InDetermination:
Measuring Uncertainty in Social Science Texts

Sarah B. Bouchat
bouchat@northwestern.edu

July 2018

Not For Circulation or Citation

Abstract: Previous work on the quantitative analysis of text emphasizes validity and error of classification schemes for text, usually via comparison to human coding approaches (e.g., Benoit, Laver, and Mikhaylov 2009 and 2012, and Grimmer and Stewart 2013). These works quantify uncertainty with respect to our inferences about information contained in text, but do not engage directly with the uncertainty characterized by the words and tone of the corpora themselves. Using techniques for identifying Type II and higher fuzzy quantifiers (e.g., "about a half," or "rather high") in natural language, this paper proposes a method for measuring uncertainty in text with broad applicability to the social sciences. The paper discusses several applications of the proposed approach to problems and techniques in social and political science.

1 Introduction

1.1 Problem & Motivation

Uncertainty estimation is not only core to generating social science inferences, it also fundamentally informs research design. While uncertainty estimates may be commonplace in quantitative analyses, characterizing the uncertainty present in qualitative research or qualitative sources of data, such as text, would provide substantial new opportunities for innovative research. Rather than only evaluating uncertainty at the level of individual variables, social science texts contain and express uncertainty about broad theoretical constructs and relationships between phenomena of interest. For example, in quantitative published works, authors may express uncertainty textually in order to explain anomalous outcomes or data limitations that threaten inference. A reliable way to quantify and evaluate this additional uncertainty would aid in the production of future research. For qualitative research or qualitative sources of data such as interviews, leveraging verbally or textually expressed uncertainty could provide an external mechanism for weighting and validating evidence.

Despite the promise of utilizing textual data for uncertainty analyses, uncertainty estimation in text analysis presents a particular problem because of the high degree of dimensionality in the data and because the structure of otherwise useful assumptions constrains analogues to sample-and-population definitions of uncertainty.

Defining the problem of uncertainty estimation for text data requires distinguishing between (at least) two levels: uncertainty within text and uncertainty about text. The first type of uncertainty—within-text—is the expressed uncertainty of the author(s) of a given document, via word choice, semantic construction, or omission. The second type of uncertainty—about text—can itself reflect the researcher's uncertainty about the measured or detected level of within-text uncertainty (i.e., how accurately automated methods detect the latent uncertainty of an author and disaggregate it from other explicit or implicit orientations), or can encompass uncertainty about the text at the level of the research construct (i.e., a level of confidence in the inferences drawn regarding the research question given the available data, and/or how well the selected data capture the concepts that the research agenda requires).

Previous work on the quantitative analysis of text emphasizes validity and error of classification schemes for text, usually via comparison to human coding approaches (e.g., Benoit, Laver, and Mikhaylov 2009; Grimmer and Stewart 2013). These works present possible ways to quantify about-text uncertainty, but do not directly measure or evaluate within-text uncertainty. While within-text uncertainty evaluation at first suggests a separate approach, emphasizing word selection/omission as well as arrangement, within-text and about-text uncertainty are not so easily disaggregated in theory. That is, current validation approaches rely on a notion that a crisp, "true" state of a document exists and that only measurement error or uncertainty interferes with inference. An author's latent disposition or intent, however, qualifies the concept that quantitative research belatedly attempts to measure. The extent to which this poses a problem for the validity of social science inference depends in part on the chosen mode of analysis—failing to detect underlying uncertainty or hedging may not interfere as much with recovering the topic(s) in a document as it would with estimating other quantities of interest, such as polarity or ideology.

By contrast, a separate stream of literature focused on international security and intelligence has dedicated significant effort to identifying terms that correlate with uncertainty ("words of estimative probability") in order to quantify risk assessments, but has not adequately addressed the ways in which authors express uncertainty intentionally, unintentionally, and via omission.

Using techniques for identifying fuzzy quantifiers (particularly Type II and higher) in natural language, this paper proposes approaches for measuring uncertainty in text with broad applicability to the social sciences. In particular, quantifying uncertainty in textual sources can aid methodological research evaluating open-ended survey questions or interview responses; facilitate the weighting of documents and the evaluation of veracity in archival research; leverage data in documents with unreliable or unknown narrators more effectively; and even lay the groundwork for the formalization of these insights from published literature in the form of fully specified Bayesian priors.

While fuzzy logic and fuzzy quantifiers have had a place in computational linguistic theory for many years, applications of the concept both witin the field and in cross-disciplinary endeavors are limited. Even the most straightforward approach of simply detecting and counting uses of fuzzy quantifiers within text runs headlong into a debate about whether, when, and which stopwords to optimally exclude prior to analysis (Manning, Raghavan, and Schütze 2008; Saif et al. 2014; Denny and Spirling 2018). Excluding words such as "all," "any," "few," "about," "some," or "much" might seem prudent to focus on words of relatively greater value in distinguishing across topics or documents, but these and other terms of fuzzy quantification serve as a fundamental basis for expressions of uncertainty about the subject matter within documents.

1.2 Overview for the Paper

The next section reviews literature relevant to uncertainty estimation from divergent perspectives in computational linguistics, computer science, mathematics, philosophy, and social science, to propose that fuzzy quantification offers significant promise in efforts to characterize uncertainty for social science text data. The following section then proposes two broad approaches to implementing uncertainty measurement using fuzzy quantifiers, and the conclusion provides a discussion of directions for additional research and refinement.

2 Literature: Uncertainty in Text

Uncertainty detection and characterization in natural language has drawn attention in computational linguistics, information fusion, and math and computer sciences, where providing theoretical underpinnings and methods have similar aims but disparate formulations, as well as social and security studies scholars, for whom the immediate practical significance is more acute. As such, there remains a gulf between the formal theorization of uncertainty within fuzzy logic frameworks dating to the 1970s and 80s, and successful applications, which are few in number and often end at the stage of detection/classification (e.g., Li, Gao, and Shavlik 2014; Jean et al. 2016; Conde-Clemente et al. 2017).

2.1 Subjectivity & Security Studies

Within NLP, subjectivity analysis—whether at the sentence or word level—has sought to identify "private states" (emotions or beliefs) from textual cues (Breck and Cardie 2014, 3). Information fusion and computational linguistics are home to many ongoing efforts to generate a definitive typology of uncertainty and ambiguity that would undergird detection efforts (Rubin, Liddy, and Kando 2006). Akkaya, Wiebe, and Mihalcea (2009) offer insights into the problem of subjective word sense disambiguation, which plagues efforts to accomplish accurate automated subjectivity analysis. Identifying words that reflect sentiments or opinions, rather than words that "cue" subjectivity but are intended objectively, is challenging precisely because subjectivity is a "private state"—much like uncertainty, it is a latent condition that can only be partially observed or measured via language (2). While the authors introduce computational approaches to correctly identifying subjectivity, the conceptual overlap between subjectivity and uncertainty is limited and the attempt to classify language into binary usage categories (subjective/objective) is very limiting in the context of attempting to characterize either the type or extent of uncertainty reflected in language. For example, the authors discuss two differing uses of the word "catch," where "What's the **catch**?" reflects a *subjective* use of the term meant to indicate a "drawback," whereas "He sold his **catch** at the market" is instead *objective* because it represents a quantity. For the purposes of uncertainty estimation, however, the indefiniteness of the quantity "a catch" is of greater importance than its objective sense (Akkaya, Wiebe, and Mihalcea 2009, 2). The subjectivity lexicon emerging from this research agenda, therefore, offers a very incomplete inventory of the types of words that lend themselves to uncertainty quantification.

Dragos (2013, 4) provides a useful overview distinguishing among types of "intrinsic" uncertainty:

(1) **Ambiguity** arises naturally in language because it can never truly capture all observations and internal experiences, but in particular emerges from polysemy or incompleteness.

(2) **Vagueness & Precision** are contrasting concepts that may reflect intentional or unintentional shifts toward or away from exactness. For example, "they have several cats" is a vaguer construction of "they have 4 cats," where the choice to use "several" in the place of a discrete quantifier may reflect a lack of knowledge about the precise number or may be an intentional choice to shift importance (e.g., if the emphasis is that they have cats rather than dogs).

Rather than focusing on strict linguistic ambiguity—the kind of uncertainty that arises from gaps between the signifier and the signified, as with polysemy and homonymy—or reality-limited referential ambiguities—such as when color descriptions are inherently biased, unclear, or culturally dependent (e.g., green-blue versus blue-green versus turquoise, or proximate versus obviative referents)—treating what Dragos calls the "vagueness" form of uncertainty as a *knowledge domain* contained within language implies several clear modes of approach, beginning with lexical definition and detection, and proceeding through quantification (Auger and Roy 2008, 1862–1863, 1865–1866). Druzdzel (1989) catalogued 178 verb phrases and modifiers indicating uncertainty (e.g., "it could be," "it seems," "impossible"), a pursuit informed by the interest of the security community in quantifying uncertain levels in natural language, stemming from the famous 1973 Rand report by Sherman Kent.

Probability (percent)	Verbal Equivalent
100	It is certain that ...
85-99	It is almost certain that ...
60-84	It is probable that ...
40-59	The chances are about even that ...
15-39	It is probable that ... not ...
1-14	It is almost certain that ... not ...
0	It is impossible that ...

Figure 1: A chart from the 1973 Kent report, as depicted in Auger and Roy (2008, 1866)

The security studies lineage of quantifying qualitatively expressed uncertainty in text arises from the immmediate, significant implications for precisely characterizing experts' assessments of the likelihood of events has immediate, significant implications (Miller et al. 2013). Recognizing the need to specify and examine uncertainty across multiple dimensions and applications, Thomson et al. (2005) developed a framework that disaggregates nine distinct types of uncertainty: accuracy, precision, completeness, consistency, lineage, currency, credibility, subjectivity, and interrelatedness (Auger and Roy 2008, 1861). As with Dragos' distinction, not all of these types are necessary or feasible for quantifying uncertainty in social science texts and applications. For example, assessing the accuracy or completeness of analyses in social science articles or in open-ended survey responses may be beyond the scope of automation, and evaluating the credibility of a source may require information not contained within the source text itself.

Beyond theoretical distinctions among types of uncertainty, Druzdzel's survey provides evidence of a human preference to express uncertainty qualitatively rather than numerically, perhaps as an instinctual desire to avoid overly certain statements. Practically speaking, this evidence underscores the significance of accurate methods for uncertainty detection and measurement—however defined—in natural language processing.

2.2 Fuzzy Logic & Fuzzy Linguistics

Mathematics and formal logic, in turn, do not lack for ways to conceptualize and reason with vagueness and uncertainty (e.g., Wygralak 1998; Roos 1990), but correlating these theoretical constructs with expressions in natural language presents significant challenges (Barwise and Cooper 1981; Zadeh 2005). In particular, common statements such as "most people" or more abstract phrases such as "there are only a finite number of stars" cannot conform to standard first-order logical conditions such as $\forall x(...x...)$ (Barwise and Cooper 1981, 160). This shortcoming indicates precisely why fuzzy logic and possibility theory have offered more promising avenues of theoretical exploration.

Fuzzy logic, while maligned by proponents of traditional logic schema across disciplines, directly leverages and incorporates the ambiguity inherent in language into formal reasoning. Unlike in classical logic, where propositions are evaluated as either true or false, fuzzy logic propositions have a "truthfulness" value that is a real number on the interval $[0, 1]$ (Atanassov and Gargov 1998, 40). While the use of fuzzy logic is often at odds with formalized probabilistic representations (e.g., Lindley 1987), its aims and application bear clarification:

> Fuzzy logic is not fuzzy. Basically, fuzzy logic is a precise logic of imprecision and approximate reasoning. More specifically, fuzzy logic may be viewed as an attempt at formalization/mechanization of two remarkable human capabilities. First, the capability to converse, reason and make rational decisions in an environment of imprecision, uncertainty, incompleteness of information, conflicting information, partiality of truth and partiality of possibility—in short, in an environment of imperfect information. And second, the capability to perform a wide variety of physical and mental tasks without any measurements and any computations. (Zadeh 2008, 2753)

Natural language epitomizes the "imperfect information" that fuzzy logic and fuzzy sets attempt to contend with. Fuzzy linguistic representations have presented significant challenges to implementation, however, with fragmented efforts either emphasizing a single test application that does not generalize, or focusing on questions beyond uncertainty per se (e.g., Cabrerizo et al. (2017) utilizing fuzzy linguistics to evaluate "consensus" in decision-making). In particular, though, the evaluation of fuzzy quantifiers has been one critical point of entry for bringing fuzzy logic theory to linguistic applications, originating in the the study of generalized quantifiers in language (Barwise and Cooper 1981). Fuzzy quantifiers, in general, are imprecise descriptors often used to construct propositions for fuzzy logic. Terms such as "most," "many," or "few" provide a general sense of the quantity in question, but only within certain bounds and sometimes with reference to a vaguely defined population (analogous to the cardinality of a fuzzy set[1]). The degree of imprecision and the reference category, in turn, characterize the type of fuzzy quantifier, or the degree of fuzziness (versus crispness). Specifically, Zadeh (1983, 150) defines a fuzzy quantifier as "a fuzzy number which provides a fuzzy characterization of the absolute or relative cardinality of one or more fuzzy or nonfuzzy sets."[2]

Distinguishing among types of quantifiers, furthermore, is both necessary for evaluating not just *whether* statements are uncertain but the *degree* to which they are, and is also an ongoing effort. Zadeh (1983) posits at least three types of fuzzy quantifiers, but broadly, type or degree varies according to the fuzziness of the predicate and the quantification in the phrase (Liu and Kerre 1998, 2):

Type I: Cardinal number extensions of two-value logic (e.g., "All buildings are

[1]The cardinality of a fuzzy set can be a real or a fuzzy number, but broadly speaking, represents the number of elements that are members of a set S. Conceptually, the cardinality of a fuzzy set is complicated, as it can itself be fuzzy and therefore require characterization only in relation to other fuzzy sets (Dhar 2013).

[2]That is, not all fuzzy numbers are fuzzy quantifiers, and whether a fuzzy number qualifies as a fuzzy quantifier is sometimes unclear.

solid," "Greater than half of the surface of the southern hemisphere is water")

Type II: Quantifiers of fuzzy sets (e.g., "Some children are tall")

Type III: Fuzzy quantifiers of crisp sets/ratios of fuzzy quantifiers (e.g., "Almost all mammals indigenous to North America are eutherian,")[3]

Type IV: Fuzzy quantifiers of possibility distributions (e.g., "Few politicians remain popular long")

Dìaz-Hermida, Bugarìn, and Barro (2003) goes further in examining the bounds of these quantifiers (semi-fuzzy quantifiers, ultra-fuzzy quantifiers, and the process of quantifier fuzzification), while others have posited a multitude of possible ways to detect and measure these types of quantifiers (e.g., Szmidt and Kacprzyk 2001; Zhai and Mendel 2011; Chen, Song, and Heo 2017; Ramos-Soto and Pereira-Farina 2017). For the purposes of this paper, "Type II" fuzzy quantifiers will serve as the primary point of entry for developing a method of uncertainty measurement and evaluation. Many critical concepts and research agendas in the social sciences can be thought of in the context of fuzzy sets: scholars often draw theoretical battle lines around definitions that constitute set members (e.g., what is a democracy versus an autocracy, what types or magnitude of violence qualify as war, etc.). Statements on the basis of empirical evaluations of these theoretically defined concepts, then, directly map onto fuzzy quantification. Likewise, fuzzy or "soft" quantification is more flexible to changing population sizes than traditional metrics, and as such may facilitate easier model comparison (Farnadi et al. 2016, 61).

2.3 Possibility Theory

More broadly, how fuzzy quantifiers are expressed and evaluated hinges on the underlying possibility distribution that defines the fuzzy number. That is, informally, overlap

[3]Likelihood ratios also qualify as a type of fuzzy quantifier.

exists in our conception of what constitutes "most" and "more than half," but how much these categories overlap depends on the ease with which we believe a number that is $> 50\%$ belongs to the set we conceptually define as "most." That is, while both 52% and 80% are contained in the set "more than half," colloquially we may believe that the descriptor "most" is more easily attributable to 80% than 52%.

Possibility theory provides a framework through which to adjudicate precisely these distinctions, occupying conceptual space between probability and fuzzy set theories. As Khoury, Karray, and Kamel (2008) articulates:

> A probability of 0 means a certainty that the event will never occur, while a probability of 1 is a certainty that the event will occur. Fuzzy set theory, on the other hand, deals with the membership of an event in a set. A membership value of 0 means that the event does not belong at all in the set, while a membership value of 1 means that the event epitomizes the set. Finally, the possibility theory expresses the ease with which an event can occur, or belong to a set. A possibility of 0 means that an event cannot occur, while a possibility of 1 means that the event is completely allowed to occur, and values between 0 and 1 represent events that are restricted but not impossible. (1531)

While applications of possibility theory to text analysis are few and far between, its direct relationship to the ambiguity of language make it conceptually compelling. For example, Khoury, Karray, and Kamel (2008) track the domain representation of particular words in a corpus via a possibilistic domain classifier. That is, with language stripped into subject-verb-object triplets, they instantiate the domain information of language in the form of possibility distributions. As an example application, they use triplets gathered from descriptions of learning objects in the SchoolNet corpus to conduct classification, calculating a possibility ($\pi(d_t)$) that indicates the ease with which a given test learning objective can belong to the domain d_t (1540). Because their test is performed with a winner-takes-all approach, they only evaluate the statistic with respect to the most-possible domain. Beyond utilizing the subject-verb-object construction, therefore, the practical utility of this approach relative to topic modeling or other alternatives is less clear. Conceptually, however, using possibility theory and fuzzy sets to both directly and indirectly characterize uncertainty is promising. Indirectly, possibility theory can generate flexible bounds or constraints on the subject matter of interest or important words or topics and can thereby represent uncertainty. Khoury et al. offer the example phrase "this room is cold," where "cold" represents a fuzzy set that bounds our expectations about the likely temperature of the room. If the temperature were 15°C, we could state that the constraint is satisfied to the degree 0.8. Framed in possibility theory terms, this would indicate that there is a *possibility* of 0.8 that the room temperature is 15°C given that it was described as "cold;" that is, 0.8 suggests the "ease" of labeling a room at that temperature as "cold" (Khoury, Karray, and Kamel 2008, 1532).

More directly, possibility theory could serve to construct a lexicon of uncertainty terms or to refine existing lexica related to subjectivity, such that terms are classified not as representing uncertainty per se, but rather the degree to which they indicate uncertainty (e.g., modifiers such as "never" are not compatible with the set of terms that characterize uncertainty whereas terms like "about" may be compatible with that set at 0.5, where polysemy accounts for other uses and membership of the term in other sets). Both practically and theoretically, the continuous character fuzzy quantifiers and possibility theory is more attractive than in the discrete labeling common for subjectivity analysis.

In this paper, I contend that fuzzy logic and fuzzy linguistic principles are not only themselves potential measures of uncertainty in text, but also can be used as measurement *tools* to assess uncertainty within and across texts. The following section presents empirical approaches to applying fuzzy logic and fuzzy quantifiers in the context of providing uncertainty estimates for textual data.

3 Methods for Uncertainty Estimation with Fuzzy Quantifiers

This section discusses two broad categories of approaches to utilizing fuzzy quantifiers in uncertainty estimation for social science. The first category focuses on dictionary-based approaches that are familiar to subjectivity analysis, while the second set proposes a more expansive adoption of fuzzy logic principles to assess uncertainty in textual data. Applications to other types of social science texts are discussed following the initial analysis.

Both broad sets of approaches to using fuzzy quantifiers ultimately require detecting fuzzy quantifiers in text, and utilizing fuzzy quantifiers as a measurement tool imposes extra weight and significance on the accuracy of measuring the quantifiers themselves. While more work is likely necessary to define and refine the set of fuzzy quantifiers most useful and relevant to estimating uncertainty of various types, this section proceeds from a practical and inclusive perspective that will be amenable to later changes in specification. No global dictionary of fuzzy quantifiers in English currently exists, so I take several steps to generate an inclusive set set of fuzzy quantifiers:

(1) Generate a list of terms used in the original Zadeh (1983) article examining fuzzy quantifiers

(2) Add items from Liu and Kerre (1998), which reviews and explains Zadeh (1983) with additional examples

(3) Add empirical examples from corpora used for testing (BioScope corpus and Farnadi et al. (2016)) and empirical applications (Conde-Clemente et al. (2017)):

    (a) Test existing list from items (1) and (2) to recover items from corpora

    (b) Hand-review Bioscope corpus sample for additional examples

The comprehensive list of fuzzy quantifiers used in this section is provided according to source in the Appendix.

3.1 Dictionary-Based Approaches

3.1.1 Fuzzy Quantifiers & Subjectivity Terms

As indicated previously, the literature that attempts to empirically evaluate uncertainty almost exclusively focuses on dictionary-based methods. These efforts largely center on subjectivity analyses that identify and classify non-objective words and phrases; rarely are fuzzy quantifiers included, let alone central to the analysis. At a more theoretical level, the subjectivity analysis approaches implemented to date do not emphasize quantifying subjectivity at the document or corpus level—this quantification would require a weighting or aggregation scheme from the word- and sentence-level detection procedure. This limits the utility of current approaches for characterizing how uncertain a given article or author might be, above and beyond simply answering the question of whether they use sentences containing uncertainty terms.

A simple first step toward bridging subjectivity analysis and fuzzy quantifier approaches is to assess whether and to what extent fuzzy quantifier terms are distinct from and improve upon subjectivity terms. For convenience, this step reproduces, in part, the 2010 shared task from CoNLL, which required the detection of hedge words (Farkas et al. 2010), and data from Vincze (2014), which implements uncertainty detection across several corpora. Vincze uses the Szeged Uncertainty corpus ("Szeged Uncertainty Corpus" 2010), which is composed of the subcorpora BioScope 2.0, FactBank 2.0, and WikiWeasel 2.0 (details in Table 3.1.1). These datasets are useful primarily because they are pre-labeled for subjectivity, but attempting to "predict" that pre-labelled value with only fuzzy quantifiers is not feasible on theoretical as well as empirical grounds. On a theoretical level, as indicated by the aforementioned literature, the type of uncertainty indicated by fuzzy quantification is often conceptually distinct from what is coded as "subjective."

One dataset is primarily occupied with automatically detecting the *type* of subjectivity at the word level, therefore requiring several aggregation assumptions in order to recover a document-level metric of "uncertainty," as well as assumptions about which of these types is most compatible with the quantifiable "uncertainty" that fuzzy quantifiers embody. On a practical level, furthermore, the rates of fuzzy quantifier usage are relatively low across all of the pre-labeled corpora, making detection and classification particularly challenging.

Corpus	Description	Labels	Size
Bioscope	Radiology reports, full biology papers, and Genia corpus abstracts	Negation, hedge words, and hedge word scope	20,924 sentences
FactBank	Newswire and broadcast data	Events coded into 4 "factuality" types by source and point of view	3,123 sentences
WikiWeasel	Paragraphs from the full Wikipedia dump	"Weasel" cues	20,745 sentences

Figure 2 illustrates the overall rate of fuzzy quantifier usage by corpus, scaled by the total number of words. While variation exists, the tight clustering around an average rate of 0 underscores the challenge of conducting a validation task comparing fuzzy quantifiers with subjectivity words. Likewise, at the document level, the average number of fuzzy quantifiers per sentence is relatively low, with similarly low variance, as shown in Figure 3. An examination of the association between fuzzy quantifiers and subjectivity words, as shown in Figure 4, is inhibited by the overall low rate of fuzzy quantifier usage, but additionally suggests that at most, fuzzy quantifiers only capture a single facet of subjectivity, rather than being a completely congruent concept. Likewise, while the narrative structure and expression of uncertainty likely differs by corpus, fuzzy quantifiers and at least some kinds of subjectivity terms more likely function as substitutes rather than complements when articulating uncertain views. As such, a correlation measure based on co-occurrence at the sentence or document level is unlikely to capture their theoretical relationship.

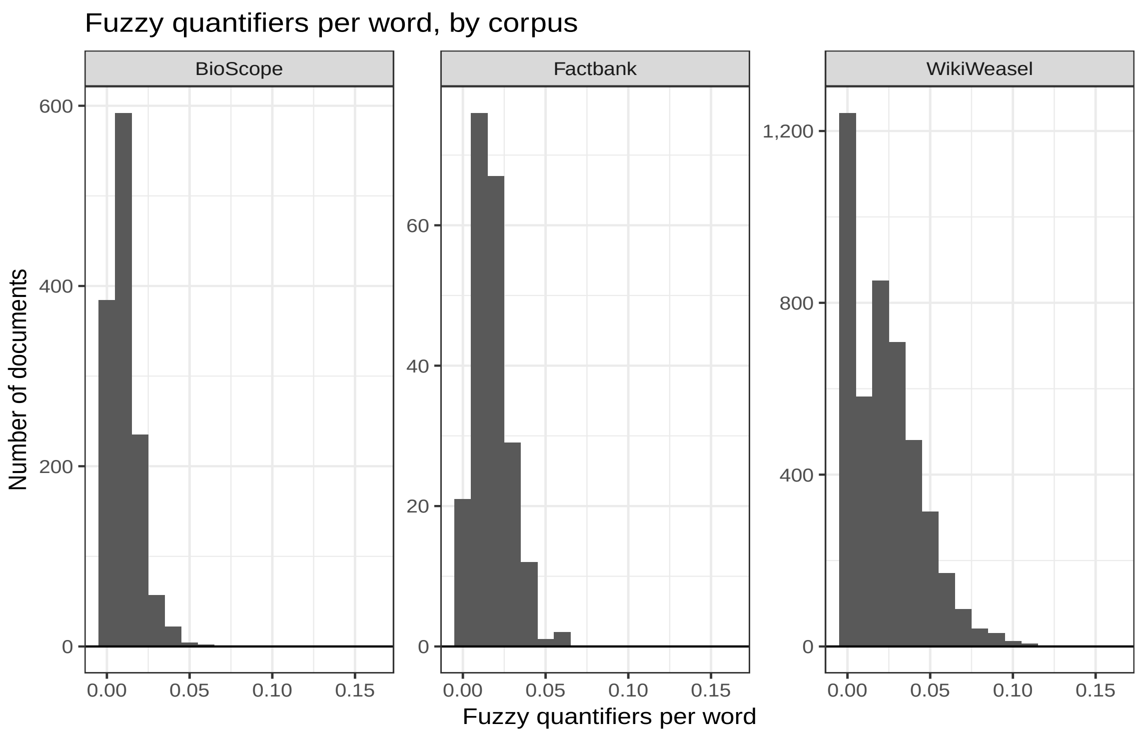

Figure 2: Per-Word Fuzzy Quantifier Rates by Corpus

3.1.2 Fuzzy Quantifiers & Word Embeddings

An alternative way of conceiving of the significance of fuzzy quantifiers as a dictionary-based metric of uncertainty is via their association with other "key words" that characterize a particular document. One advantage of this approach is that it scales and automates relatively easily without requiring human intervention to read or summarize the articles in question.

By way of illustration, this section selects two articles to evaluate fuzzy quantifier relationships with keywords. The first is Chen, Pan, and Xu (2015), which evaluates an

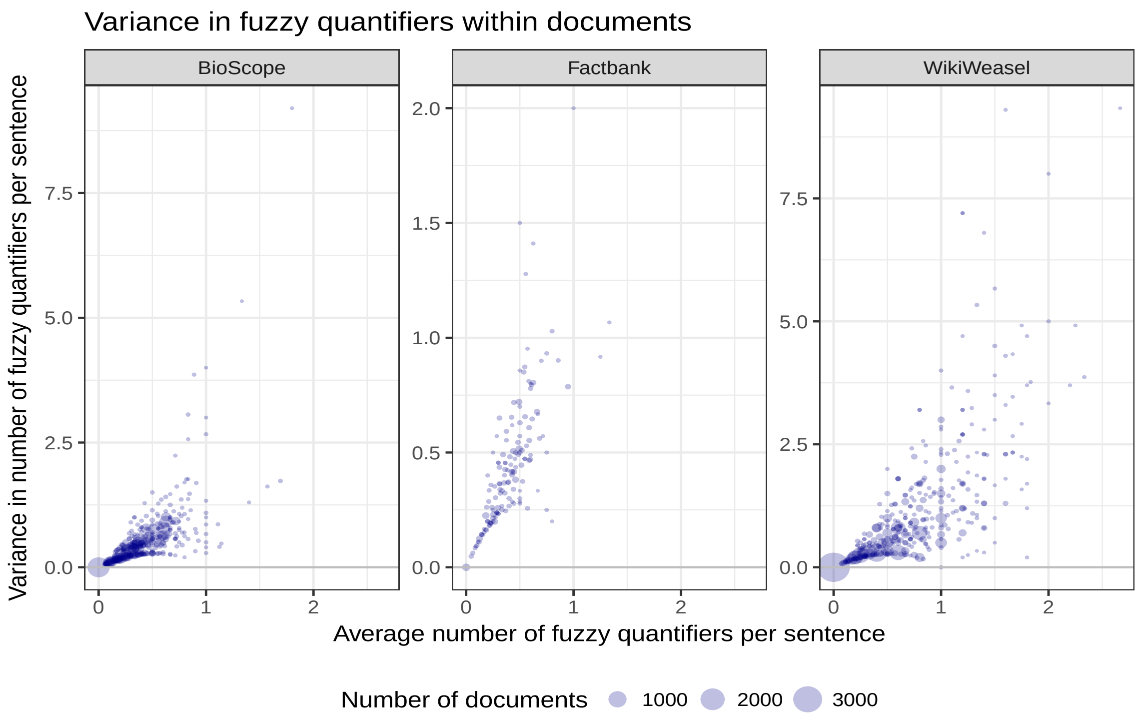

Figure 3: Fuzzy Quantifier Variation (Avg. Num. Per Sentence) by Document

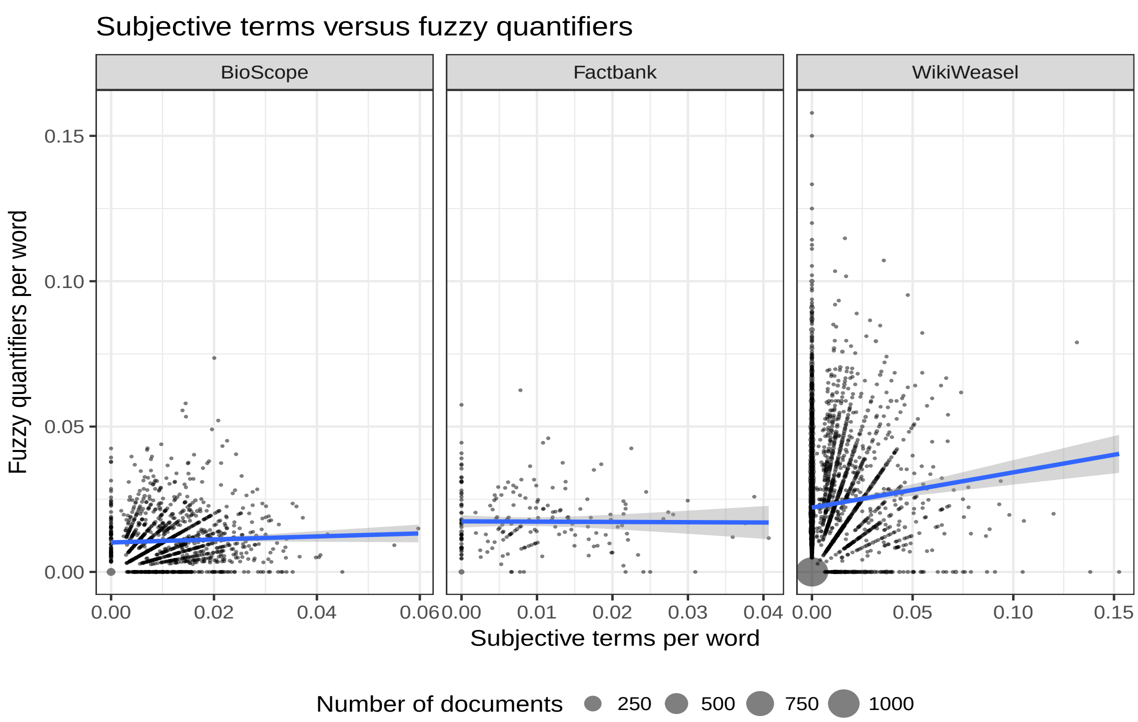

Figure 4: Correlation between Fuzzy Quantifiers and Subjectivity Terms

online field experiment across Chinese counties. Web of Science provides a set of keywords courtesy of KeyWords Plus, which is a human coded product of Thomson Reuters. A researcher seeking to automate this procedure across a body of literature, either culled directly from Web of Science or searched by keyword, could easily extract terms for each article selected by the same mechanism. For this article, the reported keywords are "institutions," "accountability," "politicians," "legislators," "protest," "policy," and "power." The second is a recent paper from *arXiv*, where keywords are automatically tagged within LaTeX source text. In constructing a "deconfounder" to disentangle multiple possible causes within observational data, Wang and Blei (2018) tackle the "strong ignorability" assumption of having accounted for all possible confounders, testing their method with synthetic health outcomes data. The stated "keywords" for the article, which are author-selected rather than editor-selected as with KeyWords Plus, are "causal inference," "strong ignorability," and "probabilistic models."

Notably, the number of possible keywords may vary by article upon implementation, and they are not explicitly or rigorously ranked according to their importance in characterizing the articles. The source and nature of these keywords is a limiting factor for generating meaningful, comparable uncertainty metrics across documents in a given corpus. For unedited papers or author-selected keywords such as those from *arXiv*, keywords may tend to be overly specific and proliferate, making direct comparisons across papers challenging (i.e., directly stating that Paper A appears more uncertain/offers more qualified statements regarding keyword x than Paper B is impossible if they do not share a keyword). Likewise, the editor-selected keywords offered by services such as KeyWords Plus or by journals may not directly correspond to the language of the articles, but rather to important disciplinary categories. With the paper by Chen, Pan, and Xu, for example, neither "legislators" nor "politicians" appears in the paper itself. For the purposes of the illustration below, these are substituted with "officials," but this type of substitution would interfere with the automation of these associations at scale.

Using GloVe embeddings, cosine distance measures the co-occurrence of fuzzy quantifier terms with each of the keywords for both papers (Pennington, Socher, and Manning 2014). Figures 6 and 7 depict the top fuzzy quantifier associations with each keyword by article, sorted by cosine distance, where -1 indicates a strong negative association and 1 represents a strong positive association; observations close to zero suggest orthogonal relationships. As the figures show, overall, fuzzy quantifier terms lack strong associations with most keywords. This highlights a few necessary conceptual steps for refinement. First, depending on the application, accurate measurement might require disaggregating fuzzy quantifiers from other keywords that are topically oriented. For example, the main point of the Wang and Blei paper is to evaluate the *multiple* causal pathways that exist for any given observational dataset, suggesting that the relatively stronger association between "multiple" and "causal inference" does not suggest greater certainty or a certainty regarding degree, but is rather subject-matter oriented. This point is underscored in the comparison to a baseline using pre-trained embeddings below.

Likewise, many fuzzy quantifier terms represent common "levels" of uncertainty or qualitative types of uncertainty, as indicated in part by the typology that Zadeh offers. For example, "almost" and "around" may serve similar functions and have similar set memberships, meaning that their associations could plausibly be collapsed to a single measure. "Many" and "few," in contrast, are opposing and possibly disjoint sets. When these terms are exclusively applied to or associated with particular keywords of interest, inferences about the expressed levels of uncertainty are clear; when both are applied to a single term, however, describing or quantifying the underlying level of uncertainty requires clearer statements about their underlying possibility distributions. While additional work is necessary to refine and validate the preliminary work that does exist, Sherman Kent's original work to characterize the possibility distributions (probability distributions) underlying individuals' assessments of the meaning of certain fuzzy quantifier terms serves as one possible way forward. Kent's assessment, undergirding the probabilities listed in Figure 1, relied on surveys of 23 Navy SEALS interpreting intelligence reports, allowing him to generate distributions of responses, which were later binned into his preferred bounds for interpretation, as shown in Figure 5 (Pherson and Pherson 2012, 187). Similar analysis has been done informally with members of r/SampleSize (zonination 2017) and could be conducted more rigorously at a larger scale.

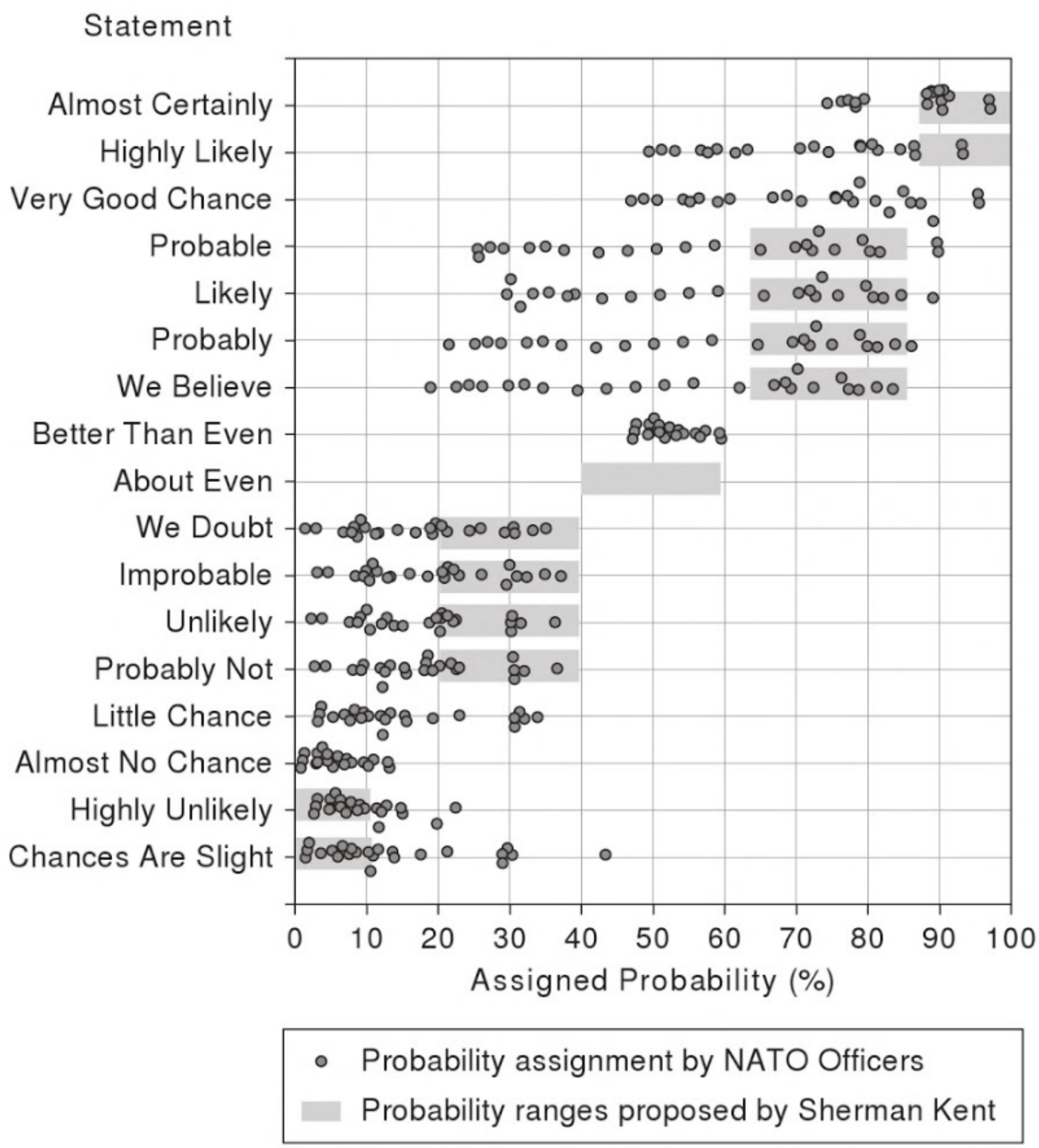

Figure 5: Reproduction of Sherman Kent's Perception of Uncertainty Measures from (Pherson and Pherson 2012)

In addition to this question of degree or magnitude, an issue of aggregation again arises along two dimensions. First, for multiple quantifiers with stronger relationships to an individual keyword, how should their differing levels of certainty be quantified, combined, and expressed? For example, qualitatively, the word "protest" in Chen, Pan, and Xu (2015) has a relatively stronger positive association via co-occurrence with "between" and a relatively stronger negative association with "none" and "less than." One interpretation is that the authors express some degree of certainty that "protests" in their empirical framework are a non-zero occurrence, and their rate is uncertain but bounded in expectation. While narratively useful, perhaps, even if accurate, this conclusion provides a very limited basis for making clear claims about the *amount* of uncertainty the authors express.

Second, at the document level, it is not clear whether or in what way it might be appropriate to aggregate differing types or levels of uncertainty related to multiple key terms of interest in an effort to characterize the "overall" uncertainty of the article in question, whether that characterization applied to the subject matter, research question, or researcher themselves. That is, if quantifiers of relatively greater certainty applied to term A and quantifiers of relatively lesser certainty applied to term B, should the aggregate level of uncertainty reflect an average, or a weighted combination, or something more reflective of the variation present?

In both figures, furthermore, the association between fuzzy quantifiers and keywords is suspect because embeddings are derived from a single paper with only a few thousand words, relative to typical pre-trained embeddings using more than a million words. One approach to rectifying this issue, as discussed later, would be to build a corpus specific to the social science applications of interest, with the aim of reflecting the type of language most likely to be used in situations where estimating text-based uncertainty is desired. Short of that solution, however, pre-trained embeddings offer a direct

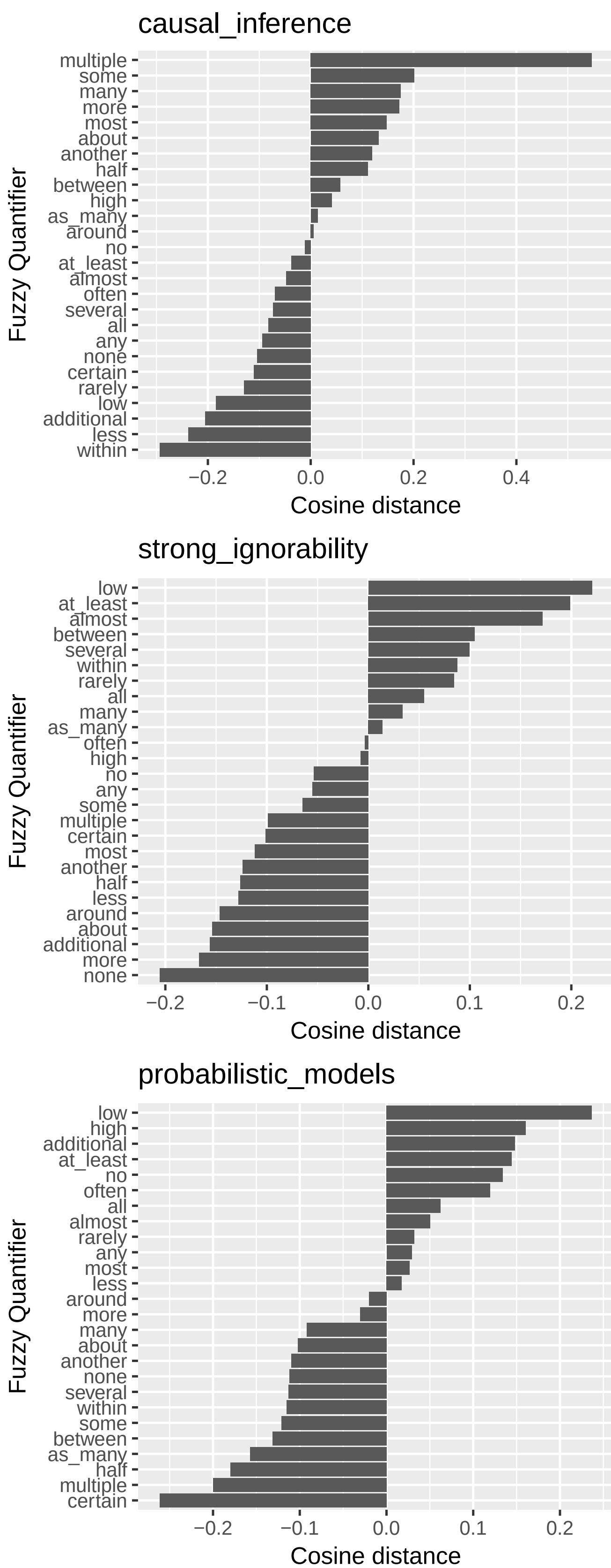

Figure 6: GloVe-generated Cosine Distances between Fuzzy Quantifiers and Article Keywords

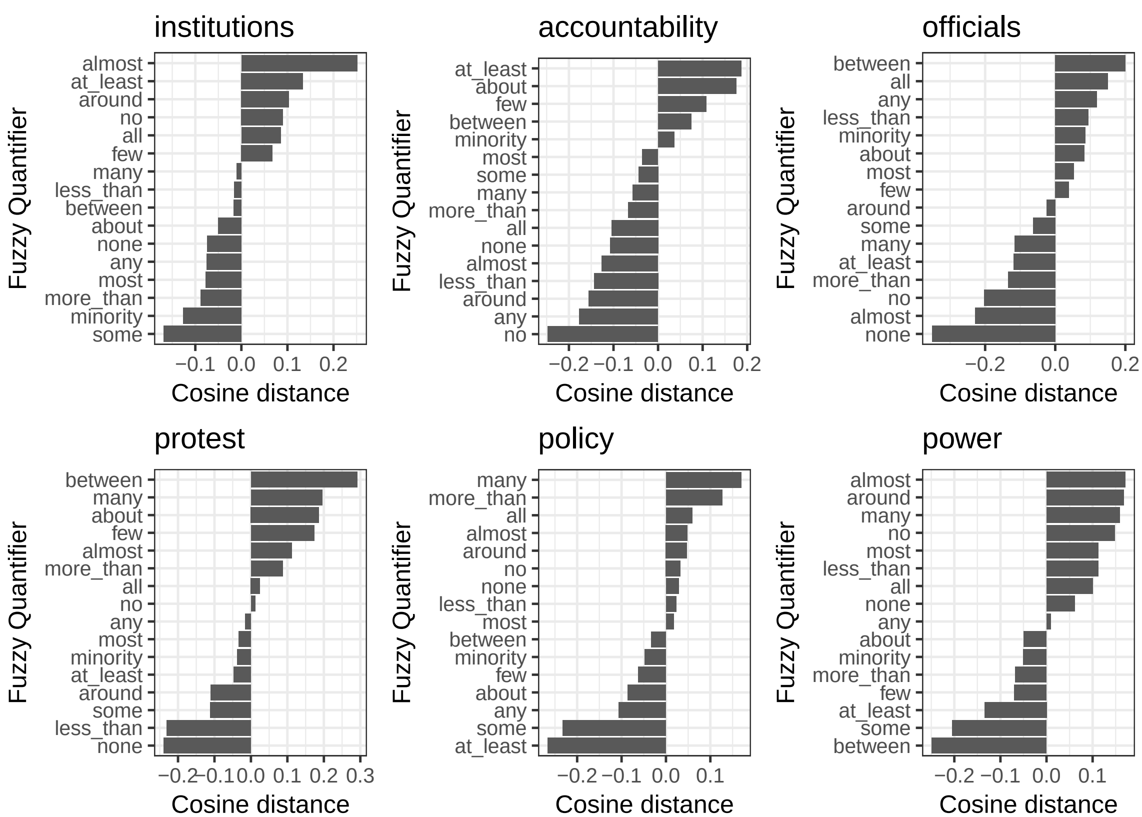

Figure 7: GloVe-generated Cosine Distances between Fuzzy Quantifiers and Article Keywords

way of evaluating whether the types of relationships we would like to identify between fuzzy quantifiers and keywords of interest are more general, inherent qualifications (e.g., about unknowable subjects or quantities) or are reflections of a given paper or author's uncertainty about a subject matter.

To demonstrate, I draw on GloVe pre-trained embeddings using text from the 2014 Wikipedia and Gigaword 5 set, with approximately 6 billion tokens (Stanford NLP Group 2018). Pre-trained embeddings ideally reflect the form and style of semantic relationships pertaining to the research question, and in this case Wikipedia entries plausibly provide a useful baseline assessment for social scientific writing, as the stated aim to provide reliable information, as well as the norms surrounding citation of published works, mirror those of academic writing. Unlike an individual social science article, however, Wikipedians likely use language that is more legible to a lay audience, and also likely reflect an average level of uncertainty about truly uncertain topics because of the contribution of multiple writers and editors.[4] Evaluating individual articles' use of fuzzy quantifiers with respect to keywords of interest against this baseline, then, could provide a more accurate assessment of the expressed level of uncertainty, albeit bundling measurement uncertainty inherent in the embeddings estimated at the individual article level.

Figure 8 provides an example of this type of comparison using the Wang and Blei article, but also its pitfalls in an exclusively dictionary-based approach. As might be expected, the keywords in that paper are not common parlance and not featured in a wide variety of Wikipedia subjects and articles. Of the keywords discussed previously, only "probabilistic models" occurs in both the baseline set of pre-trained embeddings and the article of interest. This figure illustrates the cosine distance, by fuzzy quantifier, of each qualifying term against that single keyword, as well as the change in cosine

[4] A brief assessment of the usage of fuzzy quantifiers and potential contention of topics is featured in the Appendix.

distance from the pre-trained baseline to the article.

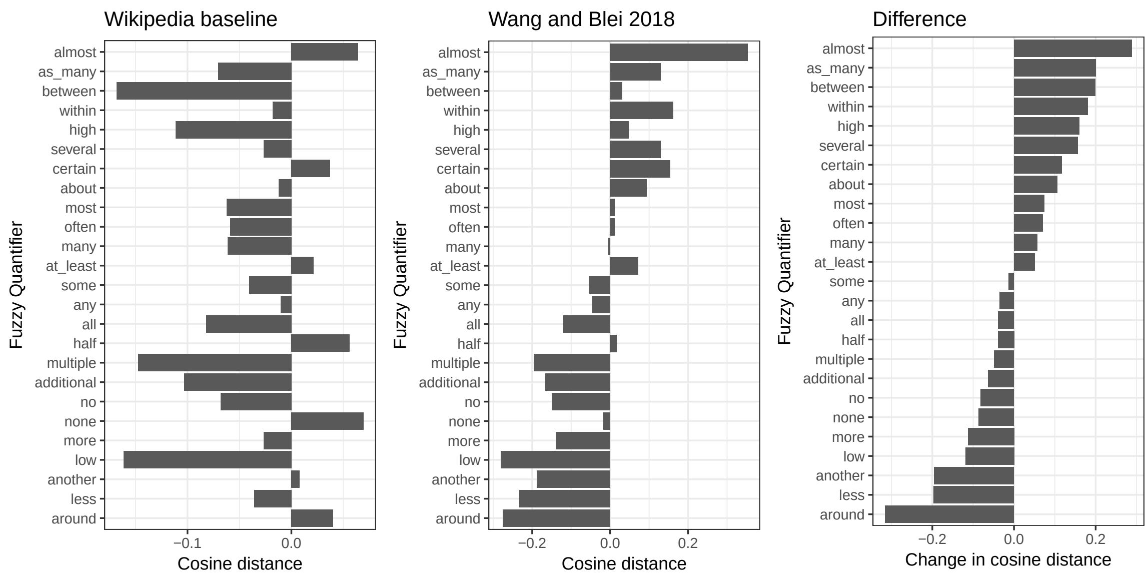

Figure 8: GloVe-generated Cosine Distances between Fuzzy Quantifiers and Article Keywords: Wikipedia Baseline and Difference for "Probabilistic Models" Wang and Blei (2018)

This difference metric, while feasible to construct when sufficient keyword prevalence exists, suffers from the same issue previously discussed regarding uncertainty aggregation across different fuzzy quantifiers. The change in cosine distance between baseline and individual article, on one hand, gives a clear assessment of how much more or less certain a particular article is relative to the Wikipedia "baseline," where positive values in the change reflect *more uncertainty* (more association with fuzzy quantifier terms reflects greater qualification and therefore greater uncertainty in this dictionary-based framework). This brief illustration demonstrates that substantial variation between fuzzy quantifier usage in an individual paper relative to a baseline metric is possible, and that a distance metric between these associations could capture variation in this uncertainty at the individual fuzzy quantifier level with respect to a single keyword. In this example, though, the change in cosine distance for "almost" is positive, while the change in cosine distance for "around" is an almost identically negative value. Strict aggregation of each of these fuzzy quantifier terms to generate a single document-level uncertainty measure would be problematic on theoretical and semantic grounds, but in addition, this approach only using individual keywords is subject to significant variation based on the selection of keywords and their generality for a given paper.

3.1.3 Discussion

Taken together, these fuzzy quantifier refinements on dictionary-based approaches offer tractable points of entry for evaluating uncertainty in larger scale corpora while remaining within methodological constructs familiar to text analysis. As indicated previously, however, additional theoretical work is necessary to assess the extent to which fuzzy quantifiers simply add information to previously existing subjectivity approaches, or whether they represent a distinct category or class of uncertainty assessment. Even having decided this pivotal theoretical issue, however, concerns about how best to weight and aggregate counts of fuzzy quantifiers in order to characterize total uncertainty at the document level (or by author, or by corpus), remain.

3.2 Fuzzy Proposition Evaluation

Rather than relying exclusively on dictonary-based approaches and effectively "counting" occurrences of fuzzy quantification, which presents not only theoretical but also practical challenges, a different approach would more firmly embed uncertainty estimation of text in a fuzzy logic framework prior to leveraging fuzzy quantifiers themselves. Zadeh (1983) offers an example of "test-score semantics," which itself views "everything that relates to natural languages [as] a matter of degree" (152). In a test-score setting, Zadeh proposes generating an *explanatory database frame* (EDF) that contains relational information for pieces of text that can be evaluated via test-scores against propositional statements. These test-scores, then, provide document-level quantities that can measure uncertainty both directly and indirectly. This section applies his hypothetical example to a social science problem for illustrative purposes.

By way of a proposition to evaluate with text, Zadeh offers the hypothetical that "Over the past few years Nick earned far more than most of his close friends," where the relations this requires cataloguing in an EDF include:

INCOME: listing the name of friends, amount they earned, and year

FRIEND: listing the name of friends and a value μ that represents the degree to which the person is a friend of Nick

FEW: μ representing the degree to which some *number* qualifies as "few"

MOST: μ representing the degree to which some *proportion* qualifies as "most"

FAR MORE: two income values and μ indicating the extent to which the first income value qualifies as "far more" than the second income value (Zadeh 1983, 153).

The information contained in each of these relationships and the elastic constraints imposed by the value(s) of μ facilitate evaluating the "truthfulness" or possibility of the propositional statement. This process very readily applies to social science domains, where this proposition assessment can be thought of metaphorically as hypothesis testing for relationships between variables of theoretical importance. Even so, Zadeh's article is primarily directed toward demonstrating that the concept of fuzzy sets applies to natural language, and that test score operations and cardinality are likewise *possible*. The article does not, therefore, emphasize ways of measuring or characterizing μ, the parameter that largely determines what inferences one might draw about uncertainty statements within text.

Farnadi et al. (2016) attempts to overcome this shortfall in an application to statistical relational learning (SRL). The paper proposes one family of quantifier mappings that correlate soft (fuzzy) quantifiers with quantitative metrics. Specifically, the paper defines a mapping $\tilde{Q}$ for soft quantifier Q such that:

$$\tilde{Q}_{[\alpha,\beta]}(x) = \begin{cases} 0 & \text{if } x < \alpha; \\ \frac{(x-\alpha)}{\beta-\alpha} & \text{if } \alpha \leqslant x < \beta; \\ 1 & \text{if } x \leqslant \beta. \end{cases}$$

This mapping means that, for example, one could define $\tilde{Q}_{Few} = \tilde{Q}_{[0.1,0.4]}$ (Farnadi et al. 2016, 64). This formulation is guided primarily by functional convenience rather than theoretical argument or empirical evidence, and while their later proof-of-concept analysis tunes the mapping and keeps it largely the same, the paper does not provide a general framework for establishing threshholds or mappings for any given fuzzy quantifier.

The section that follows combines Zadeh's general approach with Fernadi's guidance about one possible mapping in order to demonstrate the utility of this general method for evaluating uncertainty in a social science context.

3.2.1 Example: Democratic Peace

To illustrate the value of this approach for social science research, I offer the following example evaluating a small subset of the democratic peace literature. Suppose that a researcher sought to evaluate the proposition:

$$p \triangleq \text{"Most democratic countries are peaceful"}$$

"Most" serves as a fuzzy quantifier on the fuzzy set "democratic countries," which must also meet the fuzzy condition (intersect with the fuzzy set) "peaceful." In order to evaluate this claim, I construct an EDF with the following components:

COUNTRY [Name]

DEMOCRATIC [Country; Year; μ]

PEACEFUL [Country; Year; μ]

MOST [Proportion; μ]

where COUNTRY is a list of country names; DEMOCRATIC contains those same country names alongside years, and a value μ_D with the Polity score corresponding to that country-year (normalized to be in [0,1]); PEACEFUL contains μ_P, a decreasing function of battle deaths (normalized to be in [0,1]); and MOST (μ_M) is composed of a series of conditions, following guidance specifically defining "most" (in relation to "few") in Farnadi et al. (2016):

$$\mu_M = \begin{cases} 0 & \text{if } p < 0.25; \\ \frac{(p-0.25)}{0.5} & \text{if } 0.25 \leqslant p < 0.75; \\ 1 & \text{if } p \geqslant 0.75. \end{cases}$$

With these conditions in place, I then conduct what Zadeh calls a "test procedure:" evaluating the EDF for degrees of compliance with the constraints specified above. This test procedure yields a set of "scores" for each of the relations (Zadeh 1983, 153). These scores are then aggregated into an overall test score, τ, which "is a vector which serves as a measure of the compatibility of the semantic entity with an instantiation of EDF" (153). That is, given our data (corpus), and given our constraints, τ represents the degree of agreement or consistency between our proposition and text. This degree of agreement or possibility, in turn, can be thought of as reflecting our uncertainty about the veracity of the proposition we sought to test with our data.

Zadeh does not define the procedure for constructing these scores specifically; rather, the paper appears to assume a general consensus or other source of information that would aid in evaluating the EDF (an assumption perhaps more plausible with simple calculations of income among friends or children's relative heights). For our example, evaluating p and establishing test scores without the use of a natural language corpus might imply several steps:

(1) Test DEMOCRATIC constraint

- Test the DEMOCRATIC constraint for each country
- Sum country values. Suppose that $\sum \mu_D(\mathtt{country_i}) = 52.3$, for example.

(2) Test PEACEFUL constraint

- Test the PEACEFUL constraint for each country
- Sum PEACEFUL values with DEMOCRATIC constraint: $\sum (\mu_D(\mathtt{country_i}) * \mu_P(\mathtt{country_i}))$. Suppose that value summed to 29.8, for example.

(3) Calculate the proportion of peaceful countries within the set of democracies

$\frac{29.8}{52.3} = 0.57$

(4) Calculate the test score for MOST constraint

$\mu_M(0.57) = \frac{0.57-0.25}{0.5} = 0.64$

This procedure suggests, then that given our definition of "most," the possibility that 57% of democratic countries are peaceful is given by:

$$\text{Poss(Peaceful(Democratic(Countries))} = 0.57) = \mu_M(0.57) = 0.64$$

With textual data, on the other hand, the task is to derive metrics that relate the variables and values in our EDF in order to test our proposition against its constraints. For this proposition, we can produce scores on associations using a word embedding approach. For each article, this implies first specifying the set of countries c of the total number of countries C referenced in the text (hand-coded in this example). Then, we can derive a set of associations between "democratic" and each country (d_c) and "peaceful" and each country (f_c), which in turn will allow us to evaluate the constraints as above against the "most" definition, again as above. This procedure relies heavily on the word choice in the original proposition to be tested. Practically, the procedure involves the following steps:

For documents $i \in J$,

(1) Discard if no mentions of "democratic" AND "peaceful" AND $\geqslant 1$ [country]

(2) Using GloVe, generate word embeddings

(3) For each country c:

   (a) Calculate cosine distance between c and "democratic" = d_c, rescale $\in [0, 1]$

   (b) Calculate cosine distance between c and "peaceful" = f_c, rescale $\in [0, 1]$

(4) Calculate democratic country values for document $d_i = \sum(d_c)$

(5) Calculate democratic AND peaceful country values for document $q_i = \sum(d_c * f_c)$

(6) Calculate proportion of democratic countries that are peaceful for document $r_i = \frac{q_i}{d}$

(7) Evaluate against the MOST condition: $\mu_M(r) = \tau$

The primary assumptions in this procedure surround the use of cosine distance in concepts, as evaluated via word embeddings/co-occurrence of terms, as an appropriate metric for the degree of fit within a fuzzy set. That is, in this setup, d_{c_i} is thought to measure how "democratic" document i argues or expresses or assumes country c to be (or, the ease with which the document might find that country compatible with our definition of "democracy"). This assumption is further complicated at the theoretical level because the constraint evaluations are conducted within-document rather than across-document; only the conception of "most" is held constant across all documents in the corpus. That is, for a given corpus, this framework does not restrict that possibility that each document contains its own conception of democracy or a differing notion of its fuzzy boundaries.

To further illustrate, I implement this procedure with a set of published papers. Tomz and Weeks (2013), Maoz and Russett (1993), and Mesquita et al. (1999) are papers that broadly support the "democratic peace theory," that democracies are unlikely to go to war with one another. As such, the expectation is that these papers will be more "certain" in their assessment of our proposition (that is, they should reflect a higher possibility of the statement). Layne (1994) and Rosato (2003), meanwhile, dispute the theory, and therefore are unlikely to comport with our proposition as stated with a high degree of certainty. Finally, Wimmer, Cederman, and Min (2009) and Fearon and Laitin (2003) are not substantively *about* democratic peace theory; both examine the role of ethnicity in civil conflict—largely as a domestic phenomenon and not in the context of interstate war. These documents should therefore have the least "certainty" about the proposition. Given the ambiguity of language, however, no document is likely to reflect 100% or 0% possibility of the proposition regardless of measurement. In the context of word embeddings, furthermore, these boundary values would present problems for theoretical interpretation (e.g., a cosine distance of 1 between "England" and "democracy" would imply that England *is* democracy, rather than being *a* democracy).[5]

Evaluating these papers individually, without reference to a baseline dataset and without pre-trained word embeddings, yields the results depicted in Figure 9. The articles are arrayed in the order expected, and furthermore indicate at least some variation among articles of a given "type." This variation is particularly useful in illustrating the value of uncertainty measurement from textual data, since these distinctions in possibility values can reflect both the weight of the evidence (e.g., which countries are chosen for discussion of peace and democracy) as well as the underlying structure of the argument (e.g., exemplary countries for "democratic" or "peaceful" are likely more proximate to those terms when articulating a thesis).

[5]It is worth noting here that cosine distance values are rescaled empirical min and max of these data, rather than from the full theoretical [-1, 1] range. Document ordering in the results remains the same but variation is compressed with rescaling from the full range.

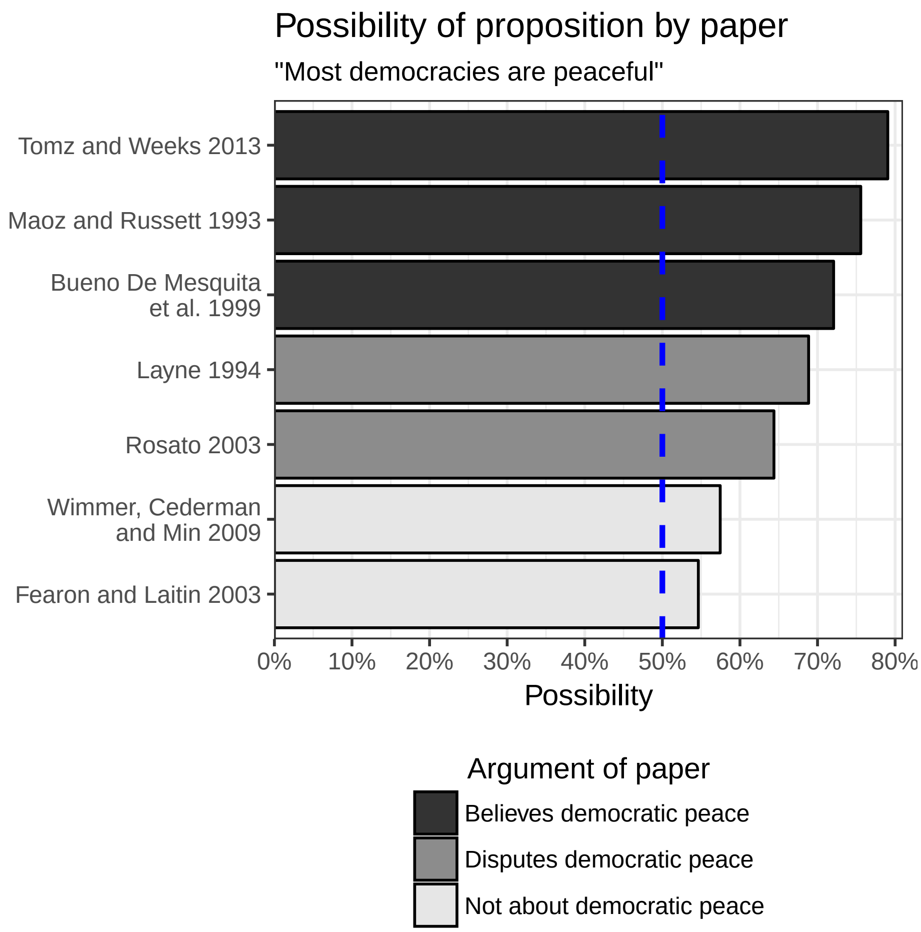

Figure 9: Proposition Testing by Article: Democratic Peace

The individual-paper results in Figure 9, however, do not benefit from the robustness that word embeddings trained on a much larger corpus offer. While constructing a social-science specific corpus from which to derive embeddings is possible in principle, evaluating uncertainty in individual papers or monographs from such a corpus would nevertheless be subject to volatility because of the sparsity of keywords, as the example in the previous section illustrated. As such, an augmentation to this approach adapts the difference metric illustrated in the previous section, but in this case used within the full fuzzy quantifier framework rather than merely via a dictionary-based comparison.

Specifically, the 7-step procedure above can be applied to the pre-trained embeddings from Wikipedia as before, and then compared to the individual embeddings calculated for each paper evaluating specifically the democratic peace, with the same difference-in-cosine-distance metric evaluating the change in (un)certainty relative to baseline. Figure 10depicts the percentage point change in certainty relative to the baseline for each article previously considered. Percentage point change is calculated on rescaled values for cosine distance as follows: for each of the individual papers under consideration, minimum and maximum values for parameters of interest (democratic and peaceful in steps 3(a) and 3(b)), with respect to any country, are linearly rescaled from $(\min, \max)$ to $[0, 1]$ to generate a conservative measure. These rescaled values are then used in the summation in steps 4 and 5. The Wikipedia embeddings are separately rescaled. Notably, this visualization includes all countries for which Wikipedia has coverage: a much larger set than that discussed by any individual paper.

Figure 11 limits the set of countries evaluated in the per-trained embeddings to those discussed by the articles under investigation, generating slightly more uncertainty with respect to the proposition. Whereas in the previous case, $r_{\text{wiki}} = 0.46$ (score= 0.37), it now evaluates to 0.42 (score= 0.23). This shift is encouraging with respect to the broader "objectivity" and/or consistency of the pre-trained baseline, since the

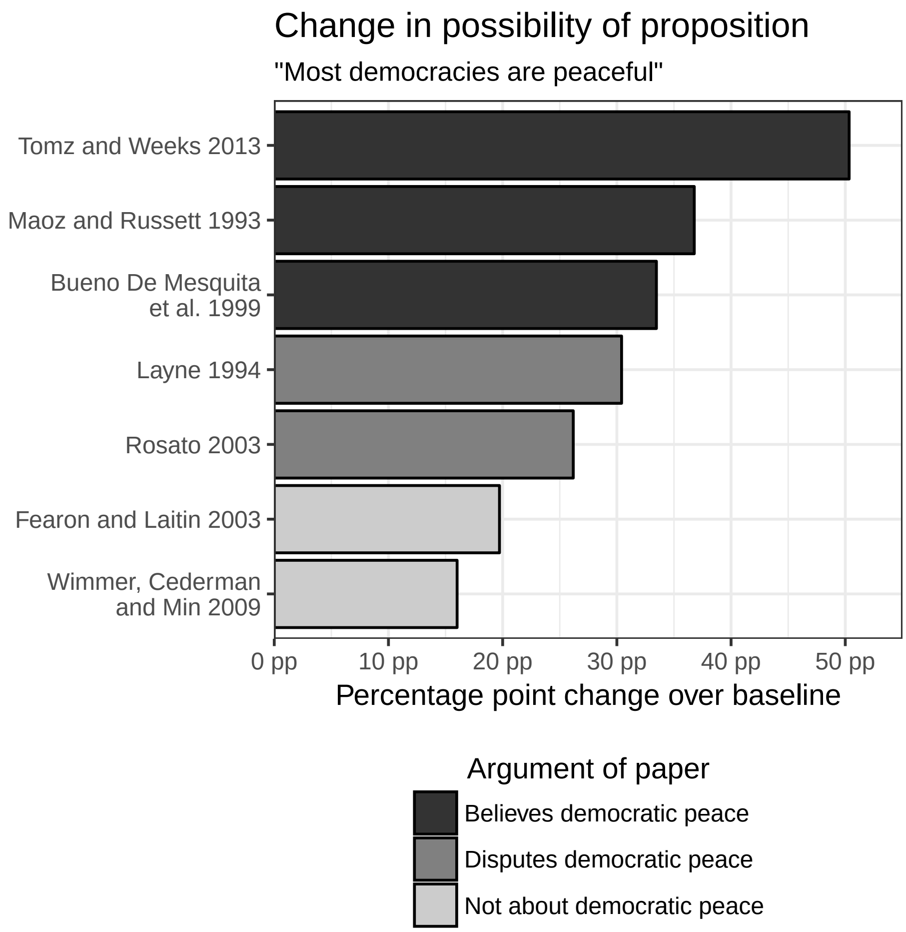

Figure 10: Article Comparison to Pre-Trained Embedding Baseline

uncertainty is greater when evaluating the proposition with respect to less "evidence." One possible explanation for greater certainty overall in academic articles is via favorable case-selection, or via omitting disconfirming evidence, suggesting that the comparison to a limited country set for the baseline is more ideal. Still, another aim is to provide an overarching uncertainty assessment for a given proposition, which does not itself require holding evidence constant; on the contrary, it may suggest authors' evaluations of divergent evidence as the basis for uncertainty (i.e., the uncertainty may be empirical rather than epistemic). Using the baseline embeddings for comparison, however, affords a corrective in terms of assessing the relative uncertainty of academic articles for a single proposition, and potentially for text sources written with different audiences in mind (e.g., a mix of academic articles and journalistic pieces).

3.2.2 Discussion

This method of evaluating uncertainty via "possibility," particularly relative to the exclusively dictionary-based approaches articulated earlier. Both approaches are subject to some degree of measurement error, particularly as it pertains to evaluating which instances of fuzzy quantifier usage are relevant to the proposition ("Most democracies do not go to war") versus those that are less so ("Most of the literature on democracy to date has emphasized..."). Yet unlike dictionary-based methods that would ultimately rely on frequency of fuzzy quantifiers to determine "uncertainty," this approach has much greater flexibility and theoretical coherence. Distinctions among documents utilizing a complete fuzzy quantifier framework are more likely related to intentional argumentative structure rather than idiosyncratic and highly susceptible to the choice of aggregation rule, as in the dictionary-based cases. That said, specifying the fuzzy quantifier constraint for this approach presents a greater theoretical challenge.

The fungibility of the "most" constraint is also potentially problematic. Placing the burden on individual researchers to specify the bounds for each fuzzy quantifier they

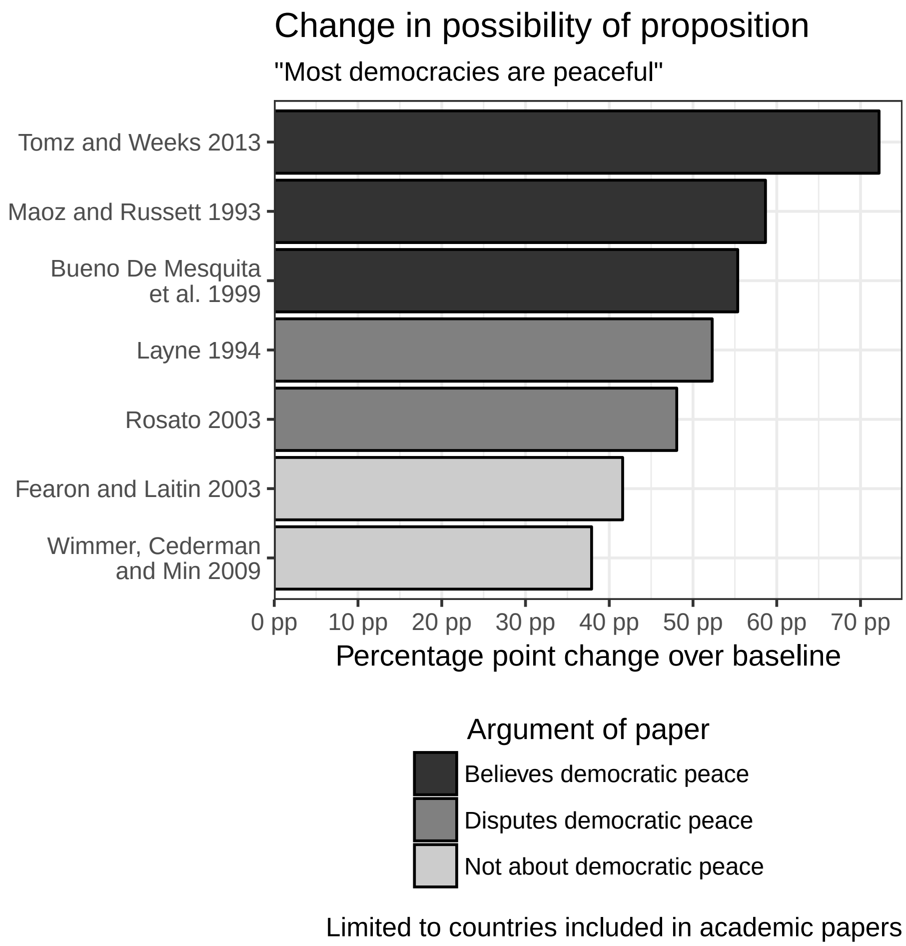

Figure 11: Article Comparison to Pre-Trained Embedding Baseline: Limited Country Set

seek to evaluate in propositions seems cumbersome and raises difficulties for generalizability across research. On another hand, as Zadeh elsewhere notes, "information is a generalized constraint on the values which a variable is allowed to take" (Zadeh 2005, 2–3). In this sense, the flexibility of these constraints makes them more easily updateable with new information and new data, or to new conceptions of what might constitute set membership. Even with a requirement to generate fuzzy quantifier constraints in each new project, the fuzzy quantifier framework at least facilitates transparent and testable statements of researchers' beliefs and arguments.

A clear direction for future improvement is in finding alternative ways to specify the fuzzy quantifier constraints. As mentioned previously, one mechanism would be through a large-scale human coding exercise or survey, in which respondents evaluated possibility bounds for given fuzzy quantifier terms. Alternatively, simulations could generate tresholds with respect to particular empirical cases (these thresholds may differ across corpora, for example), to provide general guidelines for the "best" bounds, whether those are the most inclusive or the most restrictive. Furthermore, with several large-scale corpora, it would be possible to create bounds or priors for fuzzy quantifier values, much in the same way large repositories of text make it feasible to pre-fit word embeddings. Alternatively, incorporating some degree of dictionary-based detection of fuzzy quantifier terms in relevant articles might allow for endogenously generating the fuzzy quantifier constraint from within a particular corpus or corpora.

4 Conclusions & Extensions

This paper provides a preliminary sense of the ways in which fuzzy quantifiers and fuzzy logic can undergird attempts to estimate uncertainty from text for social science applications. Aside from the pre-labeled corpora used previously for comparison to subjectivity analyses, the dictionary-based and broader fuzzy logic framework articulated here can apply to a multitude of differing data formats relevant for social science questions.

4.1 Additional Applications

The applications discussed above presume textual data coming from published research, whether in the form of papers or monographs. The methods, however, apply just as easily to other types of textual or qualitative data in social science contexts. For interview or archival data, fuzzy quantification with respect to propositions in particular could provide more accurate assessments of beliefs and uncertainty, particularly for respondents or authors who are not trained in social science methods and frameworks or statistics, or for sources from disparate time periods or locations where validation and confirmation is challenging or impossible.

Likewise, these methods could apply to open-ended survey responses, where evaluating subjects' uncertainty about questions or claims has traditionally presented a challenge. For example, survey responses with both open- and closed-ended questions could conduct the fuzzy quantifier procedure described above on textual data and correlate with levels of uncertainty expressed elsewhere for validity and scaling. A preliminary assessment using the 2016 American National Election Studies (ANES) time series data indicates that fewer than 5% of responses (per question on average) to closed-ended questions are "I don't know," but that these responses are weakly positively associated with fuzzy quantifier usage in corresponding open-ended questions (American National Election Studies 2016). More long-form open-ended survey questions and other cues for uncertainty could further confirm and refine the relationship between fuzzy quantifier terminology and construct usage and uncertainty expression.

With published works, conversely, an additional dimension to consider is editorial effects on uncertain expression. A preliminary investigation of Wikipedia articles indicated no significant relationship between fuzzy quantifier usage and the number of editors or edits for a given page (see Appendix), but for academic journal publications or other venues, editorial style or editorial board fixed effects may dictate norms around either fuzzy quantifier term usage or standards and bounds that define particular fuzzy quantifiers. Further investigating uncertainty levels not only at the document and corpus level, but also according to author and editorial staff, would further illuminate the locus from which uncertainty expression conventions arise.

4.2 Corpus Selection

Fuzzy logic and fuzzy set frameworks have numerous possible applications in other aspects of text analysis that are particularly relevant to the social sciences. It may, in fact, be useful in structuring the way text analysis is performed. While in other disciplines, corpora may be more readily defined by the task (e.g., medical journal articles evaluating breast cancer treatment prognoses, or the collected works of Ayn Rand), establishing the scope of corpora for many social and political science questions has traditionally been done by argumentation. Fuzzy set logic could apply to the inclusion of documents in a corpus, providing not only a set of possible corpus members but also establishing degrees of membership.

4.3 Multi-Lingual Applications

Evaluating uncertainty across multiple languages presents its own unique challenges (Bentz and Alikaniotis 2016). Languages differ in their conceptions of measurement and time, which can complicate efforts to precisely or quantitatively characterize arguments within the text. Languages like Amharic are in fact characterized by their distinct ambiguity, particularly in spoken language, but also in terms of word choice and order (Amare 2001), whereas languages like Burmese would generate problems for leveraging precisely the theoretical distinctions this paper attempts to articulate. In Burmese, by way of example, no distinction exists between probability and possibility; likewise, saying an event or condition is “likely” is analogous to saying it is “possible.” These distinctions are further complicated, depending on the context or type of text, by culturally dictated mechanisms for maintaining safety via vagueness or omission (e.g., if asked a sensitive question about a political situation, one might answer in Burmese, “could be” or “it’s possible”). Certain kinds of fuzzy quantification, however, are more consistent across language frames despite, or because of, their indeterminate reference sets. “Many,” “most,” or “few,” are less culturally moored assessments of uncertainty than typical subjectivity terms, and therefore may serve as a more consistent basis for establishing cross-language classification and measurement schema for uncertainty.

4.4 Refinements to Word Embeddings

Word embeddings have a series of disadvantages for this particular set of applications. Word embedding algorithms (with fasttext as one possible exception) have difficulty with words not found in the training vocabulary. While the pre-trained embeddings from GloVe circumvent this problem somewhat temporarily, extending this project to a set of social-science-specific embeddings, or applying it to very specific or esoteric papers could generate problems assessing fuzzy quantifiers in conjunction with keywords. More generally, the illustrative applications above use pre-trained embeddings (in this case, the 2014 Wikipedia dump), which demonstrate consistent efficiency and accuracy gains over simultaneous learning of embeddings, particularly with smaller corpora, but a more bespoke set of embeddings tailored to social science research—whether from published articles analogous to the PubMed databases, or via other means—might perform better in the task of evaluating truly “deviant” uses of fuzzy quantifiers on important subjects. Expanding and refining the set of embeddings would in turn potentially enable the method to extend to additional applications.

The multitude of possible extensions from a fuzzy quantifier framework for evaluating textual uncertainty suggests precisely the potential for these theoretical orientations and methodologies for a variety of social science applications. While the evidence presented here is merely suggestive, rather than definitive, it serves to indicate multiple opportunities for future expansion and refinement.

References

Akkaya, Cem, Janyce Wiebe, and Rada Mihalcea. 2009. "Subjectivity Word Sense Disambiguiation." *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing:* 190–199.

Amare, Getahun. 2001. "Towards the Analysis of Ambiguity in Amharic." *Journal of Ethiopian Studies* 34 (2): 35–56.

American National Election Studies. 2016. "THE ANES GUIDE TO PUBLIC OPINION AND ELECTORAL BEHAVIOR." *University of Michigan, Center for Political Studies [producer and distributor].* Ann Arbor, MI. `www.electionstudies.org`.

Atanassov, Krassimir, and George Gargov. 1998. "Elements of Intuitionistic Fuzzy Logic, Part I." *Fuzzy Sets and Systems* 95:39–52.

Auger, Alain, and Jean Roy. 2008. "Expression of Uncertainty in Linguistic Data." *Proceedings of the 11th International Conference on Information Fusion (FUSION):* 1860–1867.

Barwise, Jon, and Robin Cooper. 1981. "Generalized Quantifiers and Natural Language." *Linguistics and Philosophy* 4:159–219.

Benoit, Kenneth, Michael Laver, and Slava Mikhaylov. 2009. "Treating Words as Data with Error: Uncertainty in Text Statements of Policy Positions." *American Journal of Political Science* 53 (2): 495–513.

Bentz, Christian, and Dimitrios Alikaniotis. 2016. "The Word Entropy of Natural Languages." *arXiv.* `arxiv.org/1606.06996`.

Breck, Eric, and Claire Cardie. 2014. "Opinion Mining and Sentiment Analysis." In *The Oxford Handbook of Computational Linguistics,* 2nd ed., edited by Ruslan Mitov. Oxford: Oxford University Press.

Cabrerizo, F., R. Al-Hmouz, A. Morfeq, A. Balamash, M. Martìnez, and E. Herrera-Viedma. 2017. "Soft Consensus Measures in Group Decision Making Using Unbalanced Fuzzy Linguistic Information." *Soft Computing* 21:3037–3050.

Chen, Chaomei, Ming Song, and Go Eun Heo. 2017. "A Scalable and Adaptive Method for Finding Semantically Equivalent Cue Words of Uncertainty." *arXiv.* `arxiv.org/1710.08327`.

Chen, Jidong, Jennifer Pan, and Yiqing Xu. 2015. "Sources of Authoritarian Responsiveness: A Field Experiment in China." *American Journal of Political Science* 60 (2): 383–400.

Conde-Clemente, Patricia, Jose M. Alonso, Èldman O. Nunes, Angel Sanchez, and Gracian Trivino. 2017. "New Types of Computational Perceptions: Linguistic Descriptions in Deforestation Analysis." *Expert Systems with Applications* 85:46–60.

Denny, Matthew J., and Arthur Spirling. 2018. "Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It." *Political Analysis* 26 (2): 168–189.

Dhar, Mamoni. 2013. "Cardinality of Fuzzy Sets: An Overview." *International Journal of Energy, Information and Communications* 4 (1): 15–22.

Dìaz-Hermida, F., A. Bugarìn, and S. Barro. 2003. "Definition and Classification of Semi-Fuzzy Quantifiers for the Evaluation of Fuzzy Quantified Sentences." *International Journal of Approximate Reasoning* 34:49–88.

Dragos, Valentina. 2013. "An Ontological Analysis of Uncertainty in Soft Data." *16th International Conference on Information Fusion (FUSION).* https://www.researchgate.net/publication/261316990_An_ontological_analysis_of_uncertainty_in_soft_data.

Druzdzel, Marek. 1989. "Verbal Uncertainty Expressions: Literature Review." *Technical Report: CMU-EPP 1990-03-02.*

Farkas, Richàrd, Veronika Vincze, György Mòra, Jànos Csirik, and György Szarvas. 2010. "The CoNLL-2010 Shared Task: Learning to Detect Hedges and their Scope in Natural Language Text." *Conference on Computational Natural Language Learning (CoNLL).* http://www.aclweb.org/anthology/W10-3001.

Farnadi, Golnoosh, Stephen H. Bach, Marjon Blondeel, Marie-Francine Moens, Lise Getoor, and Martine De Cock. 2016. "Statistical Relational Learning with Soft Quantifiers." In *Inductive Logic Programming 25th International Conference Selected Papers,* edited by Katsumi Inoue, Hayato Ohwada, and Akihiro Yamamoto, 61–75. Springer International.

Fearon, James D., and David D. Laitin. 2003. "Ethnicity, Insurgency, and Civil War." *American Political Science Review* 97 (1): 75–90.

Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21 (3): 267–297.

Jean, Pierre-Antoine, Sèbastien Harispe, Sylvie Ranwez, Patrice Bellot, and Jacky Montmain. 2016. "Uncertainty Detection in Natural Language: A Probabilistic Model." *Proceedings of the 6th International Conference on Web Intelligence, Mining and Semantics.*

Khoury, Richard, Fakhreddine Karray, and Mohamed Kamel. 2008. "Domain Representation Using Possibility Theory: An Exploratory Study." *IEEE Transactions on Fuzzy Systems* 16 (6): 1531–1541.

Layne, Christopher. 1994. "Kant or Can't: The Myth of the Democratic Peace." *International Security* 19 (2): 5–49.

Li, Xiujun, Wei Gao, and Jude W. Shavlik. 2014. "Detecting Semantic Uncertainty by Learning Hedge Cues in Sentences Using an HMM." *Proceedings of Workshop on Semantic Matching in Information Retrieval:* 30–37.

Lindley, Dennis V. 1987. "The Probability Approach to the Treatment of Uncertainty in Artificial Intelligence and Expert Systems." *Statistical Science* 2 (1): 17–24.

Liu, Yaxin, and Etienne E. Kerre. 1998. "An Overview of Fuzzy Quantifiers I: Interpretations (Invited Review)." *Fuzzy Sets and Systems* 95:1–21.

Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schütze. 2008. *Introduction to Information Retrieval.* Cambridge: Cambridge University Press.

Maoz, Zeev, and Bruce Russett. 1993. "Normative and Structural Causes of Democratic Peace, 1946–1986." *American Political Science Review* 87 (3): 624–638.

Mesquita, Bruce Bueno de, James D. Morrow, Randolph M. Siverson, and Alastair Smith. 1999. "An Institutional Explanation of the Democratic Peace." *American Political Science Review* 93 (4): 791–807.

Miller, Ben, Ayush Shrestha, Jason Derby, Jennifer Olive, Karthikeyan Umapathy, Fuxin Li, and Yanjun Zhao. 2013. "Digging into Human Rights Violations: Data Modelling and Collective Memory." *2013 IEEE International Conference on Big Data:* 37–45.

Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. 2014. "GloVe: Global Vectors for Word Representation." `https://nlp.stanford.edu/pubs/glove.pdf`.

Pherson, Katherine Hibbs, and Randolph H. Pherson. 2012. *Critical Thinking for Strategic Intelligence.* Thousand Oaks, CA: Sage/CQ Press.

Ramos-Soto, A., and M. Pereira-Farina. 2017. "On Modeling Vagueness and Uncertainty in Data-to-Text Systems through Fuzzy Sets." *arXiv.* `arxiv.org/1710.10093`.

Roos, Nico. 1990. "How to Reason with Uncertain Knowledge." *Proceedings of the 3rd International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems:* 403–412.

Rosato, Sebastian. 2003. "The Flawed Logic of Democratic Peace Theory." *American Political Science Review* 97 (4): 585–602.

Rubin, Victoria L., Elizabeth D. Liddy, and Noriko Kando. 2006. "Certainty Identification in Texts: Categorization Model and Manual Tagging Results." Chap. 7 in *Computing Attitude and Affect in Text: Theory and Applications. The Information Retrieval Series,* edited by James Shanahan, Yan Qu, and Janyce Wiebe, 20:61–76. Springer.

Saif, Hassan, Miriam Fernandez, Yulan He, and Harith Alani. 2014. "On Stopwords, Filtering and Data Sparsity for Sentiment Analysis of Twitter." *Language Resources Evaluation Conference (LREC) Proceedings.* `http://www.lrec-conf.org/proceedings/lrec2014/pdf/292_Paper.pdf`.

Stanford NLP Group. 2018. "Pre-Trained Embeddings." *GloVe Model for Distributed Word Representation.* `https://github.com/stanfordnlp/GloVe`.

"Szeged Uncertainty Corpus." 2010. http://rgai.inf.u-szeged.hu/index.php?lang=en&page=uncertainty.

Szmidt, Eulalia, and Janusz Kacprzyk. 2001. "Entropy for Intuitionistic Fuzzy Sets." *Fuzzy Sets and Systems* 118:467–477.

Thomson, Judi, Beth Hetzler, Alan MacEachren, Mark Gahegan, and Misha Pavel. 2005. "A Typology for Visualizing Uncertainty." *Conference on Visualization and Data Analysis, IS&T/SPIE Symposium on Electronic Imaging.*

Tomz, Michael R., and Jessica L. P. Weeks. 2013. "Public Opinion and the Democratic Peace." *American Political Science Review* 107 (4): 849–865.

Vincze, Veronika. 2014. "Uncertainty Detection in Natural Language Texts." *Dissertation, Research Group on Artificial Intelligence and University of Szeged.* http://doktori.bibl.u-szeged.hu/2291/1/Vincze_Veronika_tezis.pdf.

Wang, Yixin, and David M. Blei. 2018. "The Blessings of Multiple Causes." *arXiv: 1805.06826v2* 19 June.

Wimmer, Andreas, Lars-Erik Cederman, and Brian Min. 2009. "Ethnic Politics and Armed Conflict: A Configurational Analysis of a New Global Data Set." *American Sociological Review* 74:316–337.

Wygralak, Maciej. 1998. "Vagueness and its Representations: A Unifying Look." *Mathware and Soft Computing* 5:121–131.

Zadeh, Lotfi A. 1983. "A Computational Approach to Fuzzy Quantifiers in Natural Languages." *Computers and Mathematics with Applications* 9 (1): 141–184.

———. 2005. "Toward a Generalized Theory of Uncertainty (GTU)–An Outline." *Information Sciences* 172:1–40.

———. 2008. "Is there a need for fuzzy logic?" *Information Sciences* 178:2751–2779.

Zhai, Daoyuan, and Jerry M. Mendel. 2011. "Uncertainty Measures for General Type-2 Fuzzy Sets." *Information Sciences* 181:503–518.

zonination. 2017. "Perceptions of Probability and Numbers." *Reddit, via r/samplesize.* https://github.com/zonination/perceptions.

A Fuzzy Quantifiers List

Quantifier	Source
about	Liu and Kerre (1998)
additional	Conde-Clemente et al. (2017)
all	Zadeh (1983)
almost	Zadeh (1983)
another	BioScope (hand-coded)
any	Conde-Clemente et al. (2017)
approximate	Zadeh (1983)
around	BioScope (hand-coded)
as many	Zadeh (1983)
at least	Zadeh (1983)
between	BioScope (hand-coded)
broad	BioScope (hand-coded)
certain	BioScope (hand-coded)
couple	BioScope (hand-coded)
few	Zadeh (1983), Farnadi et al. (2016)
frequently	BioScope (hand-coded)
great	BioScope (hand-coded)
half	Zadeh (1983)
high	Conde-Clemente et al. (2017)
increased	BioScope (hand-coded)
less	Zadeh (1983)
low	Conde-Clemente et al. (2017)
many	Zadeh (1983)
minority	Liu and Kerre (1998)
more	Zadeh (1983)
most	Zadeh (1983), Farnadi et al. (2016)
multiple	BioScope (hand-coded)
narrow	BioScope (hand-coded)
nearly	BioScope (hand-coded)
no	Liu and Kerre (1998)
none	Zadeh (1983)
often	Zadeh (1983)
rarely	BioScope (hand-coded)
reduced	BioScope (hand-coded)
seldom	BioScope (hand-coded)
several	Zadeh (1983)
slight	BioScope (hand-coded)
some	Zadeh (1983)
twice	BioScope (hand-coded)
ubiquitous	BioScope (hand-coded)
within	BioScope (hand-coded)

B Fuzzy Quantifiers and Edits on Wikipedia

One possible concern with the use of Wikipedia-trained word embeddings is an uneven distribution of fuzzy quantifier usage around particular topics, or arising from particular editorial styles and systems, that mean associations derived from the corpus are biased, particularly for subject matter of interest to social scientists. For example, if all articles on contentious topics have a larger number of editors or a larger number of edits, the usage of fuzzy quantifiers may proliferate (or alternatively, bottom out), and may be measuring polarization around the topic rather than uncertainty, strictly speaking, or may simply encompass imprecision of language due to a large number of contributors, which is likely theoretically different from substantive uncertainty.

The figures below provide a simple, superficial illustration for whether these types of relationships are likely to hold. Each figure uses text from articles across three broad subjects—quantum mechanics, the "War on Terror," and World War II—to assess fuzzy quantifier usage, with each serving as a different example of a "contentious" subject matter. Quantum mechanics provides an example of a topic that is both "scientific" and features some contention (with classical mechanics) for a (likely) small population of editors; the "War on Terror," in contrast, is a more contemporary topic with potentially greater polarity among editors; and World War II is one of the largest topics (by number of articles and stubs), featuring both a series of short, timeline-like stubs and longer social-historical analyses that span the spectrum of contention and uncertainty. Figure 12 indicates that, much like the other text sources examined here, fuzzy quantifiers are fairly sparse across a variety of topics, ranging from "objective" scientific subject matter to explicitly contentious issues such as the War on Terror. While not definitive, this cursory investigation indicates that fuzzy quantifier usage on Wikipedia may not differ substantially from what we should "expect" in other published or edited media formats.

Likewise, an examination of fuzzy quantifier usage by word relative to the number

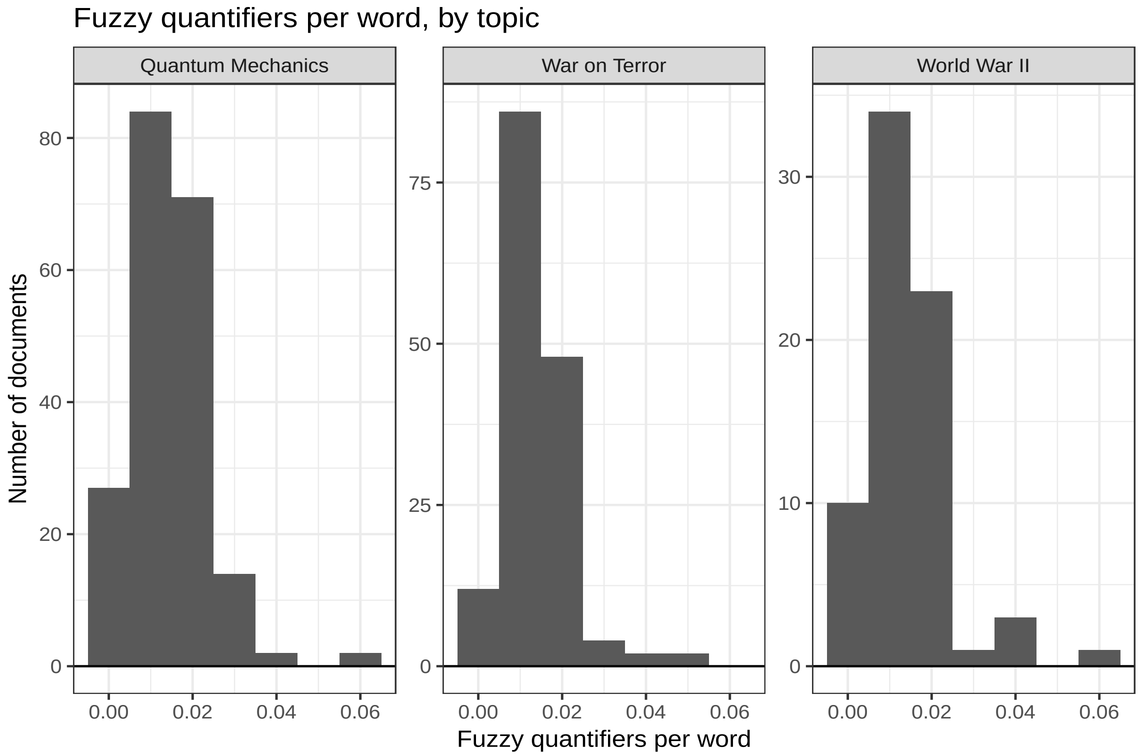

Figure 12: Wikipedia: Fuzzy Quantifier Usage by Word across 3 Example Topics

of editors and the number of edits that articles and stubs within a given topic receive indicates very a very slight, but not systematic, positive association with fuzzy quantifier usage. That these broad relationships appear largely the same (largely characterized by noise), irrespective of the "objectivity" of the subject matter in these examples, is encouraging for the prospect that fuzzy quantifier usage on Wikipedia in general should provide an adequate baseline.

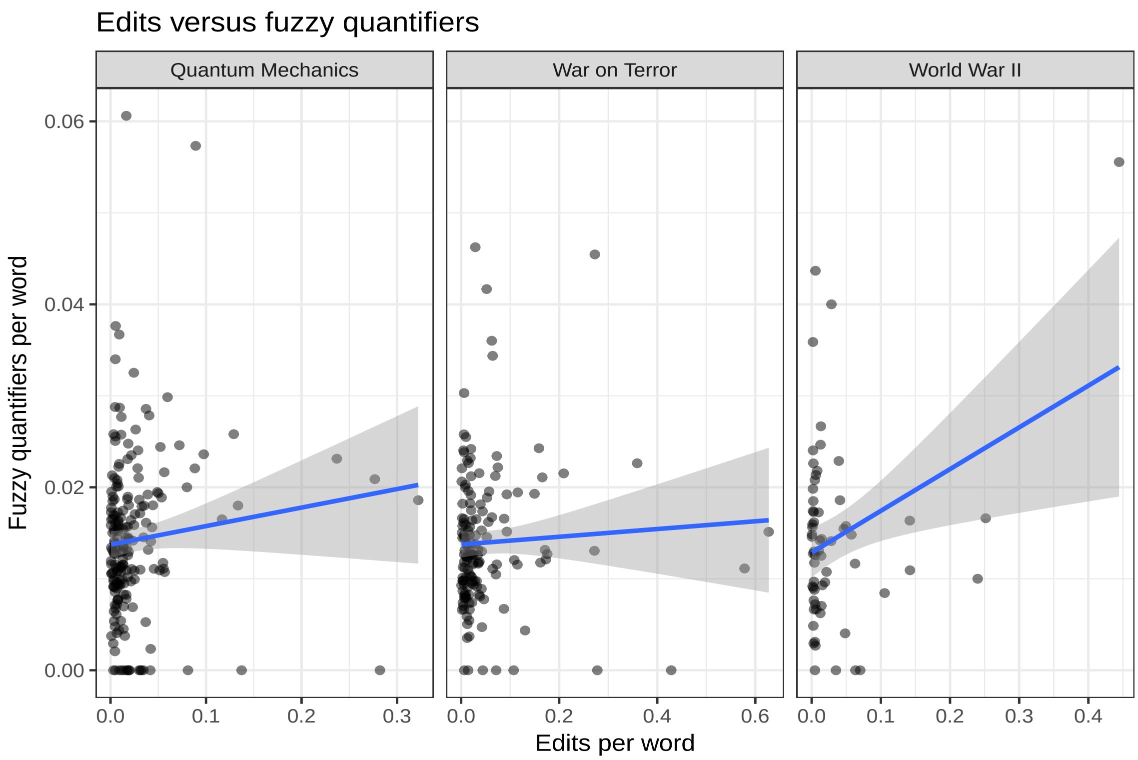

Figure 13: Wikipedia: Fuzzy Quantifier Usage vs. Number of Edits by Topic, Examples

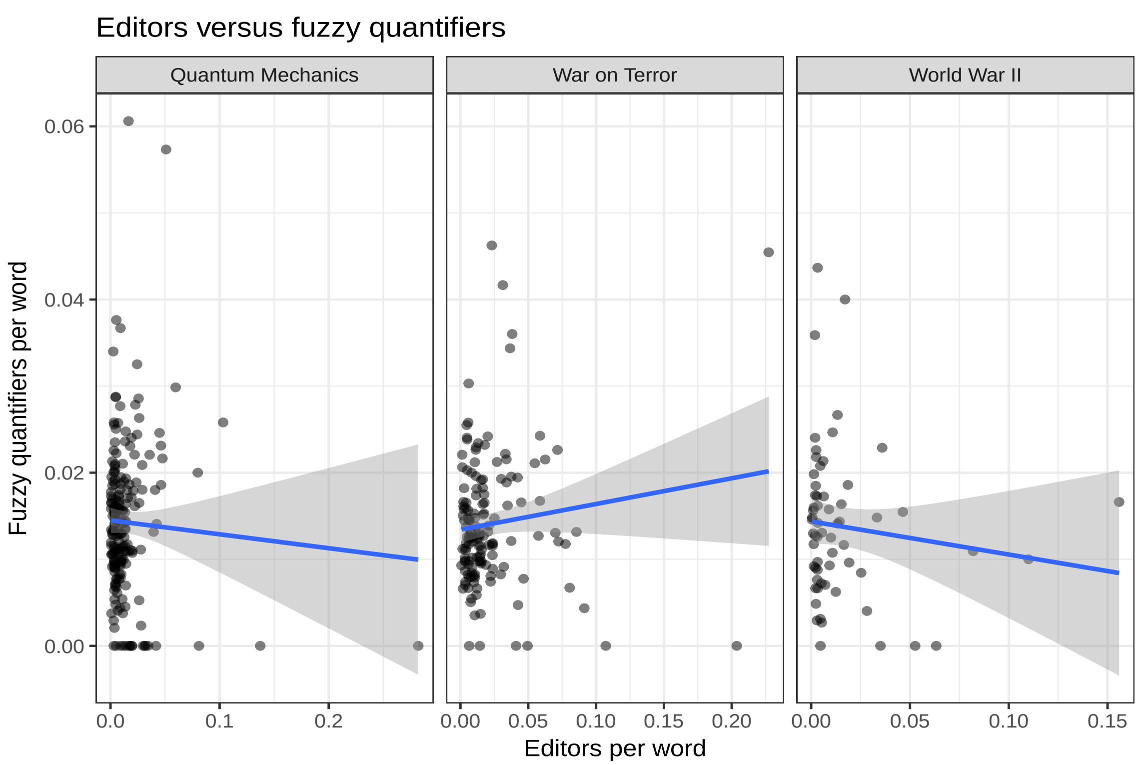

Figure 14: Wikipedia: Fuzzy Quantifier Usage by Number of Editors by Topic, Examples