Brigham Young University

BYU ScholarsArchive

2020-06-08

Gender Vs. Sex: Defining Meaning in a Modern World through use of Corpora and Semantic Surveys

Mary Elizabeth Garceau
Brigham Young University

Gender Vs. Sex: Defining Meaning in a Modern World

Through Use of Corpora and Semantic Surveys

Mary Elizabeth Garceau

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Arts

William Eggington, Chair
Wendy Baker Smemoe
Earl K. Brown

Department of Linguistics

Brigham Young University

ABSTRACT

Gender Vs. Sex: Defining Meaning in a Modern World
Through Use of Corpora and Semantic Surveys

Mary Elizabeth Garceau
Department of Linguistics, BYU

Master of Arts

Considerable resources in U.S. legal studies are devoted to determining the precise meaning of contested terms specifically in statutory interpretation. Traditional judicial approaches have defined meaning using dictionaries. This reliance has led to Mouritsen's (2010) observation that "the judicial conception of lexical meaning—i.e., what judges think about what words mean … is often [subjectively] outcome determinative." Beginning with Mouritsen's (2010) article, a movement in U.S. legal scholarship offers corpus linguistics as a more objective method to resolving contested meaning (Lee and Mouritsen, 2018). However, I assert that weaknesses still exist in contemporary applications of corpus linguistics to legal interpretation. I first review methodological differences in two corpus-based projects that attempt to resolve the meaning of the contested term, "emoluments," a high-profile Supreme Court-bound contemporary issue related to the legitimacy of the Trump presidency (Phillips and White, 2018; Cunningham and Egbert, 2019). Unfortunately, the results of these two studies are in conflict. Based upon a critique of these projects, I advocate for a more objective method of interpreting the results of corpus analyses using multiple human coders following rater reliability research models often used in sociolinguistics and second language acquisition research. In order to test our assumptions, I apply this approach to utilizing corpus linguistics to define the meaning of "sex" in two highly charged cases pending in the U.S. Supreme Court within the context of Title VII of the Civil Rights Act of 1964 which prohibits discrimination "because of. . . sex" (42 U.S.C. § 2000e-2(a)(1). The first case, Harris Funeral Home v. EEOC, questions if "sex" encompasses "gender identity;" while the second, Altitude v. Zarda, asks if the meaning of "sex" includes "sexual orientation." I discuss results of this research model and its implications to further corpus linguistic applications to the law.

TABLE OF CONTENTS

1 INTRODUCTION

Social and political tensions now surround the words *gender* and *sex*. For example, a recent blog post by the Powerline blog embodies these tensions by stating that "'gender' went from a boring word that was a prissy way of saying 'sex' to a leftist battleground, in a very short period of time" (Hinderaker, 2019). This statement is accompanied by a chart showing the frequency of the use of the word "gender" by the New York Times, from the years 1970-2017 (Figure 1)[1].
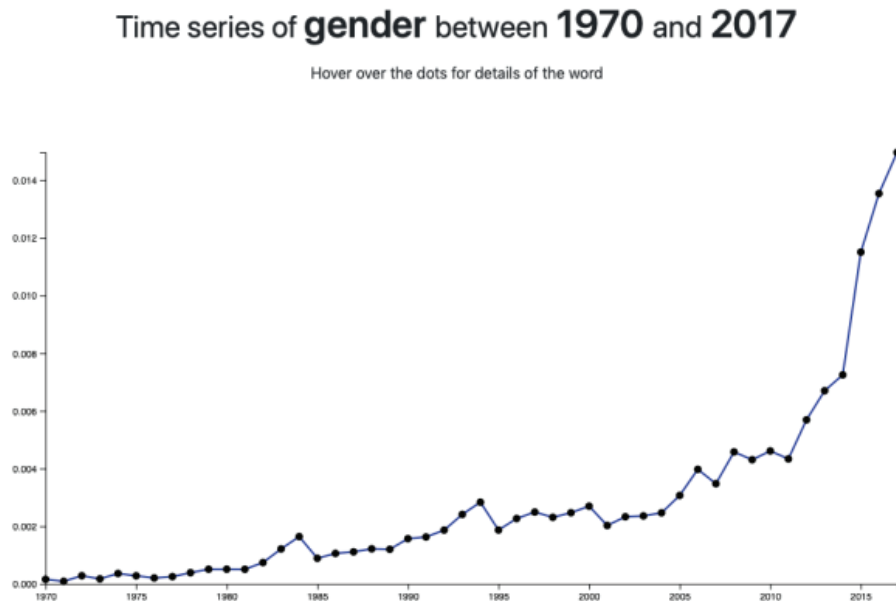
Figure 1: Time Series of Gender between 1970 and 2017

---

[1] The description from the New York Times Media Analytics website reports this database as, "A site that allows users to query a corpus of all articles written by the New York Times between 1970 and 2018. The Timeline functionality allows for tracking the frequency of word usage over time in the New York Times data corpus" (2020). There is no mention about the data being normalized to a base number, however. As can be seen in the chart, a positive spike in frequency occurs around the year 2016 reflecting a stark change from past usage. Nevertheless, if these proposed "spikes" in usage are taken from raw data, instead of comparing the total occurrences of the word gender with the total number of words in the target text, then this data may not represent reality.

Despite some issues with the way the data is portrayed, it is certainly obvious that there is more current usage of the word *gender*, as further evidenced by data from the Corpus of Contemporary American English (COCA) (Davies, 2020). COCA is a corpus with over 1 billion words taken from newspapers, magazines, television, books, articles, and the internet. A simple search of the term *gender* on this corpus including charts that show the year of occurrence, show that from 1990-1994 there are 3979 total occurrences of the word *gender*, compared to 2015-2019 which shows 6343 occurrences of *gender*.

Not only is there more usage of the word *gender*, but there is also much confusion and discussion as to the actual meanings of the word *gender* vs. the meanings of the word *sex*. Contemporary confusions regarding these meanings can lead to real-life (and death) consequences. For example, in May 2019, an article from a self-proclaimed "conservative blog" entitled "Pregnant Female Identifying As Male Goes To ER With Urgent Symptoms, Is Listed As 'Obese Man.' She Receives No Urgent Care, Gives Birth To A Dead Baby" describes tragic consequences that can result from such confusion (Parker, 2019). This incident occurred because, on the medical record, the patient listed "male" as their[2] sex. Because "male" was listed, the patient was not examined for pregnancy. The patient is now suing the hospital for negligence. This is not the only incident of its kind, nor will it be the last, thus prompting the question: what are the contemporary differences, semantically, between *gender* and *sex*? In the incident described above, if we assume that *sex* concerns biological attributes, then the correct option to list would have been "female" regardless of the individual's psychological identity.

An additional cause for concern regarding the meanings of *sex* and *gender* lies in the field of women's sports, specifically that transgender women may be allowed to compete in women's

---

[2] For convenience, I am using gender neutral plural "their" to refer to a gender-neutral singular male/female.

sports. Recently, the U.S. House Judiciary Committee held a hearing on the Equality Act. This bill forbids discrimination based on sexual orientation and gender identity, similar to existing rights on the basis of religion or race (Steinmetz, 2019). The implications of this bill are that, in the near future, "biological men" may compete against "biological women" in women's sports divisions. Note that the terms "biological men" and "biological women" were used in lieu of the words *gender* or *sex* when arguing the issue. Republican Rep. Doug Collins argued that, "Allowing men to compete against women in women's sports… is demoralizing" (Steinmetz, 2019). There is a general feeling that if biological men are allowed to compete in biological women's sports, there will, in effect, be no more biological "women's sports."

As can be gleaned from the incidents and political discussions described above, it is highly likely that it will be ultimately left to the U.S. Supreme Court to resolve the confusion regarding the contested meanings of *gender* and *sex* and other related terms in statutory interpretation.

Consequently, the research questions for this study are:

1. Are there differences in meaning and usage between the words *gender* and *sex* in modern society, and, if so,

2. How are those differences defined?

This research has been performed under an inductive approach. I examined the patterns that arose from the data, with the intent of having an unbiased perspective. In particular, this research explores the meaning of *gender* vs. *sex* regarding the two pending Supreme Court cases concerning *R.G. & G.R. Harris Funeral Homes v. EEOC* and *Altitude Express v. Zarda*. It should be noted, however, that this research does not aim to take sides, or to argue about intended meanings of *gender* and *sex*. Instead, this research aims to merely present the data

available and comment on the meaning that arises from that data. This research holds interest for those in forensic linguistics, sociolinguistics, anthropology, and semantics. Additionally, exploring the meaning of the words *gender* and *sex* can be educational for lawmakers and organizations.

In this review of related literature, I will first discuss current U.S. Supreme Court cases. This will be followed by literature on the corpus as a tool for contested meaning, and past studies that have used the corpus in legal interpretation. Finally, I will review research that discusses the use of participant data to increase validity and reliability, and the use of questionnaires in semantic fieldwork.

2.1 Current Supreme Court Cases

There currently exist two highly relevant cases concerning the meanings of *gender* and *sex* pending in the U.S. Supreme Court. The first case (*R.G. & G.R. Harris Funeral Homes v. EEOC*) questions if the word *sex* can encompass "gender identity," while the second (*Altitude Express v. Zarda*) asks if the meaning of *sex* includes "sexual orientation." The case *R.G. & G.R. Harris Funeral Homes v. EEOC* involves an employee of Harris Funeral Homes, Aimee Stephens, who was originally employed as a man in 2007, but after six years, self-identified as a woman (Howe, 2019). Subsequently, Stephens wished to come to work in women's clothing, but was terminated after this request. The Supreme Court agreed to examine the case, based on the 1989 Supreme Court decision on the *Price Waterhouse v. Hopkins* case. That case stipulates that companies cannot discriminate based on a stereotype of what a man or woman should look like.

The second case, *Altitude Express v. Zarda*, involves Donald Zarda who went to federal court contending that he was fired from a New York skydiving company, Altitude Express, on the basis that he was gay (Howe, 2019). Zarda claimed a violation of Title VII of the Civil Rights Act of 1964. Initially, Zarda's claim was denied. However, the full U.S. Court of Appeals for the 2nd Circuit revisited the case, claiming that Title VII includes discrimination based on sexual orientation, as a subset of sex discrimination. It is apparent from these examples that further

interpretation of the meaning of the words *gender* and *sex* would be very helpful for social contexts as well as for court rulings.

2.2 Corpus for Contested Meaning

As can be gleaned from the incidents, legal cases, and political discussions described above, it is highly likely that it will be ultimately left to the U.S. Supreme Court to resolve the confusion regarding the contested meanings of *gender* and *sex* and other related terms in statutory interpretation. Previous literature regarding the discovery of contested meanings that have informed various courts provide excellent research methods and models (Lee and Mouritsen, 2017; Phillips and White, 2018; Cunningham and Egbert, 2019). Central to these research models is the incorporation of multiple corpora such as the Corpus of Contemporary American English[3] and the BYU Law Corpora.

The many corpora created by Mark Davies are valuable interfaces created specifically for linguistic research[4]. I chose a select few of his corpora for my research in particular because they provide many tools to aide users as they sift through language data. These tools include, but are by no means limited to: incorporation of part of speech (POS) taggers, a collection of collocates and key words in context (KWIC), charts to give comparison over lengths of time, and the ability to collect an accurate sampling of data for a word or phrase when there is a large amount of data. –A KWIC index like the one provided on COCA allows users to search for a target term and then to see that target term surrounded by context. The corpora created by Davies have many advantages in addition to these tools. For example, one advantage is speed. Because the

---

[3] This corpus, along with the other mentioned corpora (created by Mark Davies), can be found at https://www.english-corpora.org/coca/
[4] These include, but are not limited to, the Corpus of Contemporary American English, the Corpus of Historical American English, the iWeb Corpus, the TIME Corpus, and the Corpus of Global Web-based English.

architecture of his corpora indexes data into tables, it then takes a matter of seconds to retrieve data. An example given by Davies (2009) explains that it takes "just 2.1 seconds [to find] the 168,000 tokens of *white*" in the over [1 billion-word] Corpus of Contemporary American English. A further advantage in association with speed is that any number of additional features can be added to the search with virtually no slow in retrieval time. These features include word forms, Part of Speech (POS), lemma, synonyms, etc. Users can also view sections of data, such as genre or year. This enables the researcher to see usage patterns more clearly. Davies expounds on the use of his corpora—the Corpus of Contemporary American English in particular—as an accurate tool to monitor and track language changes (2010).

The Corpus of Contemporary American English (henceforth referred to as COCA), is a corpus currently containing more than 1 billion words[5], and stretching from the year 1990 to the year 2019. In his article *The Corpus of Contemporary American English as the first reliable monitor corpus of English*, Davies explains the benefits of a "monitor corpus" (such as COCA) as opposed to a "static corpus" (2010). A "monitor corpus" is a title that refers to a corpus that is updated regularly. A "static corpus" refers to one that has not been updated since its creation, such as the Brown Corpus. This stagnant state of the Brown Corpus could be considered a disadvantage because there is no new data to examine. An additional drawback to the Brown Corpus, when compared to the COCA, is its size. COCA, as mentioned earlier, is an over 1-billion-word corpus while the Brown Corpus has only 1 million words. The large collection of data that COCA enjoys means that this corpus gives a much more accurate representation of language usage and change, because there is more language to examine. Additionally, while static corpora maintain value, a monitor corpus, as previously mentioned, further allows for

---

[5] As of May 2020

7

research on language change. As explained by Davies, a corpus such as COCA could answer questions such as if and when the words *gender* and *sex* have either increased or decreased in usage (respectively); and how these words have changed in meaning over recent decades. In supplement to COCA, Davies has also created the Corpus of Historical American English (or COHA). This corpus, while static, gives researchers access to data spanning back to 1810, which when coupled with the monitor corpus COCA gives even greater insight to language change. Further, as mentioned previously, COCA and COHA can also reveal key words in context (or KWIC) as well as collocates of the words *gender* and *sex*.

Goldfarb (2017) further gives a good overview of the helpfulness of corpora (corpus linguistics) in defining meaning in legal cases, as opposed to the past reliance on dictionaries as a tool. In particular, Goldfarb makes the insightful argument that words are not defined in isolation, but by the context of surrounding words in the sentence. One fallacy of dictionaries is that words are defined in isolation, with several given meanings for each word. However, when words are examined collectively (such as in concordance lines in a corpus), the meaning "depends largely on what the rest of the sentence says" (2017). Therefore, Goldfarb suggests that words may not be the basic units of meaning, but rather that phrases or expressions make up the basic units of meaning. In addition to words being intended as basic units of meaning in dictionaries, dictionaries also prescribe usage to words; the senses listed are seen as the only senses of the word, and those senses are clearly defined. However, corpus data is drawn from real language use in context, and therefore gives a reliable representation (depending on strategic

corpora choice and refrainment from "cherry-picking" data[6]) of the many definitions and usages of a word within a phrase or expression (See Goldfarb 2017: 4-5).

Ironically enough, the work performed in the past by lexicographers (and sometimes even the work performed by lexicographers today) certainly involved a form of "cherry-picking.[7]" Lexicographers collected examples of language usage which were compiled into "citation files," and then used to compose the senses of words for dictionary entries. These "citation files" were collected by hand and included entries from magazines, books, newspapers, etc. However, because unusual words or meanings hold linguistic markedness, these usages naturally would hold more attraction for lexicographers, and the ordinary meanings would become underrepresented (Goldfarb 2017: 18-19).

On the other hand, Goldfarb (2017: 19-20) argues that computerized corpora offer a few key advantages in comparison to this subjective assortment of "citation files." First, corpora are made up of a variety of text sources, sometimes including spoken speech; for example, COCA's sources include magazines and newspapers, but also academic texts, fiction and non-fiction novels, and televised news programs. Additionally, because the total size of the corpus is known, it is possible to enumerate frequencies. Secondly, the presentation of the data in computerized corpora is much easier to sift through and analyze than the small slips of paper in which "citation files" were recorded. The user face of corpora such as COCA, COHA, or iWeb, for example, allow users to easily discover patterns of usage. A quick glance down the page at the concordance lines reveals repeated meanings, collocates, and syntax.

---

[6] This term refers to choosing to only present data that will make ones' argument stronger, while excluding data that will weaken ones' argument.

[7] However, this is not true of every lexicographer today. Many contemporary dictionary lexicographers have begun to use corpus data to inform their definitions and usage patterns. For example, see the statement from the Oxford Languages group: https://languages.oup.com/about-us/how-we-create-language-content/

2.3 Studies Using Corpora for Legal Interpretation

In accordance with Goldfarb's insights, there have been several previous studies regarding interpretation of legal meaning using corpora. Key among these are the article by Mouritsen (2010)[8] exploring the definition of *carry* in association with the *Muscarello v. United States* case; as well as the recent delving into the emolument clauses (Cunningham and Egbert, 2019; Phillips and White, 2018). The former article emphasizes the importance of using corpus-based research to define statutory terms; in contrast to the common practice of using various dictionaries or contemporary and historical sources. This article was among the first of its kind to suggest using corpus-based research for statutory interpretation because of the accurate representation of a general population's language usage, as well as the lack of native speaker intuition. The later research conducted by Phillips and White (2018), and Cunningham and Egbert (2019) outline models for resolving contested meaning; providing excellent research methods and models to act as a guide. These research models will be reviewed below.

Mouritsen (2010) discusses the weaknesses of dependence on dictionaries in the court room, and subsequently suggests the use of corpus data as a reliable alternative. His article thoroughly examines and explains the deficiencies of the dictionary in the context of the Supreme Court case *Muscarello v. United States*. In the Muscarello case, Frank Muscarello went to a set location to sell marijuana. He had a handgun locked in the glove compartment of his truck, but the handgun remained in his glove box until the time of his arrest. The issue in the Muscarello case was the interpretation of the phrase "carries a firearm," and whether this term was intended by Congress to include the idea of conveyance in a vehicle. By extending the

---

[8] Which, incidentally, is referenced in Goldfarb's article, *A Lawyer's Introduction to Meaning in the Framework of Corpus Linguistics*.

meaning to include conveyance in a vehicle, Muscarello's likely ten-to-sixteen-month sentence would be extended to a mandatory five years, plus an additional six-to-twelve months for his underlying drug offense.

To resolve the issue, Congress and the Court utilized increasingly desperate examples (as described by Mouritsen, 2010) to prove their individual arguments, including relying heavily upon various dictionaries; and subsequently only presented examples that supported their individual cases (or in other words, they "cherry picked" their data). Congress intended to interpret the conveyance meaning, and upon realizing this, the Court then presented "the King James Bible, Robinson Crusoe, and Moby Dick, as well as two electronic newspaper databases and several unabridged dictionaries" (Mouritsen, 2010). In retaliation, Congress relied upon resources such as "a legal dictionary, … translations of the Bible, …a quotation from the television show *M*A*S*H*, and one of the Supreme Court's very few references to *Sesame Street*." However, none of these resources give an accurate representation of language usage in contemporary day-to-day life.

In fact, Mouritsen states that dictionaries should not be the source on what is correct in language (an opinion shared by Goldfarb), since most contemporary dictionaries are not concerned about prescribing standards of usage. Further, the definitions of terms are not necessarily common. A regular theme in the Muscarello case was that the "first, or primary meaning" of a word listed in a dictionary indicated that it was the most correct, or common, definition, when in fact this is not true. For example, in the first edition of the 1987 Random House dictionary, editors claimed that "the most frequently encountered meaning appears as the first definition." However, the word *generally* was added to this claim at the publication of the second edition. This is because the editors recognized that their rankings did not always

11

accurately reflect the frequencies of uses in corpora. Mouritsen (2010: 22) showcases this fault by examining the word *deal*. Mouritsen claims that the most used definition of the word *deal* is used to describe a particular amount in a transaction, i.e., *a great deal*. Although Random House claims to rank its definitions by statistical frequency, this specific definition is listed as twenty-first. The reason that this ranking is listed so dramatically low on this list is that human intuition about the frequency of lexical items is often unreliable. This fallacy is expertly expressed by McEnery and Wilson (2003: 12-14); stating that information about lexical frequency "is not susceptible to recovery via introspection."

To resolve cases of contested meaning, Mouritsen suggests turning to computerized corpora. Mouritsen's research here is invaluable, because it is compiled of quantitative, empirical evidence, made valid by the fact that it can be easily replicated. As Mouritsen states, the law is meant to reflect the common usage of the people it regulates; therefore Mouritsen chose to use COCA which gives modern usage of real language in context—in accordance to the recency of the Muscarello case (2010). Mouritsen's methods included searching COCA for all usages of the word *carry* in verbal form, such as *carry*, *carries*, *carrying*, and *carried*[9]. Additionally, because the case did not merely concern the meaning of the word *carry* as a basic unit of meaning, but instead the meaning of the word *carry* in the context of the phrasal unit of meaning *carry a firearm*, Mouritsen examined meaning and frequency of the word *carry* in association with the word *firearm*. Next, COCA allows for random sampling of the KWIC, as previously defined. Mouritsen generated 500 randomized concordance lines that showed several different meanings of the word *carry*. Finally, Mouritsen looked at collocations of the words *carry* and *firearm*. Collocations are valuable because instead of using a random normalized sampling of the

---

[9] This was done by inputting the following query into COCA: [carry]. [v*]

instances of usage, the collocation data looks at all instances of usage, and then gives a read-out of the statistics. From these methods, Mouritsen concluded that the ordinary meaning of *carry* in association with the Muscarello case was undoubtedly that of "carrying upon one's person".

As effectively demonstrated by Mouritsen (2010, 2019), corpus linguistics is a highly useful tool for legal interpretation. It may be especially needed in some instances in association with what is known as the "Plain Meaning Rule." This is a rule that judges often turn to; a method that requires the court to consider only the text of a contract and exclude all outside material unless that contract is ambiguous. However, if the contract is found to be ambiguous, then the court may utilize outside sources (a dictionary, for example) to define meaning (Merriam-Webster's Law Dictionary, 2020). This reliance on dictionaries presents several problems, some of which were discussed above. Although another issue is that the court may decide to ignore a "well attested alternative sense of the word it was purporting to define" if this sense does not support their argument. Mouritsen (2019: 4) presents the example of an ecological tourism company that did not defend a negligence action which came from the drowning of a customer. This death happened while snorkeling. Since the insurance contract stated that it "does not apply to 'bodily injury' [including death] to any person while practicing for or participating in any sports or athletic contest…" the question then became: is snorkeling a sport? The court proceeded to give examples of dictionary definitions which defined a *sport* as a rule based athletic competition. This meant that snorkeling, therefore, was not within the plain meaning of the word *sport*. This methodology, of course, was problematic because in fact the same dictionaries that the court used to define plain meaning of the word *sport* also defined it as a

13

"recreational activity" with no reference to a competition. Thus, here is another scenario in which corpus linguistics could[10] and should be used for legal interpretation.

A later corpus-based study performed by Phillips and White (2018) followed Mouritsen's example of his analysis of the *Muscarello v. United States* case; to discover possible meanings of the word *emolument*. Research regarding the meaning of *emolument* stems from the federal litigation against President Trump. Phillips and Whites' research methods included comparing data across multiple corpora and interpreting the findings. In this litigation, it was argued that there exist two meanings for the word *emolument* "—(1) a broad, general sense that covers any profit, benefit, advantage, or gain one obtains, whether tangible or not, from any source; and, (2) the legally-authorized compensation or monetizable benefits from public office, employment, or service" (Phillips and White, 2018: 5). Using this information, Phillips and White randomly sampled "hits" from each corpus, and then performed a holistic and subjective latent analysis of the usage. This analysis examined usage in context, with about 150-200 words surrounding the target word. The target words were then interpreted individually, based on the two previously presented senses of *emolument*. As part of the interpreting process, the authors discussed their results and reached a 70% agreement rate. This mutual agreement process was undertaken to determine intended meaning when it was not always clear from the corpora's key words in context analysis.

While helping to forge the way for future research in corpus-based analysis of statutory terms, there exist a few weaknesses to the study by Phillips and White. One evident weakness of this study is that it was performed with preconceived notions of possible definitions of the word

---

[10] There are, of course, some limitations to using corpus linguistics in the court room as highlighted in this example by the Plain Meaning Rule stipulations.

emoluments. Cunningham and Egbert (2019) actually address this in their later study, asserting that to "begin with the unquestioned assumption that there were two distinct meanings for emolument in 18th century America, then to frame the analysis narrowly" creates a validity problem, and that "this 'narrow meaning/broad meaning' assumption has no scientific basis." Indeed, Phillips and White divided each usage of *emoluments* into one of these categories, rather than allow the data to present the meaning. Instead of this preconceived notion, it seems prudent to analyze each concordance line with regards to what the syntax displays. A further weakness lies in the validity and reliability of the inter-rater interpretation of the KWIC. Phillips and White interpreted their data, but only the two researchers participated in the interpreting process thus leading to possible subjectivity and bias in their interpreting. While it is true that Phillips and White achieved a 70% agreement rate, this percentage is based on just two participants; and participants that are trained in semantic analysis. Conducting a survey that includes more participants combined with items that can achieve an agreement rate of 90 or 95% (based on Neyman and Fisher's confidence intervals[11]) would give more confidence to the research. This is because this high level of inter-rater agreement would effectively demonstrate a confident meaning of the word. On the other hand, a lower agreement rate suggests that the word is difficult to define—that it may have other possible meanings, or the meaning may change based on demographics. That being said, Phillips and White's research method does provide valuable information in that it examines the semantics of the target word based upon standard corpus methodologies when applied to legal research.

---

[11] Neyman and Fisher created the confidence intervals in the 1930's. The confidence intervals show what percentage a population can be expected to agree upon presented results. For example, a confidence level of 95% would mean that 95% of a population can be expected to agree with presented results (Lehmann, 1994).

15

Cunningham and Egbert (2019) conducted a similar study concerning the *emoluments* clauses, although this study presents itself as a "scientific" analysis. This claim is in response to the article written by Phillips and White and refers to Cunningham and Egbert's approach to their data. In contrast to Phillips and White's methodology—wherein they approached the target word *emoluments* with two preconceived meanings—Cunningham and Egbert began their research with no prior ideas about the target word. Their "scientific" method included examining the parts of speech directly surrounding the target word—specifically words that modified or described the target. The three analyses they conducted included examining: 1. preceding adjectives or prepositional phrases, 2. coordinated lists, and 3. modification as an object. Cunningham and Egbert allowed their study to reflect existing meaning, and for this reason their data could be determined to be more reliable.

While the above mentioned advocates for the relatively new practice of corpus linguistics in association with statutory interpretation (Lee and Mouritsen, 2017; Goldfarb, 2017; Cunningham and Egbert, 2019; Phillips and White, 2018) argue a good case in opposition to drawing language data from dictionaries, this line of research must be approached with the proper mindset and reasonable precautions. It is far too easy to participate in "cherry-picking" corpus data especially during the data interpreting process; leading one to find exactly what one wishes in order to prove their point. Instead, data must be examined in its entirety (with a reasonable sample of normalized hits), with researchers simply reporting their findings. This can be done through analyzing the syntax surrounding the target word; along with collocates, KWIC, and parts of speech, to name a few methods. Cunningham and Egbert (2019) performed their "scientific" study well in this respect as they approached their study with no preconceived

notions of the meanings of the word *emoluments*. Instead, their findings arose from careful examination of the data present, making their research methods an excellent model.

2.4 Participant Data to Increase Validity and Reliability

However, in an effort to add more validity and reliability to the interpretation of corpus data, I suggest that multiple participants should be asked to discern meaning, and interpret the data, apart from the researchers. Likewise, along with the study conducted by Phillips and White, Cunningham and Egbert were the only two participants that interpreted their data. This presents a drawback to their study. Further, although there is more research that has used corpora to examine meaning (Lee and Mouritsen 2017), none of the previous research—including Phillips and White's nor Cunningham and Egbert's—has utilized a large group of participants to aide in interpreting the data.

Accordingly, I hypothesize that using a larger number of participants to undertake the interpreting of KWIC terms derived from following a standard corpus linguistic research framework, and then comparing each participant's interpreted results to determine inter-rater reliability for each item would yield more objective results. Indeed, Lee and Mouritsen (2018) allude to this approach in their suggestions for future research when they assess corpus linguistic results in judging the contested meaning of statutory phrases such as "carry a weapon," and "vehicle in the park." For example, they suggest that one "construct a survey aimed at assessing not just the first sort of vehicle that comes to mind but also the range of meanings encompassed within a prohibition on vehicles in the park. Survey data could give us quantitative information about these notions of ordinary meaning" (2017). Building on this suggestion, as stated in the introduction, I propose to use a group of approximately 40 initial participants, followed by a

second group of approximately 100 participants, to interpret relevant data derived from a corpus linguistic study of the contested terms *sex* and *gender*.

2.5 The Use of Questionnaires in Semantic Fieldwork

The use of questionnaires or surveys to perform semantic fieldwork is a relatively new approach. Typically, theoretical syntax and semantics research has been performed by just the author of a paper, or else with some additional feedback from colleagues (Gibson, et. al 2011). However, a non-quantitative study that relies on the intuition of one or a few people is prone to major biases and creates room for subjectivity. Administering a semantic survey to a group of participants greatly lessens that subjectivity. VanderKlok (2014) advocates for semantic surveys, because, while the semantics of some units of language are clear, others are less understood; and a semantic survey can provide more information. This is because the survey should solicit judgements from a wider, diverse set of speakers. This method also provides a way to collect data quantitatively. Accordingly, VanderKlok (2014) introduces a few methods for creating questionnaires: elicitation in an acceptability/judgement task; a semi-forced task, and a rating task based on a Likert scale. I will now review VanderKlok's semi-forced and rating-scale tasks, as these questionnaires are most applicable to my research.

VanderKlok conducted a semi-forced choice task to ascertain acceptable uses of modality in the Javanese language. A questionnaire was administered to 15 participants between the ages of 19-51 years old. Before the participants took the survey, they were given instructions along with four practice examples. The participants were assured that there were no right or wrong answers. The survey was given over a PowerPoint Presentation using the "record narration" function. This was to record how much time a participant spent on each item; however, there were no limits to time. In each task, the participants were given two target sentences ((a) or (b))

18

per examined context. Participants were then asked to choose the target sentence that was the most appropriate for the given context. The options given to participants were (a), (b), both (a) and (b), neither, or to give an alternate answer. The participants were asked to make judgements on a total of 41 items. This seems to be a reasonable number of items—keeping in mind that a possible drawback of questionnaires can be respondent fatigue. When participants experience fatigue over the course of a long questionnaire, they may become less invested in their answers (Dahl, 1985). Nevertheless, VanderKlok did not experience issues with fatigue; claiming that had there been fatigue, surely more selections of the neither option would be apparent towards the end of the survey.

I find this semi-forced task to be conducive to my research for a few different reasons. First, this task can be applied to a large group of participants[12]. Second, unlike a Likert scale rating task, the semi-forced task can allow participants to offer an explanation other than the options provided. This is a valuable option in support of validity because this option steers the research away from bias based on the author's intuitions or social constructs. Another benefit of the way in which VanderKlok ran her study is that the instructions and practice questions given to the group provides inter-rater reliability. This brief overview before the questionnaire allows for explicit explanation of each of the given options. Without this instruction, participants are likely to interpret the researcher's intentions behind the options according to their individual intuitions. In addition, like the two other methods VanderKlok provides, this semi-forced task gives quantitative evidence in support of the research questions.

The rating task conducted by VanderKlok had two versions of the questionnaire and was conducted with a larger set of participants. Version A was administered to a group of 10

---

[12] Although VanderKlok utilized only 15 participants in her questionnaire, 40 does not seem that great of a number.

19

participants, and Version B to a second group of 10 participants. The ages of participants ranged from 17-50 years old. Just as in the semi-forced choice task, VanderKlok presented instructions along with four practice examples, although this step may not be necessary with a rating task. This is because Likert type scales are straightforward[13]; in VanderKlok's case with a 1 being a completely appropriate use of language, and 5 being a completely inappropriate use of language. The participants were asked to examine examples of language in context, and then to use the scale to rate the sample on whether it was an appropriate use of the language or not. Again, it was reiterated that there were no right or wrong answers. This survey was also conducted using a PowerPoint Presentation. However, as explained below, this present research uses Amazon's Mechanical Turk.

Mechanical Turk, as demonstrated by Gibson, et. al., is a viable option for linguists to collect behavioral data, as opposed to the "armchair linguistics" methods of the past (2011). This is because in recent decades it has been very time consuming, and sometimes expensive, to collect semantic data from a large pool of diverse participants. However, Mechanical Turk allows for many linguistics surveys to be completed within a day for a reasonable fee. The use of this program provides an easy fix to previous issues of semantics research. As mentioned earlier, research in theoretical semantics often involved the intuition of just the author of a paper. These results could not be generalized for several reasons: because the number of participants (oftentimes one) was too low; there often were not many items being tested for meaning; and there existed cognitive biases in the researcher and participant(s). Gibson et. al advocate for future research to adopt the standards that are common in all behavioral sciences which include:

---

[13] Although, there are always differences in personal opinion of the severity level of, for instance, a rating of 1 versus a rating of 2, instruction beforehand would not account for these differences.

(a) many experimental participants, none of which have knowledge of the research questions or hypotheses, (b) many experimental materials, (c) distractor materials, and (d) present the items in a different random order to each participant. These standards were difficult to achieve using methods of the past but are made possible by using the program Mechanical Turk. Combining the use of questionnaires to explore the semantic meanings of the terms *gender* and *sex* with the use of corpus data (being discriminatory with the selection of the chosen corpora) can present "scientific" evidence that is less biased than one or a few researchers interpreting corpus data.

3 METHODOLOGY AND RESULTS

As previously discussed, Phillips and White (2018) and Cunningham and Egbert (2019) provide excellent models for future research that can be expounded upon with further improvements and experimenting. I modeled much of my research based upon their methods. Both articles examined the intended meaning of the word *emoluments*. Methods included accessing every usage of the word in question using three different corpora: the Corpus of Founding Era American English (COFEA), Founders Online, and Hein[14]. These three corpora were sufficient for their research, because they examine American founding era English. Since

---

[14] COFEA was created by the J. Reuben Law School at Brigham Young University and was launched in 2018. COFEA totals more than 138,800,000 words combined from texts written between 1760 and 1799. COFEA is available at https://lawncl.byu.edu/. The texts in COFEA include the texts from Founders Online and HeinOnline. Founders Online contains copies of over 90,000 records from George Washington, Benjamin Franklin, John Adams, Thomas Jefferson, Andrew Hamilton, and James Madison. Founders Online is available at https://founders.archives.gov/. HeinOnline contains over 300 legal materials published during the founding era. HeinOnline is available at https://heinonline.org/.

my research concerns more modern language, I selected my three corpora accordingly. Following the collection of data, both sets of researchers (Phillips and White; Cunningham and Egbert) categorized their data into instances when the word emolument was pluralized; when it was modified by an adjective, when it was included in a coordinated list, and finally when it was modified as an object of verbs. For the purposes of my research, I examined similar constructions. My research questions can be answered using empirical data, and the analysis of the data has been spread between both qualitative and quantitative analysis—both accomplished through corpus study, and subsequent semantic survey administration.

3.1 Initial Corpora Analysis

To secure reliable language data, corpora must be chosen with care and with thought towards answering the specific research question. As shown above with reference to Mark Davie's article (2010) on using COHA (because of its usefulness as a "monitor corpus"), not all corpora are created equal or with the same purposes in mind. To examine the differences between the words *gender* and *sex*, I have deliberately chosen three different corpora through which I analyzed modern language data: COCA, iWeb, and COHA[15]. COCA, as discussed previously, can provide data on the meaning and usage of the words *gender* and *sex* over the last 30 or so years. Therefore, this data reflects recent language change and shifts in meaning. I will now give brief explanation of the other two corpora that I utilized in my research.

---

[15] In some cases of statutory interpretation, lawyers give specific definitions of terms within their documents, and these terms can only be defined within the confines of those interpretations. Indeed, one argument against the use of pre-established corpora could be that they do not apply within particular contexts. However, the use of a corpus based on set documents could easily be created using an application such as AntConc or Word Cruncher.

iWeb is a recently published corpus[16] that, as the name suggests, draws its 14-billion-word collection from sources across the web, using 22 million web pages. The overview on the iWeb corpus page explains that the websites (e.g. blogs, newspapers, educational resources) included were chosen systematically, allowing a user to create a virtual corpus for any topic (therefore allowing for a virtual corpus focusing on the topics of *gender* and *sex*). This corpus provides a massive amount of data to give evidence of how *gender* and *sex* are used in contemporary English.

COHA, is "the largest structured corpus of historical English" with more than 400 million words drawn from texts from the 1810s-2000s (Davies, 2010-). Thus, COHA provides insights on how *gender* and *sex* were used historically in the United States, displaying contrasts in meaning to the contemporary usages, as also collected from COCA and iWeb. COHA was used by both Phillips and White (2018); and Cunningham and Egbert (2019) in their research models, and while collecting data from this corpus, I followed their examples. I used four main categories in my searches on the various corpora: determiners before the target words, adjectives before the target words, the target words as adjectives, and collocations of the words *gender,* and *sex,*. After experimenting with my searches, I chose these categories because they demonstrated the widest variety of usage of the target words. This careful examination of the concordance lines allowed for three specific categories of usage to arise from the data: physical anatomy, individual identity, and an action (*sex*).

Following my initial observations, I also planned for a qualitative analysis to be performed by approximately 40 linguistics students in both the undergraduate and graduate

---

[16] Another creation of Mark Davie's released in May 2018.

programs at BYU, in the form of a semantic survey. I modeled this semantic survey after the case study on the use of questionnaires in semantic fieldwork done by VanderKlok (2014).

The items for the semantic survey were chosen randomly from the corpora concordance lines. Using the four categories of items that I used to search the corpora, I gathered between 1-3 examples of each category from each of the three corpora. There is a helpful feature on the corpora created by Mark Davies that allows the user to generate a random sampling of the results. I therefore generated a random sampling of 100 of the results, and then took one of the concordance lines from this sampling to use in the surveys. As I gathered items for the first survey, I wanted to ensure that the items came from a variety of years and sources. Therefore, if, for example, I obtained an item from iWeb that was from a blog in 2010, then on the next random sampling of 100, I chose an item that came from, for example, a newspaper in 1998.

3.2 Survey Construction

The "interpreters" were presented with a semi-forced task composed of the items from the corpora concordance lines. This task asked them to decide the speaker's intentional meaning of either the word *gender* or *sex* within each context. Additionally, there were two separate sets of data to present to participants for interpreting. Each participant was given only one of these sets; and therefore, participants did not all have the same data. To increase validity, using two versions of this survey as implemented by VanderKlok allowed for a margin of error among the survey items. For instance, if one or more items on Version A were unreliable, then a Version B would provide more data in place of the less reliable survey items.

Before participants were presented with the questionnaires, they received a brief training. The training message included instructions on how to complete the questionnaire. For example, I explained the options for "interpreting," complete with an example of the intended meaning for

each option. I then performed an example item to further clarify the process. For each task, participants were given the three options: physical anatomy, individual identity, and an action; as well as the two additional options: unclear and other. The "other" option gives participants the opportunity to present a different answer in the case that they find the given options inapplicable. Further, these options were be presented with the instruction: "if more than one option is applicable, please choose the option that best represents intended meaning." This training ensured that each participant had an equal understanding of the task, thus reducing ambiguity in the completion of the task.

In addition to this survey given to linguistics students—and in an effort to increase both validity and reduce bias with a greater extent—another quantitative sociolinguistic analysis based on the results of the first was conducted among a large group of laypeople using Amazon's Mechanical Turk, an affordable program that gives the administrator the ability to select the demographics of the survey participants and obtain very quick results (Gibson, et. al. 2011). The rationality behind using a rating task on Mechanical Turk is based on a few reasons. First, Mechanical Turk gives me the option to obtain data from an even greater number of participants, at minimal cost. Using this tool, I could easily solicit data from a few hundred participants. Further, because this survey was conducted on a much larger scale than the initial 40 participants, the data was much easier and less time-consuming to analyze using a rating task. Indeed, this is a pressing reason to use the rating task. However, this rating task was informed by the data collected from the first semi-forced choice survey, giving it greater validity and reliability. Based on the input from linguistics students with various training, this rating task would not be subject to the author's biases or intuition. Next, when using Mechanical Turk, it is impossible to administer instructions and practice tasks to the participants. In my opinion, a

25

rating task would not necessitate this training, and so the data obtained from Mechanical Turk would still be deemed reliable.

3.3 Methodology and Results: Survey 1

Linguistic Student Survey

Participants

There were forty-two participants total, recruited from either the linguistics graduate or undergraduate programs at Brigham Young University (BYU). The participants were 25 females and 17 males between the ages of 18 and 36. Although I initially asked for age and gender demographics, after analyzing the data this information (somewhat surprisingly) did not prove to be significant.

Initially, students were invited to participate in this survey through an email, with the promise of a $10 CougarCash[19] reward. The goal was to have 25 participants at the first administration of the survey; although because the survey was administered on a weekday evening, only 8 participants arrived. After this first attempt, the survey was administered to 34 students in Dr. Eggington's 201 introductory undergraduate linguistics course during the Winter 2020 semester in exchange for extra credit in the class.

Survey Design

The goal of the Linguistic Student Survey was to enlist the aid of individuals somewhat trained in semantic "interpreting" to help create a more valid and reliable survey to introduce on Mechanical Turk to a larger group of laypeople. The overall purpose of this project—and the two surveys—was to recruit participants to determine the meaning of the terms *sex* and *gender*

---

[19] This is the name of the money system used for student accounts at BYU. I received USD $80.00 from the BYU Linguistics Department for the purpose of administering the Linguistics Student Survey.

(rather than relying on the intuition of one researcher and a handful of colleagues); and the Linguistic Student Survey further helped to categorize meaning before seeking opinions from a larger group[20].

Forty-two Linguistic Student Surveys were administered in total. However, this survey had two versions; 20 copies of survey A and 22 copies of survey B were administered. The purpose of the two versions was to test more items from the corpora concordance lines. This, in turn, would allow for any questions that were unreliable to be eliminated, and additionally allow a greater chance for reliable questions to surface. Each version of the survey included 24 items, for a total of 48 items that were tested. As mentioned previously this survey was administered to two groups on two separate occasions; however, in each group participants were administered either a survey A or survey B option.

The categories included in these 24 items were as follows:

1. DETERMINER (*sex*) (*gender*)[21]
examples: that sex, each gender
2. ADJECTIVE (*sex*) (*gender*)[22]
examples: the fairer sex, traditional gender
3. (*sex*) (*gender*) ADJECTIVE[23]
examples: sex specific, gender neutral
4. COLLOCATES OF *sex,* and *gender,*
examples: sex, age; gender, race

As previously mentioned, these concordance lines were taken from three sources: COCA, iWeb, and COHA. Between 1-3 examples of each category were derived from each source to be

---

[20] This method was influenced by VanderKlok (2014).
[21] For an exact search on the corpora, you may use the search term: _d* sex or _d* gender
[22] For an exact search on the corpora, you may use the search term: _j* sex or _j* gender
[23] For an exact search on the corpora, you may use the search term: sex_j* or gender_j*

27

included in the surveys. One caveat is that I did not pull examples of (*sex*) (*gender*) ADJECTIVE from COHA, simply because there were hardly any usages of these terms in the historical American English concordance lines included on COHA.

The survey began with a cover page, which asked for the participant's age and gender[24], as well as their BYU ID number (this was purely for the purpose of depositing $10 into their account for compensation; although some participants received extra credit for having participated in the survey). Additionally, a caveat expressed to participants that age and gender were only elicited for the purpose of inter-rater reliability. Further included on this cover page were directions for taking the survey (see example 1). Finally, a sentence at the bottom of the page encouraged participants to ask any questions they might have before beginning the survey.

Ex. 1:

Directions:

For each item, you will read an example sentence that contains a target word (which will be <u>underlined</u>). Each item will be accompanied by five options that offer possible meaning of the target word. You must choose only one option; the option that BEST reflects the author's intended meaning of the target word.

Example:

Read the following sentence:

"Four-and-a-half-year-old nursery-school children played with same-<u>sex</u> playmates three times as much as they did with cross-<u>sex</u> playmates in mixed groups." -Omni (Science Magazine), 1990

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy
B. Individual identity
C. An action
D. Both A and B

---

[24] Although the findings presented in the results section of this paper make my use of gender instead of sex a bit ironic

E. The meaning is unclear
F. Other (please list) _____

Before administering the official Sex and Gender Semantic Surveys A and B, I compiled a short test version to distribute among colleagues to receive feedback. This version was composed of five items, which was examined by 4 colleagues in the BYU linguistics and TESOL graduate programs (2 professors and 2 students). Further, I presented this 5-question version to a senior linguistics class at BYU.

Based on the suggestions from these early participants, I added the following improvements to the official surveys: Instead of asking the participant to choose "the option that BEST reflects the meaning of the target word," participants were asked to choose "the option that BEST reflects the author's intended meaning of the target word." Next, an option was added to include "Both A & B" (physical anatomy and individual identity), giving the participant 6 answer options instead of 5. Finally, the year and source of each concordance line was added, based on the suggestion that the author's intended meaning could be dependent on that information. I feel that these improvements added to the validity of the surveys.

Further, I included the option of "an action" on the survey for a few reasons. First, "an action" is a plausible meaning of the term *sex* that I observed during my searches on the corpora. Second, items eliciting this interpretation acted as a distractor for the other items and could demonstrate the participant's investment in the survey task. For example, "an action" is not a plausible meaning of the term *gender*. However, if a participant were to choose the option "an action" to interpret *gender*, this could indicate that the participant's data was less reliable.

After collecting the data for survey one from two sessions of in-person assessment, I input the data onto two Microsoft Excel spreadsheets (one for each version of the survey). This

29

data analysis of the first survey helped in composing the second survey, which was administered on Amazon's Mechanical Turk program.

Results

In the question of the meaning of *sex* as physical anatomy or individual identity, *sex* was only ever classified by the majority as physical anatomy. There is one question on Survey A (question 9) where 55% of participants classified *sex* as both A & B (physical anatomy and individual identity). However, this is the only occurrence where the majority classified the term *sex* as both A & B. The question reads: "… federal laws prohibit harassment based on race, <u>sex</u>, color, national origin, religion, age, and genetic information." This is the item that concerns the federal law of equal opportunities. The reason behind the differing answers on this item seems to be purely political.

The items that achieved the highest agreement rates between participants were those that included determiners before the target words (N=8). Then, items with adjectives placed before the target words (N=4). One could therefore surmise that items which use determiners before the target words are simpler for raters to categorize and agree upon.

Another observation is that the majority of answers chosen on the surveys were option A (physical anatomy). On survey A, 43% of answers were option A, and on survey B it was 54%. The next most popular answer was option D (both A & B) with 24% on survey A and 19% on survey B. Options B (individual identity), C (an action), E (unclear), and F (other) were hardly used, in comparison. The statistics are shown below in Table 1:

Table 1: Percentage of Answer Choices, Surveys A and B

	A	B	C	D	E	F
Survey A	43%	11%	16%	24%	4%	1%
Survey B	54%	14%	9%	19%	2%	1%

I will further present the results of surveys A and B by using the categories as described above. When examined in this way, the results presented interesting patterns. In addition, for the purposes of this analysis (and any analyses following) the word *majority* is defined as the greatest number of participants that responded with the same answer choice for one item[25].

<u>Determiner placed before the target words:</u>

With items using the determiner *that* before *sex*, most raters chose the option an action. On the other hand, for items using the determiner phrase of *same sex*, the majority of raters chose physical anatomy.

Three out of the four items that used a determiner before gender (such as *that*, *each*, and *same*) were decided by the majority to mean physical anatomy. However, on one option that used the phrase *of gender*, 70% of raters chose A & B (both physical anatomy and individual identity). On the item in question, the topic was Gender Inclusive Housing. This is the kind of context could suggest individual identity. Yet it is interesting to note that on all items that placed a determiner before the target terms, none of the raters chose the option of individual identity.

<u>Adjectives placed before the target words:</u>

---

[25] The majority was anywhere between 36 and 100% on a given item.

Here again, most of the items that contained an adjective before the term *sex* had a majority classification of physical anatomy. This was true for the phrases *opposite sex* and *fair sex*. Yet, the adjectives *more* and *safe* denoted a majority meaning of an action.

When an adjective such as masculine or the quantifier one was placed before the term *gender*, most participants chose the option of physical anatomy. I believe this is because a binary context was given. Although, on the term *traditional gender roles*, the majority of participants chose the option both A & B (physical anatomy and individual identity). Perhaps this is owing to the use of the noun roles, which can suggest fluidity.

<u>The target words used as adjectives:</u>

When *sex* was used as an adjective in the concordance line, most participants categorized the meaning as physical anatomy.

There were two main categorizations for the use of *gender* as an adjective. Those were physical anatomy and individual identity; and this is the only category wherein *gender* was categorized by most participants to mean individual identity. The terms that referenced physical anatomy were *gender schools* and *gender neutral* (when referencing baby shower gifts). On the other hand, when the term gender neutral was used in context with terms such as *a-sexual* and *questioning*; as well as referencing names, then *gender neutral* was categorized by most participants to mean individual identity. Most participants also categorized the phrases *gender role* and *gender filter* as individual identity.

There was one question where *gender differences* was chosen to mean individual identity, and one where it was chosen to mean physical anatomy. The difference often lies in the context. The first item, wherein participants decided that gender meant individual identity, read, "Recent studies suggest that <u>gender</u> differences in trust and trustworthiness can be attributable to the ways

in which men and women find meaning for themselves in relationships..." The second item, categorized as physical anatomy, read, "There is also evidence that the experience of divorce…impedes the capacity of children…to form stable, lasting relationships. And there are significant <u>gender</u> differences: Boys tend to 'act out' their feelings of abandonment and resentment, while girls turn inward and become depressed." Perhaps because the first item concerns adults, it was categorized as individual identity. However, the second example concerns children, so perhaps this is why raters chose physical anatomy.

<u>Collocates of *sex,* and *gender,*</u>

When *sex,* was placed in a list of collocates (which included demographic information), most raters chose physical anatomy. This follows the previous patterns of the term *sex* to mean physical anatomy more than individual identity. Although, there is one exception to this rule that I will discuss shortly.

When *gender,* was placed in a list of collocates (also demographic information), two items were majority rated to reference physical anatomy while the other two items were majority rated to reference both A & B (physical anatomy and individual identity). All four items included shared collocates of age and race.

The items majority rated physical anatomy included the collocates of *sexuality* and *health status* respectively. The items that were rated A & B included the collocates of *marital status* and *sexual orientation* respectively. I cannot determine the difference between the terms *sexuality* and *sexual orientation*. I would have thought that raters would also chose A & B for the item that included the term *sexuality*.

<u>Items Chosen for the Mechanical Turk Survey</u>

33

There were 14 questions combined between the two surveys that achieved 80% agreement or higher between the raters[26]. I decided on the criteria of 80% agreement rate or higher because while I would have liked to use questions that had 90-95% agreement, 80% is still a high number and this criteria allowed for more questions to be included on the next survey to be conducted on Amazon's Mechanical Turk. An interesting yet not altogether surprising find was that participants agreed 100% on their selection of option C (an action) for 4 items. The reason this finding is not surprising is because the option "an action" denotes a verb. Additionally, even when *sex* is used in noun form, the nature of the context suggests "an action" (see example 2).

> Example 2:
> 14. "Yes, this one about AIDS is very popular. It speaks of the disease for which there is no cure and talks about the need to have only one wife and to practice safe <u>sex</u>." – National Geographic (Magazine), 2009
>
> 24. "There is probably no tribe in which formal marriage alone is sexually satisfactory. Perhaps it would be more correct to say that <u>sex</u> is not the most significant fact of marriage." -PBS General Anthropology (Academic journal), 1938

Included now are the other questions between the two surveys that achieved 80% agreement or higher between the raters:

Survey A:

8. "Although Lexi is the only girl on an all-male team, it seems that <u>gender</u> is the only thing that sets them apart." - Professional School Counseling (Magazine), 2010

Raters chose option A: physical anatomy

10. "At 6 years old, I knew I was attracted, drawn very heavily to other people my same <u>sex</u>." -Slate Magazine, 2017

---

[26] Based on Neyman and Fisher's confidence intervals that show what percentage a population can be expected to agree upon presented results. The confidence level of 80% means that 80% of a given population would be expected to agree on the results. (Lehmann, 1994).

Raters chose option A: physical anatomy

11. "If there's a truism about Hollywood, it is that <u>sex</u> sells." -New York Times (Newspaper), 2009

Raters chose option C: an action

21. "God's Word places great emphasis upon <u>sex</u> only within marriage; purity was to be maintained within marriage..." – cuttingedge.org (religious article), 2007

Raters chose option C: an action

22. "… this report was colored by the unwillingness of the women-reformers to admit that their own <u>sex</u> is physically weaker than the sex which votes..." –New York Times (Newspaper), 1875

Raters chose option A: physical anatomy

Survey B:

5. "Data were analyzed with Cox proportional hazard models adjusted for child age, child <u>sex</u>, birth order, parents' socioeconomic status…" -PBS NewsHour (Talk Show), 2016

Raters chose option A: physical anatomy

11. "Some studies in non-athletes have found cortisol levels to…differ between men and women… it [is] important to examine the unique cortisol responses…within each <u>gender</u>." -Journal of Sport Behavior (Academic journal), 2019

Raters chose option A: physical anatomy

12. "Please know that I HEAR YOU and I feel your pain, " [Holliday] wrote in her apology letter to the lesbian, bisexual, gay, transgender, <u>gender</u> neutral, asexual, questioning, non-binary, gender fluid crowd." -Fox News (Newspaper), 2017

Raters chose option B: individual identity

13. "Patients were matched with respect to age and <u>sex</u>, disease, and the kind of surgery that- the extent of the disease and the kind of surgery they required." -NPR (Talk Show), 1995

Raters chose option A: physical anatomy

18. "Peter Pan…was bound to become a musical in time-and doubtless in time for Mary Martin to play Peter. She looks as boyish as can be expected of any grownup of the opposite <u>sex</u>." -TIME (Magazine), 1954

Raters chose option A: physical anatomy

19. "Often male [dolphins] are related to 2 or 3 individuals of the same <u>gender</u> and create partnerships for cooperation purposes." -dolphins-world.com (Blog), 2017

Raters chose option A: physical anatomy

23. "I've seen clients who feel that if they have more <u>sex</u> with more women than any of their friends, they are somehow proving their masculinity," -Essence (Magazine), 2012

Raters chose option C: an action

As shown above, most of the items that achieved 80% or higher inter-rater reliability were answered with options A (physical anatomy) and C (an action). Only one question, question 12 on survey B, was answered with option B (individual identity). Prior analysis of results shows that most participants interpreted *sex* to mean physical anatomy over individual identity, whereas *gender* can mean either physical anatomy or individual identity based on grammar or context. In addition, only *sex* was interpreted as an action, where *gender* was never chosen to mean an action. Therefore, it makes sense that most items that achieved 80% or higher contain the term *sex*, because this term is more straightforward.

When analyzing the concordance lines included in these items, ten out of the fourteen use the term *sex*. That the raters were able to achieve high agreement rates on items that include the term *sex* is telling. I believe that one reason for this could be that the term *sex* has been in use much longer than the term *gender*. Corpus data from COHA shows that the term *sex* was the only term used for physical anatomy during the 1800s and up to the mid-1900s. From my research on COCA, the term *gender* did not begin to see much usage until the mid-1990s. On the other hand, of the three items that use the term *gender* the target word is used in a binary context.

I believe that this binary context (i.e. use of adjectives such as all-male) is helpful in interpreting the intended meaning of the target word.

Based on this data analysis, I took 14 questions that achieved 80% or higher inter-rater reliability on the first survey to use on the Mechanical Turk survey. Additionally, I chose 6 other questions to potentially give more variety to the Mechanical Turk survey. This is because, as mentioned before, most of the 14 questions elicited options A or C. Perhaps the reason for this agreement on items that elicited options A or C could be because these are more familiar contexts. –It may also be possible that the items that had a majority response of either B (individual identity) or D (both A and B) (although less than 80% agreement) are more of the unknown contexts that are important to examine and thereon gain further data. Further support for the reasoning behind including items that elicited option D is that this option was the second most chosen option, behind option A. As stated previously, 24% of answers on survey A were this option, and 19% on survey B.

The following 6 items were chosen to be additionally included on the Mechanical Turk survey, in order to give more representation to the term *gender* on the survey:

Survey A:

2. "A demographic questionnaire included age, gender, marital status, living situation, geographic region, race/ethnicity…" -Environmental Health Perspectives Journal, 2013

45% of raters chose option D: Physical anatomy and Individual identity

9. "… federal laws prohibit harassment based on race, sex, color, national origin, religion, age, and genetic information." – eeoc.gov (government website), 2020

55% of raters chose option D: Physical anatomy and Individual identity

12. "Authors have suggested that gender roles and experiences with racism and discrimination likely contribute to risk among non-dominant groups." - Academy of Entrepreneurship Journal, 2016

40% of raters chose option D.

13. "The Gender Inclusive Housing (GIH) option will provide a welcoming space for students to choose their roommates, regardless of <u>gender</u>, to promote a healthy and safe living and learning environment." – uis.edu (university page), 2020

70% of raters chose option D: Physical anatomy and Individual identity

Survey B:

6. "Secular progressives…have their own list of books they think young people should not read--for instance, books they claim are tinged with racism or jingoism or that depict traditional <u>gender</u> roles." -USA Today (Magazine), 2013

41% of raters chose option D: Physical anatomy and Individual identity

8. "Whether it be their race, <u>gender</u>, religion, ethnicity, disability, or sexual orientation, a hate group expresses prejudice against people with a particular identity." - Phsychology.org (Academic article), 2018

45% of raters chose option D: Physical anatomy and Individual identity

3.4 Methodology and Results: Survey 2

<u>Mechanical Turk Survey</u>

<u>Participants</u>

The initial plan was that there were one hundred participants total; although I did not realize how participants function on Mechanical Turk until after I conducted the survey. Therefore, there are likely more than one hundred workers who participated in this survey. After the survey responses were complete, I discovered that Mechanical Turk workers are paid by the assignment (or question) and not be the task (the entire survey). This means that although the survey had 20 questions total, some surveys were not completed by the same individual. Therefore, although I wanted responses from one hundred people in total, there were more than

one hundred participants because Mechanical Turk required one hundred workers per assignment and not per task. This is a small distinction that may have affected the data.

Another factor that may impact the data from this survey is that I did not collect any demographic data. There is no way of knowing the age of participants, their gender, where they are from, or their education level. However, I feel certain that the pool of participants was a diverse one. An article by Difallah et. al (2018) gives an overview of the demographics of the workers on Mechanical Turk, based on an (at the time) ongoing survey that was conducted for 28+ months on the said platform. According to Difallah et. al, "most of the workers are from the USA (75%), with India (16%) being second, followed by Canada (1.1%), Great Britain (0.7%), Philippines (0.35%), and Germany (0.27%)." Results further indicated that the workers were generally balanced between males and females. However, workers tend to be younger: "20% of the MTurk workers are born after 1990, 60% are born after 1980, and 80% are born after 1970." Perhaps the only drawback for my research based on this data is that my hypothesis is based on the hope that this data will help determine court rulings for issues that are ongoing in the United States—and a percentage of the participants of my Mechanical Turk survey may be from outside of the United States.

Participants were recruited on Mechanical Turk with a title: "Sex and Gender Semantic Survey;" a short description: "Read a sentence and identify the meaning of the TARGET word;" and some keywords: "sentiment, text, survey." They were promised USD $0.12 per assignment (Mechanical Turk's word for question), for a total of USD $2.40[28] if they completed the 20 assignments in the task (in this case, a survey).

---

[28] I received USD $350.00 from the BYU Linguistics Department for the purpose of administering the Mechanical Turk Survey.

Survey Design

The Mechanical Turk Survey was a 20-question online survey, derived from surveys A and B from the Linguistic Student Survey. Further, it was modeled after the Linguistic Student survey. The main difference between the Mechanical Turk Survey and the Linguistic Student Survey is that each participant on Mechanical Turk was given the same version of the survey; there was no survey B. This is because one of the purposes of the Linguistic Student Survey was to test many items to find a smaller group of items with high inter-rater reliability. The hope was that the items with high inter-rater reliability between trained linguistics students would lead to a more valid survey for laypeople to participate in.

This survey was very straightforward with no cover page, but with the directions listed in a panel to the left of the screen. The directions here were briefer:

> *Directions: You will read an example sentence that contains a TARGET word in BOLD. Each item will be accompanied by four options that offer possible meaning of the TARGET word. You must choose the option that best reflects the author's intended meaning of the TARGET word.*

The conciseness of the directions was so that workers could see the full directions on their screen without the necessity of scrolling down the page. There was an option for workers to click on a description for longer instructions. These longer instructions included the example prompt given on the cover page of the Linguistic Student Survey.

There were further differences in formatting for the Mechanical Turk Survey. Instead of underlining the target word, Mechanical Turk's design for the survey necessitated that the target word be in ALL CAPS. Additionally, the options for the worker to select were given on the right side of the item. To choose their answer, the worker clicked on their chosen option and then clicked on the submit button. This option of clicking the submit button after clicking on their

chosen answer allowed for workers to make sure they had chosen the correct option before moving on to the next item.

Based on the results of the Linguistic Student Survey, the answer options for the Mechanical Turk Survey were reduced from six to four. Instead of including options *E. The meaning is unclear* and *F. Other (please list)*, only the four previous, pre-determined options were kept. There were a few different reasons for this choice. First, options E and F were not used frequently on the Linguistic Student Survey. Additionally, these options were included on said survey to be a marker of items that should possibly be revised or deleted. Finally, option F would prove to be more difficult to include on Mechanical Turk (owing to the write-in option), as well as more time-consuming for the participants. Therefore, only four options were provided for participants.

Results

The results of each question in the Mechanical Turk Survey were much more spread than the results of the Linguistic Student Survey had been. Although, like in the Linguistic Student Survey, option A of physical anatomy was the most used option with 31% of answers being option A. However, this was closely followed by option B, individual identity, with 30%. This was somewhat unexpected based on the Linguistic Student Survey, where option B was chosen about half this much. See Table 2 for the percentage of answer choices.

Table 2: Percentage of Answer Choices, Mechanical Turk Survey

	A	B	C	D
Survey	31%	30%	22%	17%

41

Because option D includes both options A and B, this means that either option A or option B was chosen approximately 78% of the time.

Table 3 gives the survey answers in more detail: including each question with the target word and the most popular answer to each question. It is further organized by category (DETERMINER before *sex* or *gender*, etc.).

Table 3: Results from the Mechanical Turk Survey

Key:
DET: DETERMINER (*sex*) (*gender*)
ADJ B: ADJECTIVE (*sex*) (*gender*)
ADJ: (*sex*) (*gender*) ADJECTIVE
COL: COLLOCATES OF *sex,* and *gender,*

Category	Sentence	Answer
DET	There is probably no tribe in which formal marriage alone is sexually satisfactory. Perhaps it would be more correct to say that SEX is not the most significant fact of marriage	72% An Action
DET	God's Word places great emphasis upon SEX only within marriage	71% An Action
DET	If there's a truism about Hollywood, it is that SEX sells.	60% An Action
DET	colored by the unwillingness of the women-reformers to admit that their own sex is physically weaker than the SEX which votes	39% Physical Anatomy
DET	Some studies in non-athletes have found cortisol levels to…differ between men and women …it [is] important to examine the unique cortisol responses…within each GENDER	55% Physical Anatomy
DET	Although Lexi is the only girl on an all-male team, it seems that GENDER is the only thing that sets them apart.	44% Physical Anatomy
ADJ B	clients who feel that if they have more SEX with more women than any of their friends, they are somehow proving their masculinity	69% An Action
ADJ B	It…talks about the need to have only one wife and to practice safe SEX.	67% An Action
ADJ B	She looks as boyish as can be expected of any grownup of the opposite SEX.	45% Physical Anatomy
ADJ B	At 6 years old, I knew I was attracted, drawn very heavily to other people my same SEX	44% Physical Anatomy
ADJ B	books they claim are tinged with racism or jingoism or that depict traditional GENDER roles	57% Individual Identity
ADJ B	Often male [dolphins] are related to 2 or 3 individuals of the same GENDER	53% Physical Anatomy
ADJ	[Holliday] wrote in her apology letter to the lesbian, bisexual, gay, transgender, GENDER neutral	65% Individual Identity
ADJ	Authors have suggested that GENDER roles and experiences with racism and discrimination likely contribute to risk among non-dominant groups	53% Individual Identity
ADJ	The GENDER Inclusive Housing (GIH) option will provide a welcoming space for students to choose their roommates, regardless of GENDER,	39% Individual Identity
COL	Patients were matched with respect to age and SEX, disease,	62% Physical Anatomy
COL	Data were analyzed…adjusted for child age, child SEX, birth order	60% Physical Anatomy
COL	federal laws prohibit harassment based on race, SEX, color	33% Physical Anatomy*
COL	Whether it be their race, GENDER, religion, ethnicity, disability, or sexual orientation	55% Individual Identity
COL	A demographic questionnaire included age, GENDER, marital status,	46% Individual Identity

*Number 17 has interesting results; although 33% chose Physical Anatomy, this was closely followed by a 32% selection of Both A & B. Additionally, 27% chose option B, Individual Identity

From Table 3, one can see that there are 8 questions in total that achieved a 60%

agreement rate or higher between participants. The standard of agreement rate was lower in this

43

survey than it was in the Linguistic Student Survey, where a total of 14 questions achieved 80% agreement rate or higher between participants. This could be for a variety of reasons: First, the linguistics students were at least moderately trained in analyzing semantic data and categorizing meaning, because of the nature of their studies. Second, there were only 42 participants of the Linguistic Student Survey compared to the 100+ participants of the Mechanical Turk Survey. Third, there were 48 items total between both Linguistic Student Surveys, and only 20 items on the Mechanical Turk Survey. Finally, the participants of the Linguistic Student Survey may have been from a more conservative background since they attend a private religious university— whereas participants on Mechanical Turk are presumably of a more diverse background.

I will again analyze the results that stemmed from the categories that I used to search terms on the corpora. The results in this section mostly mirror the data in the same section of the Linguistic Student Survey results (adding to the validity of both surveys).

Determiner placed before the target words

The only change from the previous results is that instead of marking A&B (both physical anatomy and individual identity) for the determiner of *gender*, most participants chose individual identity. This is a pattern throughout the data.

Adjectives placed before the target words

Again, the results here were the same as the previous section, excepting that individual identity was chosen in place of A&B when faced with the phrase *traditional gender roles*.

The target words used as adjectives

Most participants chose the option of individual identity instead of A&B when presented the phrase *gender roles*. However for the phrase *gender neutral* (when in context with words

44

such as *a-sexual* and *questioning*), the majority of Linguistic Student Survey participants and Mechanical Turk participants agreed that the phrase referred to individual identity.

<u>Collocates of *sex,* and *gender,*</u>

Most Mechanical Turk participants agreed with Linguistics Student Survey participants that *sex,* when placed in a list of demographic data referred to physical anatomy. Again, the only change was that instead of choosing A&B on demographic items using the word *gender*, many Mechanical Turk participants chose the option of individual identity.

Again, the question involving *sex,* in a list of demographics which referenced federal law had some interesting results, which I will discuss shortly.

In general, the average answer time of participants was one minute and forty-four seconds. However, out of the more than one hundred participants in the survey, there were at least nine participants who consistently took more than 10 minutes to complete an assignment; even upwards of 20 to 30 minutes. There are a few issues with this average time per question. First, there is no way to know how the participant spent that time. Perhaps they left the assignment to do another activity before answering. Perhaps they were attempting to complete multiple assignments on Mechanical Turk at the same time. Any of the options poses a problem because the lengthy time suggests that the worker was not giving their full attention to the assignment. Therefore, the work of participants who consistently took more than 10 minutes to respond had to be excluded to maintain the integrity of the overall data[29]. Fortunately, these

---

[29] It was not until after I administered the survey and collected the data that I realized that I could set a time limit for each individual assignment (for example, 2 minutes to complete the question). I did set a time limit of 35 minutes, which I thought was for the entire survey. The issue of participants taking longer than 10 minutes to answer a question could have been avoided if I had understood the design of Mechanical Turk better.

outliers were replaced with other workers who completed the assignments in a timely manner, and at no additional cost.

There are some drawbacks to using the Mechanical Turk program, which I will address further in the discussion section. First, although Mechanical Turk was never intended to be a main source of income for individuals[30], several news articles give an insight to workers who consistently rely on Mechanical Turk as their only source of income (Semuels, 2018; Kessler, 2018). The article by Kessler highlights the panic for dependent workers to snatch up the higher paying tasks and get through them as quickly as possible; the only thought being to earn money as quickly and as easily as possible[31]. This means that without doubt, some of the data that I obtained from Mechanical Turk is not as accurate as the data obtained from the Linguistics Student Survey[32].

---

[30] Amazon's recruiting page for Mechanical Turk invites readers to "Make money in your spare time," and there are no benefits or retirement packages offered to workers (Amazon Mechanical Turk Inc., 2018).

[31] In fact, after rejecting the data of one worker, I received over 40 harassment emails within 12 hours, each containing the same message repeatedly asking me to reverse my rejection.

[32] This may be highlighted by the fact that some workers categorized gender as an action in a small percentage of answers.

4 DISCUSSION

The goal of using the semantic surveys was to discover common meaning of the terms *sex* and *gender*, as well as any rules that govern those meanings. Overall, the two surveys (Linguistic Students Survey and Mechanical Turk Survey) were much alike in their data. As I mentioned before, this fact gives validity to the survey items used[33].

A small difference between the two surveys is that the Mechanical Turk Survey participants chose the option of individual identity more frequently than the Linguistic Student Survey participants (30% as compared to 11-14%). The fact that Brigham Young University is a Church school could have influenced this data, owing to the generally conservative demographic. However, participants of the Linguistic Student Survey did choose both A & B (physical anatomy and individual identity) in instances where Mechanical Turk participants chose individual identity. The main difference between the two survey participant groups was that Mechanical Turk participants favored individual identity for *gender* instead of both A & B. Therefore, this small deviation in participant data does not affect the results that much[34].

The main findings that arose from the data came from the examination of each category that items were discovered in on the corpora. From examining the responses from participants of

---

[33] One could question whether the use of linguistics students as participants could make the data less valid; however, the fact that the results of the two surveys closely mirrored each other disputes this idea in my research. In addition, the use of linguistics students as participants was helpful in indicating which questions had the highest agreement rates, which then informed the Mechanical Turk Survey.

[34] There is one item wherein the responses may have been affected by the differing demographics. It reads: "Often male [dolphins] are related to 2 or 3 individuals of the same <u>gender</u> and create partnerships for cooperation purposes." Participants of the Linguistics Student Survey interpreted the meaning of *gender* as Physical Anatomy with 95% agreement. On the other hand, participants of the Mechanical Turk Survey responded with 53% Physical Anatomy, yet 30% Individual Identity. What is striking to me is that 30% of participants thought of dolphins as having an individual identity in addition to their physical anatomy.

each use of the target word in context, I found that the term *sex* is most commonly used to refer to either physical anatomy or to an action, and that the term *gender* can be used to refer to individual identity, but also to physical anatomy. Further, the way that these words are classified can depend on the grammar and context of the sentence.

When a determiner was placed before the term *sex*, it was categorized as physical anatomy when used with a determiner such as *same*. However, when the determiner was, for example, *that*, then *sex* was classified as an action. The questions that achieved the highest agreement rate among participants of both surveys concerned the option an action. I believe this is because an action is the most familiar and easily recognized usage of the word *sex*.

Participants also agreed that when a determiner was placed before the word *gender*, this most often meant physical anatomy. The exception was when there was the context of Gender Inclusive Housing.

When an adjective was placed before the *gender*, such as *masculine*, *gender* was classified as physical anatomy. This is probably because of the binary nature of the context. Although when the phrase *traditional gender* was used, this was classified as individual identity. This could be because people see traditional gender as fluid and changing.

Further, the term *gender* as an adjective when used in the phrase *gender role* was classified as individual identity. This could be similar to the phrase *traditional gender*, which could be seen as fluid. Yet when *gender* was an adjective before words such as school, then *gender* was classified as physical anatomy.

Thus, the results from both surveys show that *gender* can be used to refer to either physical anatomy or individual identity—again, depending on grammar and context. In the context of a list of demographic information, however, most participants concluded that *gender*

refers to individual identity. Five of the questions shared on both surveys give insight on how *sex* and *gender* are interpreted when listed with other demographic data.

Perhaps the most striking observation was the use of *sex* to often refer to physical anatomy, but never individual identity. The term *sex* is often used on medical forms and in a medical context. I would conjecture that a reason participants most commonly chose the response of physical anatomy to describe the word *sex* is because they are familiar with the medical context.

There is, however, one exception to this rule that was discovered in both surveys. As mentioned previously in the Mechanical Turk Survey results section, question 17 was a bit of an anomaly when compared to the other questions concerning *sex,* in a list of demographic data. The answer results on this question were split, with 33% of participants choosing physical anatomy, 27% of participants choosing individual identity, and 32% choosing both physical anatomy and individual identity. I suspect that this is because question 17 references federal laws. Indeed, 55% of participants from the Linguistic Student Survey B chose both physical anatomy and individual identity to answer this question. The full text of question 17 is: "federal laws prohibit harassment based on race, sex, color, national origin, religion, age, and genetic information," taken from eeoc.gov (government website), 2020. Eeoc.gov is the United States Equal Employment Opportunity Commission. I believe that because the issue of equal employment regardless of sexual orientation has been such a big topic of debate in recent years, that many people have begun to think of the term *sex* as referring to individual identity, or both individual identity and physical anatomy, within a legal context. Indeed, it seems to be a value for many Americans that individuals should be treated equally regardless of both physical anatomy and individual identity.

49

There are four items included on both surveys that suggest that the presence of binary contexts results in the majority categorization of physical anatomy. The items are listed below:

A. "… this report was colored by the unwillingness of the women-reformers to admit that their own <u>sex</u> is physically weaker than the sex which votes..." –New York Times (Newspaper), 1875

B. "Peter Pan…was bound to become a musical in time-and doubtless in time for Mary Martin to play Peter. She looks as boyish as can be expected of any grownup of the opposite <u>sex</u>." -TIME (Magazine), 1954

C. "Although Lexi is the only girl on an all-male team, it seems that <u>gender</u> is the only thing that sets them apart." - Professional School Counseling (Magazine), 2010

D. "Some studies in non-athletes have found cortisol levels to…differ between men and women… it [is] important to examine the unique cortisol responses…within each <u>gender</u>." -Journal of Sport Behavior (Academic journal), 2019

In each item, there is a binary option for the terms *sex* and *gender*. The older examples from 1875 and 1954 respectively use the term *sex* to refer to physical anatomy, while the more recent examples from 2010 and 2019 use the term *gender* to do the same. In my research on COHA, the term *gender* was not widely used[35]. However, the term *gender* has 45,913 hits on COCA, and the term has been rising steadily since the 1990s. Owing to the differing usages of the terms *sex* and *gender* depending on the year of the items in question, I analyzed the items based on year to see if this factor affected meaning categorizations. However, I would attest that the year of the item did not affect its categorization. This is because the patterns previously explained are present in all examples regardless of year.

The fact that the term *gender* was not used consistently before the last couple of decades would suggest some semantic space intrusion on the term *sex*. For example, *gender* is now a term that can be used commonly to mean physical anatomy as well. However, it is difficult—from the

---

[35] It was difficult to obtain items for the surveys using the term *gender* from COHA because that term only had a few hits. From 1830-1950, *gender* has 261 hits. In comparison, the term *sex* has 8,685 hits from 1830-1950.

data that this study elicited—to pinpoint the time when the term *gender* began to intrude on the semantic space of the term *sex*. This is because the pattern of responses from participants remained consistent no matter the year of the item. Therefore, these patterns could simply reflect the current interpretation of *gender*, but not any past interpretations.

Mechanical Turk was an excellent resource[36], as suggested by Gibson, et. al. (2011) although there were a few features that the I found not to be user friendly. I discovered one feature when I began the process of rejecting the work of nine individuals[37]. Mechanical Turk has two options that the requester, or the researcher, can use if there is a quality issue with a worker's performance; the requester can decide to reject the work or block the worker (or both). I was hoping that the option to block a worker would also allow the requester to reject their work, but this option was not available. Therefore, I had to use both methods of rejection and blocking. Rejecting the workers assignments posed a further problem: there was not an option to reject every assignment the worker had done. Instead, I had to search for each assignment by the target worker and individually reject them. A lot of time was spent in this endeavor. Although, a benefit of Mechanical Turk is that when a worker's assignment is rejected, data for that assignment can be immediately replaced by that of a new worker.

Overall, however, Mechanical Turk was the best option to gain the data I wanted in a short period of time. Initially, the data from 100 participants per assignment (with a total of 20 assignments) was received within an hour. Additionally, the data was cheap at USD \$.12 per assignment, totaling at USD \$2.40 per survey. Another great feature is that once the data is

---

[36] One caveat of Mechanical Turk for some users is that there is some minimal coding involved for the requester. This is present in the survey design.
[37] These were individuals mentioned previously in the results section, who spent more than 10 minutes to complete each assignment.

received, the requester can download a CSV file with all the data organized in a spreadsheet. This makes the task of analysis very easy.

5 CONCLUSION

When undertaking the project of assigning meaning to the terms *sex* and *gender* within given contexts, I had no expectation for the results that would surface from the surveys I conducted. It was my intention to discover what would arise from the data; accordingly, my two research questions were:

1. Are there differences in meaning and usage between the words *gender* and *sex* in modern society, and, if so,

2. How are those differences defined?

Corpus linguistic analysis that informed widely distributed semantic surveys lead to definitive answers to these questions. After finalizing my research and examining the data from the surveys, I can confirm that there are differences in meaning and usage between the words *gender* and *sex*, within certain contexts; and I can explain how these differences, as discovered in my semantic surveys, are defined.

Firstly, *sex* was never classified by the majority as referring to individual identity, except in just one circumstance. It was consistently classified by the majority to mean physical anatomy, or otherwise an action. The term *sex* might be used to describe an action when certain determiners or adjectives are placed before the word. Otherwise, when *sex* is used as an adjective or in a list of demographic data, it always refers to physical anatomy. Further, the high inter-rater agreement rate on items using the term *sex* could be because this term has been in use much longer than the term *gender*.

The term *gender*, on the other hand, can refer to either individual identity or physical anatomy, depending on grammar and context within the sentence. In most cases, when *gender* is preceded by a determiner, participants referred to *gender* as physical anatomy. When an

adjective which used a binary context was placed before *gender*, or when *gender* was used as an adjective in a binary context, participants referred to *gender* as physical anatomy. However, in any context involving the phrase *gender role*, this was agreed to mean individual identity. Lastly, the data shows that when *gender* is used in a list of demographic data, it generally always refers to individual identity.

Further, the data shows that the year of the item does not affect the participants' interpretation of the terms based on the above described parameters; nor does the sex of the participant. The Linguistics Student Survey showed that whether the participants were male or female, they chose options in much the same way. Further, both the Linguistics Student Survey and the Mechanical Turk Survey show that the patterns of usage did not deviate based on the year of the source.

The data that shows the classification of the terms *sex* and *gender* to mean physical anatomy based on the presence of binary context may help inform the concern regarding the future of women's sports (as highlighted in the introduction). Currently, there are two distinctions within sports regarding the terms *sex* and *gender*: men's sports and women's sports. Because of the binary nature of this context and based on the response data from the semantic surveys conducted, one could conclude that in the case of women's sports, the terms *sex* and *gender* should refer to physical anatomy.

As mentioned before, a fascinating observation comes from participants' categorization of the terms *sex* and *gender* when they are included in a list of demographic data. On both the Mechanical Turk Survey and the Linguistics Student Survey, participants agreed that the term *sex,* when listed with demographic data, refers to physical anatomy, and the term *gender,* most often refers to individual identity. There is at least 60% agreement between raters on this on the

Mechanical Turk Survey, and at least 91% agreement on this on the Linguistic Student Survey. This information may suggest to those involved in writing demographic questionnaires that *sex* should be the term used when eliciting responses about a person's physical anatomy and *gender* should be the term for eliciting responses concerning a person's individual identity. However, in the article referenced at the beginning of this paper: "Pregnant Female Identifying As Male Goes To ER With Urgent Symptoms, Is Listed As 'Obese Man.' She Receives No Urgent Care, Gives Birth To A Dead Baby" even though the hospital used the term *sex* on their admission paperwork, this patient interpreted *sex* to mean individual identity (Parker, 2019). Although this case is very tragic, based on the data obtained in this study the hospital should not necessarily lose the suit based on negligence, if the basis for the case lies on the term *sex* within a demographic context.

There was, of course, an exception to this demographics pattern regarding the term *sex* on the item which involved the federal law restricting discrimination. Title VII of the Civil Rights Act of 1964 declares that, "It shall be an unlawful employment practice for an employer …to fail or refuse to hire or to discharge any individual, or otherwise to discriminate against any individual with respect to his compensation, terms, conditions, or privileges of employment, because of such individual's race, color, religion, *sex*, or national origin…" (emphasis added) (U.S. Equal Employment Opportunity Commission, 2020). This law was written in a time when, according to COHA and COCA, the term *sex* was much more widely used than the term *gender*. I believe that, to compensate for the time in which this law was written (i.e. a time when the term *gender* was not widely used), the participants of both surveys allow that in this context the term *sex* means either individual identity or *both* physical anatomy and individual identity—instead of

following the previously established pattern in which *sex* means physical anatomy when included in a list of demographics.

It is clear from this data that participants intend for Title VII to extend to both physical anatomy and individual identity in reference to the term *sex* within the law. This would mean that the two cases detailed in the introduction of this paper, *Altitude Express v. Zarda* and *R.G & G.R. Harris Funeral Homes v. EEOC*, should result in Zarda and EEOC obtaining equal employment rights in accordance with Title VII.

However, it is also evident that the deviation in the response pattern in this instance is political. This is interesting, because based on the plain meaning rule, the law should use common language (Merriam-Webster's Law Dictionary, 2020). It is clear from this data that with respect to *sex* and *gender*, common usage is currently out of sync with the intent of the law.

There are a few limitations to this study. First, I did not focus on demographic data such as age, sex, or nationality; in fact, I did not elicit this information on the Mechanical Turk Survey. Although I did not find significant differences in answers on the Linguistic Student Survey based on age and sex (and therefore I did not include that information), any of the surveys could elicit different responses based on a wider demographic. For instance, a repeat of the Mechanical Turk Survey would be helpful to assure the validity of my presented data. Second, I did not effectively limit the response time for participants. Perhaps a repeat study of this nature that gives participants only one minute to respond would generate more intuitive answers, instead of allowing participants time to second guess their initial responses. A further limitation lies in the use of Mechanical Turk to achieve data. While Mechanical Turk was an excellent tool to gather data quickly and cost effectively, the articles (previously mentioned) which highlight workers with a goal to earn money as fast as possible are not reassuring in terms

of the quality of work that can be expected from some participants (Kessler, 2018 and Semuels, 2018).

There is one final limitation in this study that is perhaps the most important. Of the 48 total items tested, only one of them included text from a legal context. Consequently, this item arguably revealed the most interesting results. A replicate study using more items from legal contexts could prove to be revealing of how the layperson approaches the term *sex* within the law.

The aim of my methodology was to remain as unbiased as possible regarding the interpretation of the terms *sex* and *gender*. I believe that the use of two semantic surveys administered to approximately 142 participants aided me in this aim. However, while there were several participants involved in interpreting the meaning of the terms, I was the only researcher analyzing the results of the surveys and their implications. A further precaution for future research could be to utilize a group of collaborators to discuss the results and implications of semantic survey data.

Despite its limitations, this study includes important implications for future corpus and semantic survey-based studies concerning meaning. The model of the two surveys –the first administered to trained linguistic students, which then informed the second survey administered to laypeople—proved to be successful. Although participants of the Mechanical Turk Survey did not reach the same level of inter-rater agreement as participants of the Linguistic Student Survey, the majority answers reflected the same responses collected from the Linguistic Student Survey (thus increasing validity). Further, by examining the data of a large group of participants I was able to observe patterns in the data that may not have been possible (or reliable) to discern as a sole interpreter or within a small group of colleagues.

Future focused research on the semantic meanings of the terms *sex* and *gender* could give further enlightenment to the findings described above. Studies that repeat the methods outlined in the administered semantic surveys would give validity to the results achieved here. Additionally, it would be interesting to see what patterns emerge from different contexts involving the terms *sex* and *gender* than the categories presented here.

APPENDIX

Sex and Gender Semantic Survey A

Please provide the following information before beginning the survey:

Age*:

Gender*:

BYU ID number**:

*Age and gender are only asked for inter-rater reliability reasons.

**Your ID number will only be used to deposit the $10 CougarCash to your account.

Directions:

For each item, you will read an example sentence that contains a target word (which will be underlined). Each item will be accompanied by five options that offer possible meaning of the target word. You must choose only one option; the option that BEST reflects the author's intended meaning of the target word.

Example:

Read the following sentence:

"Four-and-a-half-year-old nursery-school children played with same-<u>sex</u> playmates three times as much as they did with cross-<u>sex</u> playmates in mixed groups." -Omni (Science Magazine), 1990

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy

B. Individual identity

C. An action

D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

Before you begin the survey, please now take a moment to ask the administrator any questions you might have. If you have no questions, you may turn the page.

You may begin:

1. Read the following sentence:

"This generational decline in sexual activity was similar across <u>sex</u>, race, religion and the presence of minor children in the household." – poz.com (blog), 2017

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy

B. Individual identity

C. An action

D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

2. Read the following sentence:

"A demographic questionnaire included age, <u>gender</u>, marital status, living situation, geographic region, race/ethnicity…" -Environmental Health Perspectives Journal, 2013

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy

B. Individual identity

C. An action

D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

3. Read the following sentence:

"It is the settled policy of a majority of this nation to recognize no political differences among men, excepting those which necessarily arise from age, <u>sex</u>, and mental sanity…" –American Whig Review (Magazine), 1848

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy

B. Individual identity

C. An action

D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

4. Read the following sentence:

"Recent studies suggest that <u>gender</u> differences in trust and trustworthiness can be attributable to the ways in which men and women find meaning for themselves in relationships..." -NPR Daybreak (Talk Show) 2005

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy

B. Individual identity

C. An action

D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

5. Read the following sentence:

"…the office of elector [is] something which calls for no qualifications beyond a human form and the masculine <u>gender</u>,…" –North American Review (Magazine), 1865

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy

B. Individual identity

C. An action

D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

6. Read the following sentence:

"There are a variety of tests examining bias on age, <u>gender</u>, race, sexuality, and more." – hci.org (blog), 2016

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy

B. Individual identity

C. An action

D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

7. Read the following sentence:

"[The court discussed] the appeal of a Colorado baker who refused to create a wedding cake for the reception of a same <u>sex</u> couple." -Omni (Science Magazine), 1990

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy
B. Individual identity
C. An action
D. Both A and B
E. The meaning is unclear
F. Other (please list) _____

8. Read the following sentence:
"Although Lexi is the only girl on an all-male team, it seems that <u>gender</u> is the only thing that sets them apart." - Professional School Counseling (Magazine), 2010
In this sentence, the <u>target</u> word is referring to:
A. Physical anatomy
B. Individual identity
C. An action
D. Both A and B
E. The meaning is unclear
F. Other (please list) _____

9. Read the following sentence:
"… federal laws prohibit harassment based on race, <u>sex</u>, color, national origin, religion, age, and genetic information." – eeoc.gov (government website), 2020
In this sentence, the <u>target</u> word is referring to:
A. Physical anatomy
B. Individual identity
C. An action
D. Both A and B
E. The meaning is unclear
F. Other (please list) _____

10. Read the following sentence:
"At 6 years old, I knew I was attracted, drawn very heavily to other people my same <u>sex</u>."
-Slate Magazine, 2017
In this sentence, the <u>target</u> word is referring to:
A. Physical anatomy
B. Individual identity
C. An action
D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

11. Read the following sentence:

"If there's a truism about Hollywood, it is that <u>sex</u> sells." -New York Times (Newspaper), 2009

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy

B. Individual identity

C. An action

D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

12. Read the following sentence:

"Authors have suggested that <u>gender</u> roles and experiences with racism and discrimination likely contribute to risk among non-dominant groups." - Academy of Entrepreneurship Journal, 2016

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy

B. Individual identity

C. An action

D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

13. Read the following sentence:

"The Gender Inclusive Housing (GIH) option will provide a welcoming space for students to choose their roommates, regardless of <u>gender,</u> to promote a healthy and safe living and learning environment." – uis.edu (university page), 2020

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy

B. Individual identity

C. An action

D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

14. Read the following sentence:

"Yes, this one about AIDS is very popular. It speaks of the disease for which there is no cure and talks about the need to have only one wife and to practice safe <u>sex</u>." – National Geographic (Magazine), 2009

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy
B. Individual identity
C. An action
D. Both A and B
E. The meaning is unclear
F. Other (please list) _____

15. Read the following sentence:

"One of the most popular of the recent tourists, and not nominally of the masculine <u>gender,</u> loads with inexhaustible ridicule the prudery of the American women." – North American Review (Magazine), 1836

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy
B. Individual identity
C. An action
D. Both A and B
E. The meaning is unclear
F. Other (please list) _____

16. Read the following sentence:

"Broward's domestic partnership policy has always been open to both same sex and opposite <u>sex</u> couples...." –southfloridagaynews.com (blog), 2015

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy
B. Individual identity
C. An action
D. Both A and B
E. The meaning is unclear
F. Other (please list) _____

17. Read the following sentence:

"[Fox was] endowed with personal attractions which made him irresistible with the fair <u>sex</u>…"
–Harpers (Magazine), 1881
In this sentence, the <u>target</u> word is referring to:
A. Physical anatomy
B. Individual identity
C. An action
D. Both A and B
E. The meaning is unclear
F. Other (please list) _____

18. Read the following sentence:
"Opponents of integrating women in combat zones long feared that <u>sex</u> would mean the end of American military prowess. But now birth control is available." -20/20 (Talk Show), 2013
In this sentence, the <u>target</u> word is referring to:
A. Physical anatomy
B. Individual identity
C. An action
D. Both A and B
E. The meaning is unclear
F. Other (please list) _____

19. Read the following sentence:
"It could be an ex-wife or ex-girlfriend, a friend of the same <u>sex</u>, or even a celebrity."
–cafemom.com (blog), 2013
In this sentence, the <u>target</u> word is referring to:
A. Physical anatomy
B. Individual identity
C. An action
D. Both A and B
E. The meaning is unclear
F. Other (please list) _____

20. Read the following sentence:
"Conversion therapy as a practice is a religious, not psychologically-based, practice…The treatment may include….reinforcement techniques that emphasize traditional <u>gender</u> role behavior."   – counseling.org (article), 2013
In this sentence, the <u>target</u> word is referring to:
A. Physical anatomy

B. Individual identity

C. An action

D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

21. Read the following sentence:

"God's Word places great emphasis upon <u>sex</u> only within marriage; purity was to be maintained within marriage.." – cuttingedge.org (religious article), 2007

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy

B. Individual identity

C. An action

D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

22. Read the following sentence:

"… this report was colored by the unwillingness of the women-reformers to admit that their own <u>sex</u> is physically weaker than the sex which votes..." –New York Times (Newspaper), 1875

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy

B. Individual identity

C. An action

D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

23. Read the following sentence:

"When same <u>gender</u> schools are established, boys and girls would follow personal pursuits that will interest, rather than trying to fit into specific stereotypes." –NYLN.org (blog), 2016

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy

B. Individual identity

C. An action

D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

24. Read the following sentence:

"The real secret in life is to be lucky enough to limp through it with a mate, preferably of the opposite <u>sex</u>, who reads you." -Colorado Springs Gazette (Newspaper), 2017

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy

B. Individual identity

C. An action

D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

Sex and Gender Semantic Survey B

Please provide the following information before beginning the survey:

Age*:

Gender*:

BYU ID number**:

*Age and gender are only asked for inter-rater reliability reasons.

**Your ID number will only be used to deposit the $10 CougarCash to your account.

Directions:

For each item, you will read an example sentence that contains a target word (which will be <u>underlined</u>). Each item will be accompanied by five options that offer possible meaning of the target word. You must choose only one option; the option that BEST reflects the author's intended meaning of the target word.

Example:

Read the following sentence:

"Four-and-a-half-year-old nursery-school children played with same-<u>sex</u> playmates three times as much as they did with cross-<u>sex</u> playmates in mixed groups." -Omni (Science Magazine), 1990

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy

B. Individual identity

C. An action

D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

   Before you begin the survey, please now take a moment to ask the administrator any questions you might have. If you have no questions, you may turn the page.

You may begin:

1. Read the following sentence:

"The results have been weighted to take account of household size and number of telephone lines into the residence and to adjust for variation in the sample relating to geographic region, sex, race, age and education." -New York Times (Newspaper), 2004

In this sentence, the target word is referring to:

A. Physical anatomy

B. Individual identity

C. An action

D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

2. Read the following sentence:

"…this already across the United States was a very emotional legal debate, whether states or whether the country should allow same sex marriages. " -CNN Insight (News Talk Show), 2004

In this sentence, the target word is referring to:

A. Physical anatomy

B. Individual identity

C. An action

D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

3. Read the following sentence:

"Even today, NASA continues to recycle the names of mythological figures and great men of history when naming spacecraft and missions….Even spacecraft with seemingly gender neutral names are coded male: Voyager and Pioneer evoke the men who heroically left home and hearth on voyages of exploration." -aeon.co (Academic article), 2016

In this sentence, the target word is referring to:

A. Physical anatomy

B. Individual identity

C. An action

D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

69

4. Read the following sentence:

"We each have a 'gender filter' that can enhance but also distort our perceptions of the world. This <u>gender</u> filter might be a factor when considering the content of reading and writing material that teachers present to students." -Reading Teacher (Academic journal), 2012

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy
B. Individual identity
C. An action
D. Both A and B
E. The meaning is unclear
F. Other (please list) _____

5. Read the following sentence:

"Data were analyzed with Cox proportional hazard models adjusted for child age, child <u>sex</u>, birth order, parents' socioeconomic status…" -PBS NewsHour (Talk Show), 2016

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy
B. Individual identity
C. An action
D. Both A and B
E. The meaning is unclear
F. Other (please list) _____

6. Read the following sentence:

"Secular progressives…have their own list of books they think young people should not read-- for instance, books they claim are tinged with racism or jingoism or that depict traditional <u>gender</u> roles." -USA Today (Magazine), 2013

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy
B. Individual identity
C. An action
D. Both A and B
E. The meaning is unclear
F. Other (please list) _____

7. Read the following sentence:

"Consistent with the perspective that jealousy is a <u>sex</u> specific evolved mate retention module, men in the present study reported more anger and more hurt in response to sexual infidelity than in response to emotional infidelity." -North American Journal of Psychology (Academic journal), 2005

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy

B. Individual identity

C. An action

D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

8. Read the following sentence:

"Whether it be their race, <u>gender</u>, religion, ethnicity, disability, or sexual orientation, a hate group expresses prejudice against people with a particular identity. " -Phsychology.org (Academic article), 2018

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy

B. Individual identity

C. An action

D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

9. Read the following sentence:

"There is also evidence that the experience of divorce…impedes the capacity of children…to form stable, lasting relationships. And there are significant <u>gender</u> differences: Boys tend to 'act out' their feelings of abandonment and resentment, while girls turn inward and become depressed." -Public Interest (Academic journal), 1996

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy

B. Individual identity

C. An action

D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

10. Read the following sentence:

"…Britney got here -- got to see some of her gifts, which were a wrought iron bassinet from her mom, a car seat and a baby bath tub, all gender neutral, of course." -CNN Showbiz (Talk Show), 2005

In this sentence, the target word is referring to:

A. Physical anatomy

B. Individual identity

C. An action

D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

11. Read the following sentence:

"Some studies in non-athletes have found cortisol levels to…differ between men and women… it [is] important to examine the unique cortisol responses…within each gender." -Journal of Sport Behavior (Academic journal), 2019

In this sentence, the target word is referring to:

A. Physical anatomy

B. Individual identity

C. An action

D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

12. Read the following sentence:

"Please know that I HEAR YOU and I feel your pain, " [Holliday] wrote in her apology letter to the lesbian, bisexual, gay, transgender, gender neutral, asexual, questioning, non-binary, gender fluid crowd." -Fox News (Newspaper), 2017

In this sentence, the target word is referring to:

A. Physical anatomy

B. Individual identity

C. An action

D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

13. Read the following sentence:

"Patients were matched with respect to age and <u>sex</u>, disease, and the kind of surgery that- the extent of the disease and the kind of surgery they required." -NPR (Talk Show), 1995

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy

B. Individual identity

C. An action

D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

14. Read the following sentence:

"Students from the same background, of the same <u>sex</u>, attending the same type of school and with the same levels of achievement at age 14 have different work and further study experiences depending on the subjects they take in Year 12" -acer.org (Academic article), 1999

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy

B. Individual identity

C. An action

D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

15. Read the following sentence:

"A rule that limits how much health insurers can vary premiums within a community based on age, <u>gender</u>, health status and other factors, in order to protect those with greater health needs from exorbitant costs." -USA Today (Newspaper), 2012

In this sentence, the <u>target</u> word is referring to:

A. Physical anatomy

B. Individual identity

C. An action

D. Both A and B

E. The meaning is unclear

F. Other (please list) _____

16. Read the following sentence:

73

"Take a look at this list of "warning signs" and see if these statements describe your relationship:...the person you're dating is jealous...of you-he or she gets angry when you text...people of the opposite <u>sex</u>..." -youngwomenshealth.org (Organization), 2020
In this sentence, the <u>target</u> word is referring to:
A. Physical anatomy
B. Individual identity
C. An action
D. Both A and B
E. The meaning is unclear
F. Other (please list) _____

17. Read the following sentence:
"In summer the time is occupied with out-of-door work, but in the comparatively idle days of winter the father, or not infrequently the mother, teaches the children of either <u>sex</u> the common branches." -Atlantic Monthly (Academic journal), 1888
In this sentence, the <u>target</u> word is referring to:
A. Physical anatomy
B. Individual identity
C. An action
D. Both A and B
E. The meaning is unclear
F. Other (please list) _____

18. Read the following sentence:
"Peter Pan...was bound to become a musical in time-and doubtless in time for Mary Martin to play Peter. She looks as boyish as can be expected of any grownup of the opposite <u>sex</u>. " -TIME (Magazine), 1954
In this sentence, the <u>target</u> word is referring to:
A. Physical anatomy
B. Individual identity
C. An action
D. Both A and B
E. The meaning is unclear
F. Other (please list) _____

19. Read the following sentence:
"Often male [dolphins] are related to 2 or 3 individuals of the same <u>gender</u> and create partnerships for cooperation purposes." -dolphins-world.com (Blog), 2017

In this sentence, the target word is referring to:
A. Physical anatomy
B. Individual identity
C. An action
D. Both A and B
E. The meaning is unclear
F. Other (please list) _____

20. Read the following sentence:
"And as everyone knows, boys and girls are so fundamentally, intrinsically different that it's simply impossible for a person of one gender to empathize with and relate to a character of an opposite gender." -huffingtonpost.com (Newspaper), 2012
In this sentence, the target word is referring to:
A. Physical anatomy
B. Individual identity
C. An action
D. Both A and B
E. The meaning is unclear
F. Other (please list) _____

21. Read the following sentence:
"[A] former Army second lieutenant… made headlines…when he shot and killed two women at a yoga studio in Tallahassee, Florida….'His interactions with the opposite sex had gotten him fired from teaching jobs, [and] booted from the Army…' the report said." -New York Post (Newspaper), 2019
In this sentence, the target word is referring to:
A. Physical anatomy
B. Individual identity
C. An action
D. Both A and B
E. The meaning is unclear
F. Other (please list) _____

22. Read the following sentence:
"In the Stanton case the U.S. Supreme Court, in the context of child support, held that the section in the Utah legislation fixing these ages denied the equal protection of the laws as guaranteed by the Fourteenth Amendment in that no valid distinction might be drawn on the basis of sex only."
-lawreform.ie (Academic article), 1977

In this sentence, the <u>target</u> word is referring to:
A. Physical anatomy
B. Individual identity
C. An action
D. Both A and B
E. The meaning is unclear
F. Other (please list) _____

23. Read the following sentence:
"I've seen clients who feel that if they have more <u>sex</u> with more women than any of their friends, they are somehow proving their masculinity," -Essence (Magazine), 2012
In this sentence, the <u>target</u> word is referring to:
A. Physical anatomy
B. Individual identity
C. An action
D. Both A and B
E. The meaning is unclear
F. Other (please list) _____

24. Read the following sentence:
"There is probably no tribe in which formal marriage alone is sexually satisfactory. Perhaps it would be more correct to say that <u>sex</u> is not the most significant fact of marriage." -PBS General Anthropology (Academic journal), 1938
In this sentence, the <u>target</u> word is referring to:
A. Physical anatomy
B. Individual identity
C. An action
D. Both A and B
E. The meaning is unclear
F. Other (please list) _____

Sex and Gender Semantic Survey: Mechanical Turk

Directions:

You will read an example sentence that contains a TARGET word in BOLD. Each item will be accompanied by four options that offer possible meaning of the TARGET word. You must choose the option that best reflects the author's intended meaning of the TARGET word.

In this sentence, the TARGET word is referring to:

"Four-and-a-half-year-old nursery-school children played with same-SEX playmates three times as much as they did with cross-SEX playmates in mixed groups." -Omni (Science Magazine), 1990

Physical anatomy

Individual identity

An action

Both A & B

In this sentence, the TARGET word is referring to:

"Some studies in non-athletes have found cortisol levels to…differ between men and women… it [is] important to examine the unique cortisol responses…within each GENDER." -Journal of Sport Behavior (Academic journal), 2019

Physical anatomy

Individual identity

An action

Both A and B

In this sentence, the TARGET word is referring to:

"Patients were matched with respect to age and SEX, disease, and the kind of surgery that- the extent of the disease and the kind of surgery they required." -NPR (Talk Show), 1995

Physical anatomy

Individual identity

An action

Both A and B

In this sentence, the TARGET word is referring to:
"There is probably no tribe in which formal marriage alone is sexually satisfactory. Perhaps it would be more correct to say that SEX is not the most significant fact of marriage." -PBS General Anthropology (Academic journal), 1938
Physical anatomy
Individual identity
An action
Both A and B

In this sentence, the TARGET word is referring to:
"Authors have suggested that GENDER roles and experiences with racism and discrimination likely contribute to risk among non-dominant groups." - Academy of Entrepreneurship Journal, 2016
Physical anatomy
Individual identity
An action
Both A and B

In this sentence, the TARGET word is referring to:
"The GENDER Inclusive Housing (GIH) option will provide a welcoming space for students to choose their roommates, regardless of GENDER, to promote a healthy and safe living and learning environment." – uis.edu (university page), 2020
Physical anatomy
Individual identity
An action
Both A and B

In this sentence, the TARGET word is referring to:

"Whether it be their race, GENDER, religion, ethnicity, disability, or sexual orientation, a hate group expresses prejudice against people with a particular identity. " -Phsychology.org (Academic article), 2018

Physical anatomy

Individual identity

An action

Both A and B

In this sentence, the TARGET word is referring to:

"If there's a truism about Hollywood, it is that SEX sells." -New York Times (Newspaper), 2009

Physical anatomy

Individual identity

An action

Both A and B

In this sentence, the TARGET word is referring to:

"A demographic questionnaire included age, GENDER, marital status, living situation, geographic region, race/ethnicity…" -Environmental Health Perspectives Journal, 2013

Physical anatomy

Individual identity

An action

Both A and B

In this sentence, the TARGET word is referring to:

"Peter Pan…was bound to become a musical in time-and doubtless in time for Mary Martin to play Peter. She looks as boyish as can be expected of any grownup of the opposite SEX. " -TIME (Magazine), 1954

Physical anatomy

Individual identity

An action

Both A and B

In this sentence, the TARGET word is referring to:

"Yes, this one about AIDS is very popular. It speaks of the disease for which there is no cure and talks about the need to have only one wife and to practice safe SEX." – National Geographic (Magazine), 2009

Physical anatomy

Individual identity

An action

Both A and B

In this sentence, the TARGET word is referring to:

"Although Lexi is the only girl on an all-male team, it seems that GENDER is the only thing that sets them apart." - Professional School Counseling (Magazine), 2010

Physical anatomy

Individual identity

An action

Both A and B

In this sentence, the TARGET word is referring to:

"At 6 years old, I knew I was attracted, drawn very heavily to other people my same SEX." - Slate Magazine, 2017

Physical anatomy

Individual identity

An action

Both A and B

In this sentence, the TARGET word is referring to:

"God's Word places great emphasis upon SEX only within marriage; purity was to be maintained within marriage.." – cuttingedge.org (religious article), 2007

Physical anatomy

Individual identity

An action

Both A and B

In this sentence, the TARGET word is referring to:

"Secular progressives…have their own list of books they think young people should not read--

for instance, books they claim are tinged with racism or jingoism or that

depict traditional GENDER roles." -USA Today (Magazine), 2013

Physical anatomy

Individual identity

An action

Both A and B

In this sentence, the TARGET word is referring to:

"… this report was colored by the unwillingness of the women-reformers to admit that

their own sex is physically weaker than the SEX which votes..." –New York Times (Newspaper),

1875

Physical anatomy

Individual identity

An action

Both A and B

In this sentence, the TARGET word is referring to:

"Please know that I hear you and I feel your pain," [Holliday] wrote in her apology letter to the

lesbian, bisexual, gay, transgender, GENDER neutral, asexual, questioning, non-binary,

GENDER fluid crowd." -Fox News (Newspaper), 2017

Physical anatomy

Individual identity

An action

Both A and B

In this sentence, the TARGET word is referring to:

"Often male [dolphins] are related to 2 or 3 individuals of the same GENDER and create partnerships for cooperation purposes." -dolphins-world.com (Blog), 2017

Physical anatomy

Individual identity

An action

Both A and B

In this sentence, the TARGET word is referring to:

"I've seen clients who feel that if they have more SEX with more women than any of their friends, they are somehow proving their masculinity," -Essence (Magazine), 2012

Physical anatomy

Individual identity

An action

Both A and B

In this sentence, the TARGET word is referring to:

"… federal laws prohibit harassment based on race, SEX, color, national origin, religion, age, and genetic information." – eeoc.gov (government website), 2020

Physical anatomy

Individual identity

An action

Both A and B

In this sentence, the TARGET word is referring to:

"Data were analyzed with Cox proportional hazard models adjusted for child age, child SEX, birth order, parents' socioeconomic status…" -PBS NewsHour (Talk Show), 2016

Physical anatomy

Individual identity

An action

Both A and B

WORKS CITED

Amazon, 2018. Amazon Mechanical Turk Inc. *Available online at https://mturk.com/*

Cunningham and Egbert. 2019. Scientific Methods for Analyzing Original Meaning: Corpus Linguistics and the Emoluments Clauses. *Georgia State University College of Law, Legal Studies Research Paper*. *2019*(2).

Davies, M. 2008- The Corpus of Contemporary American English (COCA): One billion million words, 1990-2019. *Available online at https://www.english-corpora.org/coca/*

Davies, M. 2009. The 385+ million-word Corpus of Contemporary American English (1990-2008+). *International Journal of Corpus Linguistics 14(2),* 159-190.

Davies, M. 2010. The Corpus of Contemporary American English as the first reliable monitor corpus of English. *Literary and Linguistic Computing 25(4)*. 447–464.

Davies, M. 2010- The Corpus of Historical American English (COHA): 400 million words, 1810-2009. *Available online at https://www.english-corpora.org/coha/.*

Davies, M. 2018- The 14 Billion Word iWeb Corpus. *Available online at https://www.english-corpora.org/iWeb/.*

Gibson, E., Piantadosi, S. and Fedorenko, K. 2011. Using Mechanical Turk to Obtain and Analyze English Acceptability Judgments. *Language and Linguistics Compass.* 509-524.

Goldfarb,. 2017. A Lawyer's Introduction to Meaning in the Framework of Corpus Linguistics. *BYU Law Review 1359*.

Hinderaker, J. 2019. "Tracking Timespeak Through Time." *Powerline*.

Howe, A. 2019. "Court to take up LGBT rights in the workplace (Updated)." *SCOTUSblog*.

Kessler, S. 2018. "The Crazy Hacks One Woman Used to Make Money on Mechanical Turk." *Wired.*

Lee, T. and Mouritsen, S. 2018. Judging Ordinary Meaning. *Yale Law 127*(4), 788-879.

Lehmann, E. 1994. Jerzy Neyman: A biographical memoir. National Academy of Sciences Washington D.C.

McEnery, T. and Wilson, A. 2003. Corpus Linguistics. 2d ed. *Edinburgh University Press* Merriam-Webster's Law Dictionary, 2020. *Available online https://www.merriam-webster.com/legal*

Mouritsen, S. 2010. The Dictionary is Not a Fortress. *BYU Law Review 1915(5).*

Mouritsen, S. 2019. Contract Interpretation With Corpus Linguistics. *Washington Law Review 1337,* 94.

Parker, A. 2019. "Pregnant Female Identifying As Male Goes To ER With Urgent Symptoms, Is Listed As 'Obese Man.' She Receives No Urgent Care, Gives Birth To A Dead Baby." *RedState.*

Phillips, J. and White, S. 2018. The Meaning of the Three Emoluments Clauses in the U.S. Constitution: A Corpus Linguistic Analysis of American English From 1760– 1799. *Working Paper.*

Semuels, A. 2018. "The Internet Is Enabling a New Kind of Poorly Paid Hell." *The Atlantic.*

Steinmetz, K. 2019. "Lawmakers Fixate on Sports in House Hearing on Sweeping LGBT Rights Bill." *TIME.*

U.S. Equal Employment Opportunity Commission. 2020. Title VII of the Civil Rights Act of 1964. *eeoc.gov*

Difallah, D., Filatova, E., Ipeirotis, P. 2018. Demographics and Dynamics of Mechanical Turk Workers. *Web Search and Data Mining Conference Presentation.*

VanderKlok, J. 2014. On the Use of Questionnaires in Semantic Fieldwork: a Case Study on Modality. *Proceedings of Language Documentation and Linguistic Theory.*