracts</td>
<td>93260.27</td>
<td>792.15</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Academic-Opinions</td>
<td>37850.15</td>
<td>333.16</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Academic-Statutes</td>
<td>39303.12</td>
<td>5129.02</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Academic-Contracts</td>
<td>57636.44</td>
<td>788.79</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Opinions-Statutes</td>
<td>1452.97</td>
<td>5128.47</td>
<td>.777</td>
</tr>
<tr>
<td>Opinions-Contracts</td>
<td>19786.29</td>
<td>785.19</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Statutes-Contracts</td>
<td>18333.32</td>
<td>5178.07</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Note. Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same. Significance values have been adjusted by the Bonferroni correction for multiple tests.

Here, the legal register medians were much greater than the non-legal register medians with the lowest legal register median (statutes) being 4 times greater than the highest non-legal register median (academic). The contracts register had the highest median of the set being 2 times greater than the next highest legal register median (opinions).
Section 3.1.1 established that *there-* connector PAs can create intertextual references (see example 1), contextual references (see example 3), and serve as discourse markers. Contracts also utilize *there-* connector PAs more than the other registers because not only do the agreements in contracts depend on their individual parts, they also depend on other laws and agreements found in other documents. More intertextual references yield the potential for more PAs. The other two legal registers make intertextual references as well. Supreme Court opinions reference laws pertinent to the case as well as other cases where similar decisions were made. Sections of the US Code will also refer to other sections of the US Code as if they were separate entities. *There-* connector PAs are perhaps the most problematic group of PAs in terms of creating issues for the reader. Similar to *here-* connect PAs, *there-* connector PAs have the potential to be ambiguous. This can be seen in example 11 with a common contract ending.

11. IN WITNESS WHEREOF, the Depositor and the Owner Trustee have caused their names to be signed hereto by their respective officers *thereunto* duly authorized, all as of the day and year first above written. (Contracts, Trust Agreement)

As mentioned in the literature review, a general way to understand PAs is to switch the order of the preposition and the adverb and then replace the adverb with the reference indicated by the text (Osminkin, 2020). In example (11), *thereunto* would become “unto (there or that reference)”. In example 11, the Depositor and the Owner Trustee are signing their names to the contract, and they have been duly authorized by their respective officers. However, for a person unfamiliar with this common contract ending, they may wonder what purpose *thereunto* serves. Is the word *thereunto* referencing “the Depositor”, the “Owner Trustee”, or “their respective officers”? Although those all appear earlier in the sentence, there is not a clear indication that any of these parties are being referenced by the word *thereunto*. In this case, *thereunto* is a distal reference to the act of signing a document. The sentence without the PA *thereunto* could read, “[they] have caused their names to be written here duly authorized by their respective officers to do so” in which the phrase “to do so” indicates that the referent is an action and not an object. This example shows that understanding this particular sentence for the first time would be confusing and take time to reason through.

In addition to ambiguity, *there-* connector PAs have the potential to cause issues in the cohesion of the text. Cohesion decreases when there is confusion between a deictic reference and its referent. This occurs in legal writing when there is too much distance between a PA and its referent and when there are other objects in between. Example 13 will be used to demonstrate this principle.

12. Amendments and Binding Effect. This Agreement and the Undertaking and the rights and duties of Indemnitee and the Company hereunder and thereunder may not be amended, modified or terminated except by written instrument signed and delivered by the parties hereto. This Agreement is and shall be binding upon and shall inure to the benefits of the parties *thereto* and
their respective heirs, executors, administrator, successors and assigns.

(Contracts, Indemnity Agreement)

The contract says that “This Agreement is and shall be binding upon and shall inure to the benefits of the parties thereto (to that referent)”, so what is the referent of thereto? For someone who is not familiar with PAs or legal language, they might think about the possible scenarios:

- This Agreement is and shall be binding upon and shall inure to the benefits of the parties to this Agreement?
- This Agreement is and shall be binding upon and shall inure to the benefits of the parties to the written instrument signed and delivered by the parties hereto?
- This Agreement is and shall be binding upon and shall inure to the benefits of the parties to this agreement and the undertaking and the rights and duties of Indemnitee and the Company hereunder and thereunder.
- This Agreement is and shall be binding upon and shall inure to the benefits of the parties to this agreement and the undertaking.

It is possible that a reader would need to review the sentence multiple times to understand the grouping done in sentence three. However, once the reader can see that this Agreement, the Undertaking, and the rights and duties of both parties under both documents have been grouped together into one, they will see that that is the reference for thereto. The reference is possible to find, but to get there, the reader must consider multiple options that come before it. Sifting through multiple references is what makes this incohesive, but what causes the reader to consider these options is that the referent is perhaps too far away from the PA. One way to improve this sentence would be to delete the PA and simply reinsert the intended reference. Although this adds extra wording to the contract, the edited statement (example 13) becomes clearer than the original.

13. (edited) “Amendments and Binding Effect. This Agreement and the Undertaking and the rights and duties of Indemnitee and the Company hereunder and thereunder may not be amended, modified or terminated except by written instrument signed and delivered by the parties hereto. This Agreement is and shall be binding upon and shall inure to the benefits of the parties to This Agreement and said Undertaking and the parties’ respective heirs, executors, administrator, successors and assigns.”

3.1.4 Comparison of where-connector PAs across register
The following boxplot contains the rate of occurrence per million words of the where-connector PAs within each of the six legal and non-legal registers. The results of the non-parametric ANOVA were $F(5, 171,407) = 145719.38, p < .001, \eta^2 = 0.27$. The median for each register was: TV/movies – .00; news – .00; academic – .00; US statues – 15.46; opinions – .00, contracts – 173.18.
Figure 4. The *where* PA distribution per million words by register.

Table 6. Pairwise comparisons of register for *where* PAs

<table>
<thead>
<tr>
<th>Sample 1-Sample 2</th>
<th>Test Statistic</th>
<th>SE</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>News-TV/Movies</td>
<td>363.421</td>
<td>155.267</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>News-Academic</td>
<td>6962.036</td>
<td>150.063</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>News-Opinions</td>
<td>17287.712</td>
<td>144.308</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>News-Statutes</td>
<td>66013.008</td>
<td>2802.701</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>News-Contracts</td>
<td>75415.026</td>
<td>416.779</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>TV/Movies-Academic</td>
<td>6598.615</td>
<td>190.993</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>TV/Movies-Opinions</td>
<td>16924.291</td>
<td>186.505</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>TV/Movies-Statutes</td>
<td>65649.587</td>
<td>2805.190</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>TV/Movies-Contracts</td>
<td>75051.605</td>
<td>433.203</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Academic-Opinions</td>
<td>10325.676</td>
<td>182.195</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Academic-Statutes</td>
<td>59050.972</td>
<td>2804.907</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Academic-Contracts</td>
<td>68452.990</td>
<td>431.365</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Opinions-Statutes</td>
<td>48725.296</td>
<td>2804.605</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Opinions-Contracts</td>
<td>58127.314</td>
<td>429.397</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Statutes-Contracts</td>
<td>9402.018</td>
<td>2831.729</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

*Note.* Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same. Significance values have been adjusted by the Bonferroni correction for multiple tests.

The highest medians in the *where*-connector set of PAs were in the statutes register and the contracts register. The remaining registers had a mean of 0 including the opinions register. The median of the contracts register was again the highest median being 11
times greater than the median of the next highest register (statutes). This data also shows that *where*- connector PAs are generally the least frequent types of PAs.

*Where*- connectors PAs are perhaps less likely to cause issues in ambiguity. For example, the word, “wherein” is taken to mean “in which”. The “which” is almost always specified by the object that immediately precedes it, and thus there is no confusion as to the intended referent. However, because these are the least frequent and hardly ever appeared in the non-legal registers, it may not be common knowledge that where- takes on the meaning of which when reading PAs, in which case it might be better for a legal writer to just use the phrase “in which” instead of wherein. This relativizing function of PAs appears to be particular to legal registers, especially statutes and contracts.

### 3.1.5 Analysis of individual PAs

Now that PAs have been discussed generally as distinct between legal and non-legal registers, this study also considered each individual PA found and verified their statistical differences between legal and non-legal registers. The table below categorizes the PAs by those that were significant between all legal and non-legal registers, the PAs where only one or two legal registers was significant between the non-legal registers, and the PAs that were not significant between legal and non-legal registers.

**Table 7. Differences in frequency of PA types across legal and non-legal registers**

| Some significantly different | Hereabove*, Herefor*, Herefrom*#, Hereinbelow*, Hereinunder*#, Hereeto##, Hereupon$, Thereabouts$, Thereagainst*, Thereinafter##, Thereinbefore$, Thereinto$, Thereout$, Therethrough#, Wherefore$, Whereto##, Whereunder#, Wherewith$ |
| None significantly different | Herebefore, Herefore, Heretobefore, Thereabout, Therealong, Whereabout, Whereafter, Whereat, Wherefor, Wherefrom, Whereover, Whereunto |

Note. These symbols indicate a significant difference between the specified legal register and all non-legal registers: # - Statutes, $ - Opinions, * - Contracts

There are a variety of PAs in each section, and there does not seem to be any immediate pattern to indicate why a certain word is statistically distinct between the legal and non-legal registers while the others are not. It could be that the reason that some PAs are not statistically different is because are generally very infrequent. However, for the the statistically significant PAs, part of the answer lies in the role of phraseology in legal language.
Phraseology is the idea that “Every native speaker has thousands upon thousands of multi-word units stored in memory” and asserts that when communicating with others, humans “rely on familiar chunks of language behavior” (Stubbs. 2001). PAs most distinct to legal registers tend to be those found frequently in these multi-word units. For example, three 2-grams for the word *herein* occurred more than a thousand times each without the legal registers. These 2-grams were “herein shall” found 4513 times, “herein provided” found 1964 times, and “herein contained” found 1538 times.

The fact that many significantly distinct PAs have frequent collocates supports the idea that legal language is written with collections of common phrases. In contracts this is based on the tendency for legal writers to borrow wording from similar texts to make texts that fit their purposes. As mentioned in the here- connector PA section, contracts are rarely written from scratch, so any multi-word units common in older contracts sometimes make their way into newer contracts. In the case of the PAs that are not distinctive to legal language, it seems that the author was not following a consistent pattern but using their knowledge of how to use PAs to create new ones. However, if everyone else is able to communicate ideas without the use of a certain PA, the legal writer should consider whether using a PA in that context is necessary to include or if restating the reference would be appropriate.

V. Conclusion
This study has explored the purpose and distribution of PAs across three legal and three non-legal registers and has ultimately concluded that PAs are distinctive to legal language. The data has also shown that PAs are most frequent in the contracts register due to the number of sections and supporting documents that contracts reference. On the other hand, the non-legal registers rarely use these words, and when doing so, they are often discussing legal topics or imitating older versions of English.

PAs are used in legal language to make self-references, contextual references, and intertextual references, and serve as discourse markers. The self-references and intertextual reference can apply to a physical text itself or the time before or after the text was written. A combination of here-, there-, and where- connectors and the various prepositions create PAs that facilitate these different types of references. The purpose, length, and type of document also affect how frequently certain PAs will be used, and the types of formulaic phrases employed in legal language have encouraged the continued use of PAs despite their absence in non-legal registers. See table 8 for a summary of the here-, there-, and where- PA functions.

Table 8. Summary of functions of PA types

<table>
<thead>
<tr>
<th>PA type</th>
<th>Deixis</th>
<th>Common functions in legal registers</th>
<th>Other possible functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>here-</td>
<td>exclusively proximal</td>
<td>Connects two local elements; performative marker</td>
<td>Discourse connector</td>
</tr>
<tr>
<td>there-</td>
<td>usually distal, but oftentimes proximal</td>
<td>Connects one local and one distal element; discourse connectors; connects two local elements</td>
<td>Performative marker</td>
</tr>
</tbody>
</table>
4.1.1 Suggestions
There are a number of potential issues that can cause PAs to be misinterpreted and a drafter of legal documents should consider whether each use of a PA will aid or hinder their reader from understanding the document they are drafting. These issues in ambiguity and cohesion have the potential to cause confusion for the lay person and the lawyer alike. As has been discussed above, there are a few general suggestions that, when applied, can decrease the level of ambiguity and incohesiveness that PAs create. These suggestions will be revisited below along with good examples found from the corpora.

It is important to note that as a general rule, linguists avoid being prescriptive. Many linguists believe being prescriptive impedes language variation and change, and this type of change is valued in the field of linguistics. However, legal documents seem to be an important place to control legal language, and most law schools are uniform in prescribing the usage of legal language. These suggestions are made in hopes of aiding legal language in prescriptive choices that have the potential to make legal documents easier to understand and reduce ambiguity.

4.1.2 Here- Suggestions
The first suggestion involves the use of the here- connector PAs. As mentioned in the introduction, words like herein, hereof, and hereunder are ambiguous in that the here- could refer to “here in this section”, “here in this paragraph”, “herein this condition” or “here in this document”. This is an important distinction considering the fact that lawyers like to keep conditions or sections separate from each other like with the phrase “notwithstanding the foregoing”. In order to prevent any ambiguity, I suggest that here- connector PAs only be used to refer to the entire document as a whole. This is more of an issue for contracts than Supreme Court opinions and statues considering contracts have more functional sections, but this rule should still be applicable to most legal registers regardless. For one to personally apply this principle may not completely eliminate the ambiguous nature of a PA, however, and it may be beneficial for the legal document to include a statement regarding the referential nature of here- PAs. A number financing and security agreements in the contracts register already contain a clarifying statement about here- connector PAs in their “accounting terms and other definitional provisions” and other similar sections. Example 14 is an example of said statement.

14. The words "hereof", "herein" and "hereunder" and words of similar import when used in this Agreement shall refer to this Agreement as a whole and not to any particular provision of this Agreement, Section, subsection, schedule and exhibit references are references to sections or subsections of, or schedules or exhibits to, as the case may be, this Agreement unless otherwise specified. (Contracts, Financing and Security Agreement)

It is recommended that any contract with here- PAs incorporate similar statements wherever they would fit most appropriately. Not only does this eliminate ambiguity, but
it may also help the average person understand the nature of these words before they suddenly come across it later in the legal document.

4.1.3 *There*- Suggestions
The next two suggestions involve the use of *there*- connector PAs. This type of PA creates more issues because the reference is a specific word or group of words in the text as opposed to the references to paragraphs, sections, or the entire document. Therefore, the reader is tasked with looking back and finding the specific reference being made. As mentioned in the introduction, this affects the reader’s comprehension of the text.

First, it is recommended that the referent be as close to the reference as possible so that the reader need not sift through the other objects in the sentence to know what the PA represents. Examples 15 and 16 will demonstrate this principle.

15. a United States Senator from said State shall be elected by the people thereof (USC02)

16. for the said sum of $10,505.20 1/2, with interest thereon from the date when the same was received.” (Opinions, The United States, use of James Mackey et al. Plaintiffs in Error, v. Richard S. Coxe)

Notice how there is only one other object between the PA and its referent. In example 15, the *there* in “thereof” references the “said State” while the *of* in “thereof” establishes the relationship between the referent and the intermediary object which is “the people” in this case. These things considered, it is apparent that the Senator shall be elected by the people of said state. A good rule of thumb, then, is to avoid putting additional objects between the PA and the referent in order to avoid providing the reader with more possible referents than necessary.

Second, in cases where the space between a PA and its referent contains other objects or is otherwise separated by one or more additional clauses, it would be recommended to delete the PA and replace it with the referent itself preceded by the intended preposition. PAs can be useful in shortening the length of complex sentences, especially when the PA references multiple words at a time, however, legal writers might be sacrificing the cohesiveness of the text by including PAs instead of the actual referent. To accomplish this, one can simply insert a word such as *the, that, these/those,* or *said* in between the preposition contained in the PA and the referent the PA referenced which would read, “-preposition- -helper word- -referent-” as shown in the examples below.

17. this agreement … is in furtherance of the indemnification provided in the Bylaws and any resolution pursuant thereto, and shall not be deemed a substitute therefore nor to diminish … any rights of the Indemnitee thereunder (Contracts, Indemnitee Agreement)

18. (edited) this agreement … is in furtherance of the indemnification provided in the Bylaws and any resolution pursuant to that indemnification.
and shall not be deemed a substitute for that indemnification nor to diminish … any rights of the Indemnitee under that indemnification (Contracts, Indemnitee Agreement)

19. the reasons given for the rejection of such claims might apply equally to the same claims contained in the original patent, but with respect to such claims he was functus officio. his opinion thereon was but his personal opinion (McCormick Harvesting Mach. Co. v. C. Aultman & Co. et al. Same v. Aultman-Miller Co)

20. (edited) the reasons given for the rejection of such claims might apply equally to the same claims contained in the original patent, but with respect to such claims he was functus officio. his opinion on said claims was but his personal opinion (McCormick Harvesting Mach. Co. v. C. Aultman & Co. et al. Same v. Aultman-Miller Co)

4.1.4 Where- suggestions
Finally, some brief recommendations will be made for the where- PAs. The where- PAs do not tend to cause issues in ambiguity or cohesion in legal language due to the fact that the referent tends to immediately precede the PA as in the following examples.

21. Payment of the judgment having been refused, and it appearing that the debtor had no property wherewith to satisfy the execution, the judgment creditor, as plaintiff, instituted the present suit against the defendants as the sureties of the principal, counting on the said bond as the cause of action. (Supreme Court Opinions, Babbitt v. Finn)

In example 21, “wherewith” refers to the “property” which is right next to the PA. Therefore, the where- form PAs do not cause ambiguity issues so long as the referent continues to immediately precede the PA. Although these words might not cause ambiguity, their unfamiliarity may still cause confusion. Without knowing beforehand that “wherewith” can be understood as “with which”, readers may wonder if they missed a location and spend time trying to find something that is not there. Seeing as where-connector PAs like wherewith could easily be changed to two words (with which) without significantly changing the length of the sentence, the two-word construction should be considered since it is more familiar to the lay person and does not in any way change the meaning of the sentence.

4.1.5 Limitations and further research
This study acknowledges once again the lack of research on actual human comprehension of legal language that contains PAs. Furthermore, our suggestions have not been tested and are merely based on observations made about the data. I also recognize that the legal registers used here do not represent legal language as a whole nor do they represent other legal registers that could have been studied. Additional research on the effectiveness of our suggestions evaluated by test subjects using more legal registers could help determine the utility of these recommendations.
Author (forthcoming)
Atlantic Fish Spotters Ass’n. v. Evans. 321 F.3d 220 (1st Cir. 2003)


Nichols v. Stat Radiology Medical Corporation, No. 21-1560 (6th Cir. Mar. 24, 2022)


Appendix A. Contract Sub Registers
Business management contracts (BUS):
[500 contracts]
BUS_oper - operating agreement
BUS_part - partnership agreement
BUS_liab - limited liability company agreement
BUS_fran - franchise agreement
BUS_admin - administration agreement
BUS_manage - management agreement
BUS_trust - trust agreement
BUS_merge - agreement and plan of merger / plan of reorganization assignment of contract /assignment of contract rights and obligations / assignment of contractual rights / assignment of contracts / agreement for assignment of contract proceeds / collateral assignment of contracts / transfer and assumption agreement / transfer agreement
BUS_partic - participation agreement

Sales-related contracts (SALE):
[850 contracts]
SALE_purch - purchase agreement / purchase and sale agreement
SALE_bsale - bill of sale / bill of sale and assignment / bill of sale and assignment and assumption / bill of sale and assignment agreement
SALE_escrow - escrow agreement
SALE_asset - asset purchase agreement Outlines the terms of sale of an asset
SALE_for - forward purchase agreement
SALE_dep - deposit agreement
SALE_stock - securities purchase agreement / Stock purchase agreement
SALE_opt - stock option agreement / option agreement
SALE_res - restricted stock agreement
SALE_sub - subscription agreement
SALE_under - underwriting agreement
SALE_warr - warrant agreement
SALE_agency - sales agency financing agreement
SALE_inden - indenture
SALE_purchb - purchase and sale of business agreement / business purchase agreement / purchase of business agreement / sale of business agreement / sale of business / business sale agreement / business acquisition agreement / acquisition agreement
SALE_dist - distribution agreement
SALE_lic - license agreement / licensing agreement

Employment contracts (EMP):
[450 contracts]
EMP_change - change in control agreement
EMP_exec - executive employment agreement
EMP_gen - general employment contract / employment contract
EMP_ind - independent contractor agreement
EMP_non - non-compete agreement / non-compete agreement
EMP_serv - services agreement
EMP_con - consulting agreement
EMP_warr - guarantee agreement / warranty and indemnity agreement / warranty and indemnification agreement
EMP_stat - statement of work / statement of work and account satisfaction agreement / managed services statement of work / initial statement of work project description

Investment management contracts (INV):
[250 contracts]
INV_invri - investor rights agreement
INV_reg - registration rights agreement
INV_intrad - investment advisory agreement
INV_subad - sub-advisory agreement / investment sub-advisory agreement
INV_invman - investment management agreement

Leasing and loan financing contracts (LOAN):
[500 contracts]
LL_credit (odd) / LL_loan (even) credit agreement loan agreement / loan and security agreement
LOAN_fin - financing agreement
LOAN_bus - business financing agreement
LOAN_finsec - financing and security agreement
LOAN_sec - security agreement
LOAN_pool - pooling and servicing agreement
LOAN_loan - loan financing and servicing agreement
LOAN_mort - mortgage loan purchase agreement
LOAN_equip - equipment lease agreement
LOAN_prop - property lease / property lease agreement / master lease agreement

Business assurance contracts (ASSURE):
[150 contracts]
ASSURE_ind - indemnity agreement / Protects the first party from
ASSURE_non - nondisclosure agreement
ASSURE_sett - settlement agreement

Appendix B. Extended Methodology

The process began with a python script that would return the words found with here-, there-, and where- beginnings along with the file that the words were found in. The regular expressions used to return these words were \bhere\w+\b, \bthere\w+\b, and \bwhere\w+\b. These regular expressions were meant to find every word that started with letters h-e-r-e, t-h-e-r-e, or w-h-e-r-e to be sure that every PA was captured. Among the list of words returned, there were words that were not PAs such as heresy, hereditary, whereas, and wherever. These words were removed from the list. The program also returned words that contained misspellings and other text errors such as wherepon, herei, herel, therebec, and heretofore4. To fix this, the program was altered to find what text each error was in, then each text was opened and edited manually to fix the spelling errors. In the case where it was clear which word was meant to be in the text, the spelling error was corrected. However, in a small number of cases where the correction was not clear (e.g. should theren be corrected to therein or thereon?), the correction was not made. The words that ended with numbers turned out to be footnotes, so for these words, the footnotes were taken out. There were a number of words that looked very similar to PAs and might have been used for that purpose by the author of the text, but the PAs analyzed in this study were limited to words that contain one locative adverb, 1-2 prepositions, and no additional words or letters. The word was also required to behave as an adverb to be analyzed in this study. For example, Therein consists of the locative adverb there and the preposition in. Seeing as this meets the criteria (1 locative adverb and 1-2 prepositions), therein was analyzed in this study. Similarly, the word hereinafter contains the locative adverbs here and the prepositions in and after, therefore, hereinafter was used in this study. On the other hand, the word hereinelsewhere contains the locative adverb here, and the preposition in, but also contains the additional word elsewhere which is another adverb. Consequently, hereinelsewhere was not analyzed in this study. At that point, every PA in each register was identified, and the texts had been corrected to find every instance of these words. These corrections were made to ensure that the
program would count every PA in the corpora and to verify that the list of PAs possible in these corpora was complete. With the large amount of data collected by the python script, it was necessary to do a recall analysis to check that the program successfully counted every PA in each text. The recall analysis concluded that the computer program was successful in finding most if not all PAs.

The recall analysis also called to attention that not every text in the Corpus of Supreme Court Opinions in the United States register was completely captured and recorded. For example, "id": "552.US.442" was a text in the Supreme Court opinions register where no PAs were found because the text only contained the first few paragraphs of the text. In all circumstances where there were no PAs found in a US Supreme Court opinion, the opinion was opened and checked for completeness. In the cases where the file had not completely captured all the text in the opinion, that opinion was discarded.

With the most accurate data achievable with reasonable effort, the python script was given an entire list of PAs in the form of regular expressions. The script then opened each text in the corpus, searched for and counted each PA, calculated the normalized frequency for every possible PA in that text (even if a certain PA was not found in that text), then recorded that number to a .csv file with the rows being a new text and the columns being the individual PAs.
## Appendix C. Data Tables for PAs across register

### Table 3. Descriptive Statistics for the Total Number of PAs per Register.

<table>
<thead>
<tr>
<th>Registers</th>
<th>Medians</th>
<th>Means</th>
<th>SD</th>
<th>95% Confidence Interval</th>
<th>IQR</th>
</tr>
</thead>
<tbody>
<tr>
<td>TV/Movies</td>
<td>.00</td>
<td>30.02</td>
<td>111.45</td>
<td>[28.61, 31.43]</td>
<td>.00</td>
</tr>
<tr>
<td>News</td>
<td>.00</td>
<td>44.18</td>
<td>243.83</td>
<td>[42.86, 45.78]</td>
<td>.00</td>
</tr>
<tr>
<td>Academic</td>
<td>224.20</td>
<td>378.82</td>
<td>513.56</td>
<td>[372.58, 385.05]</td>
<td>579.12</td>
</tr>
<tr>
<td>Statutes</td>
<td>961.40</td>
<td>1351.82</td>
<td>1027.21</td>
<td>[1081.73, 1621.91]</td>
<td>900.58</td>
</tr>
<tr>
<td>Opinions</td>
<td>1800.18</td>
<td>2377.68</td>
<td>2367.37</td>
<td>[2350.39, 2404.97]</td>
<td>2469.36</td>
</tr>
<tr>
<td>Contracts</td>
<td>9527.51</td>
<td>10173.25</td>
<td>4739.62</td>
<td>[9994.39, 10352.10]</td>
<td>4658.59</td>
</tr>
</tbody>
</table>

### Table 4. Descriptive Statistics for the Total Number of *Here*- connector PAs per Register.

<table>
<thead>
<tr>
<th>Registers</th>
<th>Medians</th>
<th>Means</th>
<th>SD</th>
<th>95% Confidence Interval</th>
<th>IQR</th>
</tr>
</thead>
<tbody>
<tr>
<td>TV/Movies</td>
<td>.00</td>
<td>7.82</td>
<td>53.48</td>
<td>[7.14, 8.49]</td>
<td>.00</td>
</tr>
<tr>
<td>News</td>
<td>.00</td>
<td>2.25</td>
<td>51.57</td>
<td>[1.92, 2.59]</td>
<td>.00</td>
</tr>
<tr>
<td>Academic</td>
<td>.00</td>
<td>16.55</td>
<td>87.48</td>
<td>[15.49, 17.61]</td>
<td>.00</td>
</tr>
<tr>
<td>Statutes</td>
<td>224.56</td>
<td>357.89</td>
<td>414.04</td>
<td>[249.03, 466.76]</td>
<td>262.38</td>
</tr>
<tr>
<td>Opinions</td>
<td>.00</td>
<td>355.70</td>
<td>1024.39</td>
<td>[343.89, 367.51]</td>
<td>343.76</td>
</tr>
<tr>
<td>Contracts</td>
<td>5625.33</td>
<td>6563.64</td>
<td>3893.74</td>
<td>[6416.70, 6710.57]</td>
<td>4017.96</td>
</tr>
</tbody>
</table>

### Table 5. Descriptive Statistics for the Total Number of *There*- connector PAs per Register.

<table>
<thead>
<tr>
<th>Registers</th>
<th>Medians</th>
<th>Means</th>
<th>SD</th>
<th>95% Confidence Interval</th>
<th>IQR</th>
</tr>
</thead>
<tbody>
<tr>
<td>TV/Movies</td>
<td>.00</td>
<td>20.10</td>
<td>86.97</td>
<td>[19.00, 21.20]</td>
<td>.00</td>
</tr>
<tr>
<td>News</td>
<td>.00</td>
<td>37.97</td>
<td>224.07</td>
<td>[36.50, 39.44]</td>
<td>.00</td>
</tr>
<tr>
<td>Academic</td>
<td>170.33</td>
<td>337.74</td>
<td>485.10</td>
<td>[331.86, 343.63]</td>
<td>515.46</td>
</tr>
<tr>
<td>Statutes</td>
<td>799.31</td>
<td>964.17</td>
<td>723.93</td>
<td>[773.83, 1154.52]</td>
<td>668.55</td>
</tr>
</tbody>
</table>
Table 6. Descriptive Statistics for the Total Number of *Where*- connector PAs per Register.

<table>
<thead>
<tr>
<th>Registers</th>
<th>Medians</th>
<th>Means</th>
<th>SD</th>
<th>95% Confidence Interval</th>
<th>IQR</th>
</tr>
</thead>
<tbody>
<tr>
<td>TV/Movies</td>
<td>.00</td>
<td>2.11</td>
<td>25.75</td>
<td>[1.78, 2.43]</td>
<td>.00</td>
</tr>
<tr>
<td>News</td>
<td>.00</td>
<td>3.96</td>
<td>66.59</td>
<td>[3.52, 4.40]</td>
<td>.00</td>
</tr>
<tr>
<td>Academic</td>
<td>.00</td>
<td>24.52</td>
<td>104.34</td>
<td>[23.26, 25.79]</td>
<td>.00</td>
</tr>
<tr>
<td>Statutes</td>
<td>15.46</td>
<td>29.76</td>
<td>81.22</td>
<td>[8.40, 51.11]</td>
<td>20.31</td>
</tr>
<tr>
<td>Opinions</td>
<td>.00</td>
<td>125.35</td>
<td>437.28</td>
<td>[120.31, 130.39]</td>
<td>.00</td>
</tr>
<tr>
<td>Contracts</td>
<td>173.18</td>
<td>316.01</td>
<td>416.79</td>
<td>[300.28, 331.74]</td>
<td>277.98</td>
</tr>
</tbody>
</table>