Of Men, Women, and Computers: Data-Driven Gender Modeling for Improved User Interfaces

Hugo Liu
Massachusetts Institute of Technology
hugo@media.mit.edu

Rada Mihalcea
University of North Texas
rada@cs.unt.edu

Abstract

Men and women have unique sensibilities for information, which can be tapped to create gender-sensitive user interfaces that appeal more specifically to each sex. Building on previous research in gender psychology and also in user modeling, we take a data-driven approach to understanding gender preferences by mining a large corpus of 150,000 weblog entries– half authored by men, half by women. This paper reports two kinds of contributions. First, we employ automatic language processing, semantic analysis, and reflexive ethnography to articulate gender preferences for several dimensions of gender space will provide valuable insight to user interface designers– time, color, size, socialness, affect, and cravings. Second, we employ statistical gender models to build GENDERLENS–a novel intelligent news filtering system that customizes news based on the gender of its reader. A user evaluation found that GENDERLENS successfully predicted men and women's preferences for news, with statistical significance for four out of five news genres tested.

Keywords

gender analysis, natural language processing, user interfaces

1. Introduction

Men are from Mars, women are from Venus –or so the genre of self-help relationship literature would have us believe. But there is truth even in this folk idea– in certain respects, men and women think and feel differently, and perceive, value and understand the world in their own ways. The philosophical inquiry of feminist epistemology that was begun at the turn of the 20th century sought to reveal the fundamentally different modi operandi of men and women, concluding that much of the societal and cultural knowledge that is considered to be universal and generic actually assumes patriarchal priorities and values. The redoubtable feminine theorist Virginia Woolf deconstructed these 'universal' values:

> *"But it is obvious that the values of women differ very often from the values which have been made by the other sex; naturally, this is so. Yet it is the masculine values that prevail. Speaking crudely, football and sport are 'important'; the worship of fashion, the buying of clothes 'trivial'. And these values are inevitably transferred from life to fiction. This is an important book, the critic assumes, because it deals with war. This is an insignificant book because it deals with the feelings of women in a drawing-room."*[1]

If one could uncover men and women's fundamentally different priorities and values, then it may be possible to improve communication between the genders, and to translate news and information into the language of each gender so as to appeal to their unique sensibilities. So what are the priorities and values of men and women? How does one go about uncovering these gender differences? To perform 'gender modeling', we turn to the personal writings of men and women, accessing 150,000 textual entries pulled from the weblogs authored by men and women. The sort of writing found in a weblog may be ideally suited to what we wish to discover, since weblogs often give an intimate account of personal everyday life, and personal viewpoint unto current events. More than just language and syntax, weblogs contain ample evidence of experiences and perceptions, which we attempted to uncover using corpus-based modeling and semantic analysis. Finally, there is nary any pressure for weblog writings to be formal or patriarchal– often they are non-linear and cyclical, and quite free to express the sort of candid values that feminists advocated for in their own écriture féminine movement.

In this paper, we first describe a corpus-based approach to gender modeling in the context of related work in user modeling and gender psychology. Second, we perform a statistical analysis and ethnographic study of gender difference in order to bring to light some insights that are likely to pique the interests of humanists and user interface designers alike. Third, we present GENDERLENS, a novel system for news filtering based on gender models, and present a user study of the GENDERLENS.

2. Related work

Field work in social and gender psychology has had much to say about the differences of men and women. The masculine is stereotyped as detached, rational, and aggressive, and the feminine as nurturing, gentle, and tactful [4]. While some stereotypes are unfounded, sociolinguists do affirm that some communication styles are gendered. It has been found that men and women differ on private versus public speaking, on "report talk" versus "rapport talk"–these and other facets of relational dialectics are gendered and constitute so-called "GenderLects" [25]. Given the reality of distinctions, should not intelligent user interfaces accommodate men and women's different requirements and preferences for how information is communicated to them? The role of gender differences in interfaces has a small but growing body of research. In a study on gender differences in email preferences, [6] argued that "men tend to focus discourse on hierarchy and independence, while women focus on intimacy and solidarity." In investigating gender preference in multimedia interfaces [21], researchers found that girls emphasized "writing, colors, drawing, help" while boys emphasized "control over the computer, sharp moves and many movements on the screen."

The present research also extends recent ethnographic studies addressing the role of gendered language and the "gender gap" in the blogosphere [12], the significance of gender differences in self-dis-

[1] Virginia Woolf: 1929/1989, *A room of one's own*, Harvest Books, 73-74.

ICWSM'2007 Boulder, Colorado, USA

closure strategy in teenage blogs [11], and the validity of author gender predictions based largely on syntactic words (e.g. pronouns, determiners) [10]. This research distinguishes itself from these previous efforts by 1) taking a more substantial cross-sectional sampling of the blogosphere, 2) focusing more specifically on gendered dimensions apropos information interfaces, and 3) developing a more sophisticated repertoire for semantic sense-making than has previously been achieved with syntactic words or raw word frequencies.

In this paper, we will employ computational techniques from user modeling to extend this literature with findings for specific gender preferences in dimensions relevant to user interfaces such as socialness, affect, color, size, and time.

Recent work in user modeling looks beyond explicit user preferences to model the latent aesthetic sensibilities of people [15] in order to improve intelligent user interfaces and provide more sympathetic user experiences. In the context of intelligent user interfaces, one of us has previously examined the role of affective understanding in the interface [16] and the modeling of people's attitudes [17] from their texts. While it is desirable to understand each person's affective patterns, there is also value and insight to be had in understanding the patterns and preferences of whole communities and cultures. To this end, we recently combined corpus-based computational analysis with traditional ethnographic method practiced in social psychology to study the sources of happiness in blog culture [19]. In the research reported in this paper, we further develop our corpus-based approach to include a variety of semantic analysis techniques for modeling differences between the attentional patterns and preferences of men and women, using a very large body of text recently sampled from the blogosphere.

Stereotyped-based and behavior-based modeling are two long-standing approaches to personalization and recommendation in the user modeling literature. Stereotype-based approaches, such as the GRUNDY book recommender system [22], make explicit inferences about users on the basis of their demographic characteristics, and have the advantage that their intuitions are human-readable. Instead, behavior-based approaches, such as collaborative filtering, are based on data rather than the assumptions than often underlie stereotype modeling; they work well and have enabled many intelligent recommender interfaces [9], although they have the drawback of not providing intuitive human-intelligible insights into how the system works.

When approaching gender modeling, we wanted the best of both these worlds– to articulate explicit insights about the genders as done in stereotype modeling and to derive data-driven and scalable models to back gender-sensitive intelligent interfaces as with collaborative filtering. Thus, the goal of a corpus-based modeling of gender difference is two-fold– 1) to use ethnographic methods to generate actionable insight about men and women; and 2) to use the underlying statistical models of the genders to power gender-based customization and recommendation systems. The first part of this paper makes use of semantic analysis and ethnographic methods to articulate in plain English insights about gender differences. The quality of these insights are validated against long-standing psychological research into gender difference. These insights should also have direct import for designers of gender-sensitive user interfaces. The second part of this paper makes use of a corpus-derived statistical model of gender difference to power GENDERLENS – a gender-sensitive news filtering application.

3. Approach

What do women think? And what do men enjoy? We use automatic language processing techniques to determine the dominant traits in diary-like entries authored by men and women, and use these features to model the differences between each gender' interests as expressed in logs of their day-by-day life.

The automatic classification of a text's author gender has been studied in previous research, for the purpose of authorship profiling [13, 23], or the identification of the parties in spoken dialogue [2]. This previous work has proved the feasibility of automatically classifying text by author gender, and we build upon these results in our current work. We go however beyond the task of text classification, by carefully analyzing the types and properties of the most discriminatory features in male and female authored text, and showing that these properties can be used to improve the quality of a system for information access.

Starting with a very large corpus of texts annotated for author gender, we derive the most salient features for each of the two genders, we analyze the category of these top ranked features, and consequently determine the most important dimensions of the gender space. This section describes the data set used in our experiments and the feature scoring mechanism. The main categories identified across the salient features are then analysed in the following section.

3.1 Data

The study is based on a large corpus of blogposts annotated for gender, collected from the *Blogspot* (http://www.blogspot.com) community. We chose to use *Blogspot* as opposed to other blog communities such as *LiveJournal* or *MSN-Spaces*, as it has richer blogger profile annotations including gender, age, location, occupation, and others.

Starting with the names of approximately 300,000 blogs that were updated with a new entry during two randomly selected days (July 27–28, 2006), we collected the profile page of the blog owners (bloggers) and the corresponding profile features. We discarded all the blogs maintained by more than one blogger (collective blogs), and we also discarded the blogs corresponding to bloggers who chose not to include gender information in their profile. Finally, we parsed the entries from the remaining set of blogs, and kept only the blogposts written in English and having a length within a 200–4,000 character limit. Interestingly, although a large fraction of the blogs listed on *Blogspot* are spam, the constraints that a blogger have a profile and that the size of a blogpost be within certain limits removed almost all the spam – to the point that a random hand-check of 100 blogposts revealed clean spam-free data.

The post-processing and profile-based filters left us with a total of about 160,000 blog entries annotated for gender, which after balancing between male and female authors, left us with the final set of 75,000 male blog entries and 75,000 female blog entries. Table 1 shows two sample entries written by a male and a female writer.

Male-authored blogpost
No word back from the Georges Island people on possible use of their power so I'm going to proceed with the QRP plans. Even though the QRP stuff is smaller than the 100 watt outfit, there will still be a significant amount of stuff I'll need to wrestle on to the island. I'll bring the Pelican 1510 case outfitted with the Elecraft K 2.
Female-authored blogpost
You could probably tell that I literally enjoy dressing up in costumes and crap. I just don't have the resources nor the skills to make a good costume. But I'm a resource for outlandish ideas. I remember shocking my host dad when I told him that I enjoy dressing up like that.

Table 1: *Sample blogposts authored by a male and a female writer*

One aspect of interest with respect to the *quality* of this data set was how well male and female writers can be identified based on the blogs they authored. We trained a Naive Bayes text classifier over unigram features (words) and evaluated the classification accuracy using a set of 140,000 blog entries as training data, and the remaining 10,000 blogs posts for test. The classification was measured at 71%, which is a significant improvement over the 50% accuracy associated with the naive baseline of using one gender assignment by default. As it turns out, the gender annotations in this data set are clearly separable, and therefore we can use this corpus to learn gender characteristics.

3.2 Feature scoring

Particularly relevant for our study is the ranking over the saliency of the features in the corpus. Starting with the features identified as important by the Naive Bayes classifier (a threshold of 0.3 was used in the feature selection process), we selected all those features that had a total weight exceeding a given threshold T, where a feature weight is calculated for each category (male/female) and is determined as the probability of seeing the feature in a given category. We then calculate the *gender score* of a feature as the ratio between the weight in the female-authored corpus and the total weight in the entire blog corpus. This results in a score within the [0–1] interval, with a value closer to 1 indicating a feature representative for the female-authored corpus, and a value closer to 0 corresponding to high saliency features from the male-authored blog dataset.

$$Weight_C(F) = P(F|C)) \approx \frac{Count(F)}{Count(C)} \quad (1)$$

$$GenderScore(F) = \frac{Weight_{female}(F)}{Weight_{female}(F) + Weight_{male}(F)} \quad (2)$$

For instance, assuming the feature *cake* has a weight of 0.025 for the female-authored category of blog entries, and a weight of 0.007 for the male-authored category, this results in a *gender score* of $0.025/(0.007 + 0.025) = 0.78$.

Table 2 shows the top discriminatory unigram features when a threshold T of 500 was used. A similar process has been applied to bigrams and trigrams, and sample discriminatory features are also shown in Table 2.

	Female	Male
Unigrams	knitting	microsoft
	hubby	democrats
	yarn	poker
Bigrams	my husband	my wife
	love him	of Israel
	so excited	prime minister
Trigrams	I love him	my wife and
	so much fun	of the United
	I miss my	the Bush administration

Table 2: *Top discriminatory unigram/bigram/trigram features*

4. Dimensions of the gender space

The perspectives of men and women are complex systems, but these complex systems may afford simple and elegant projections capable of producing deft insight into what the genders share and how they differ. This is the undertaking that is related in the present section. Using WordNet to cluster together dominant n-gram features into larger conceptual buckets, several promising categories –time, color, size, affect, socialness, and cravings– emerged as affording the greatest discriminatory power over our gendered blog corpus. Upon further inspection, these 'categories' would more aptly be called 'dimensions', since they are not the target of storytelling so much as its modes. After all, most all experiences and perceptions are tinted by temporality, feeling, color, size, socialness, and even metaphors for food. These dimensions can also directly inform user interface design, since interfaces have color schemes and sized elements, task flows often imply temporality and socialness, and information content is usually designed to appeal to particular affects or cravings. To illuminate how these dimensions factor into gender difference, we had to address the semantic microcosm of each dimension with separate experiments. The differences which emerge seem to bolster a common theme –the dichotomy that men think in generalities and women think in particularities.

4.1 Time

Our analysis revealed that men and women tended to experience time somehow differently. We hoped to infer how men and women value time by measuring how they talked about timed events. On what time scale? Did they focus on the past, the present moment, or on the future? We identified linguistic features corresponding to temporal expressions from everyday speech– such as "tomorrow," "saturday," – and tracked the semantic orientation of these phrases toward masculine and feminine writings in the blog corpus.

First, we compiled a list of English adverbs pertaining to time. Next, we used this lexicon to filtered over the lists of dominant n-gram features, keeping just those features containing at least one temporal lexeme. This resulted in a set of temporal expressions, which we then segregated by hand into 1) relative-time expressions (strongly deictic) such as "last week," and 2) concrete-time expressions (weakly deictic) such as "wednesday." Taken as a class, neither relative-time expressions nor concrete-time expressions held much discriminatory power for gender, meaning that men and women do not prefer to use relative-time expressions over concrete-time expressions, or vice versa. However, when we sorted each class of expressions chronologically and by time scale, a gender difference emerged.

Figure 1 illustrates this difference– representative subsets of relative-time expressions (top) and concrete-time expressions (bottom) were isolated, and were ordered chronologically and by granularity. The y-axes show the semantic orientation of each expression (-1.0=male; 0.0=neutral; +1.0=female). The graph of relative-time expressions (top) suggests that women are more likely than men to value the events of the here-and-now, from "last weekend" through to "this weekend." On the other hand, men are more likely to focus on events of the past and future months and years. The graph of concrete-time expressions (bottom) more clearly illustrate how men and women might enjoy life on time scales. Feminine writing dominates the days-of-the-week. Masculine writing prefers to focus on months-of-the-year.[2] The graphs in Figure 1 also hint that factors other than time are at play –for example, one might note feminine preference for more social times, e.g. weekends. That the month of August is talked about more by women than by men seems for a moment to break the pattern of men thinking in months, until we recall that the corpus was sampled from the blogosphere in the last few days of July– so it makes sense that women would focus on August since it is imminent and women have a preference for the imminent (cf. Figure 1, top).

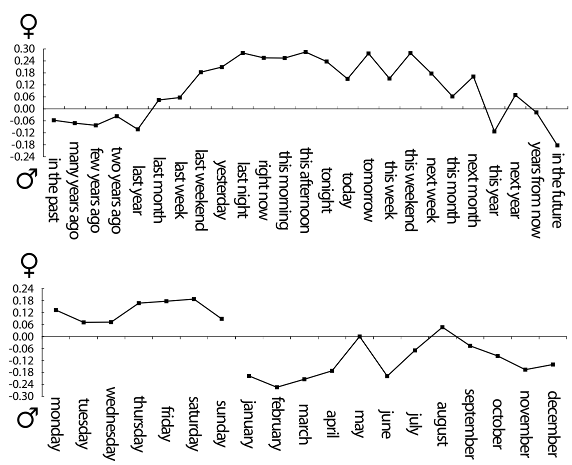

Fig. 1: *Men are from march, women are from this weekend: the role of time in feminine and masculine writings*

[2] A few of the temporal expressions were quite obviously polysemous (e.g. "may" as the month or as a modal); but for simplicity and reproducibility, we refrained from sense disambiguation of linguistic features, here, and in other parts of the study

4.2 Food

A cursory flip through the list of dominant features reveals a substantial number of references to food and eating, such as ”baking,” ”yummy,” and ”cookies,” which were some of the most feminine words in the corpus. Literally, what this shows is that women were more likely to write about food and eating than were men; the affective valence underlying these speech acts were not taken into account so each utterance about food can be motivated by either desire or disgust. This notwithstanding, one can pragmatically infer that writing about food is evidence of preoccupation with, valuing of, or attending to food. From this premise, we made formal calculations to gauge the extent of men’s and women’s preoccupation with food in general, and to gauge their interests in particular types of food.

Our experiment was to utilize the ontology of food terms from the machine-readable dictionary, WordNet [20], to summarize male and female interest in various subcategories of food. WordNet version 2.1 defines three senses for the word ”food” – food as nutrient (1st sense, with direct hyponyms such as ”foodstuff,” ”beverage,” and ”nutriment”), solid food (2nd sense, with direct hyponyms such as ”chocolate,” ”pasta,” and ”yogurt”), and food for thought (3rd sense). The first two senses being relevant, we followed their hyponym relations (subsumed concepts) down to the leaf concepts and harvested food expressions associated with each food subcategory[3]. For example, $food_1$[4]’s hyponym tree contained 1394 concepts, generating 2047 unique food expressions; $food_2$’s hyponym tree contained 1109 concepts, accounting for 1632 unique food expressions. By taking a food category’s subsumed expressions to represent its linguistic context, we can ’summarize’ the semantic orientation of each kind of food as being male-leaning or female-leaning. A simple baseline approach is used, as outlined in [27]. Each food category’s expressions are mapped onto features in the blog corpus, resulting in a discrimination score. The semantic orientation of the food category is defined as the average of these discrimination scores. The results of this analysis are shown in Figure 2.

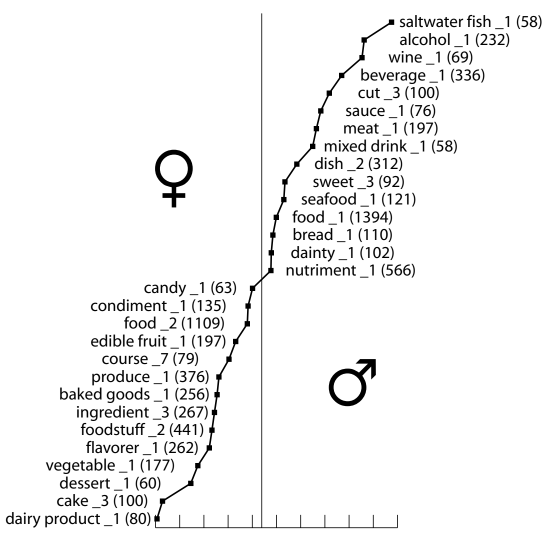

Fig. 2: *Food preferences*

The graph labels can be read as ”concept_sensenumber (number of subsumed hyponyms).” To produce this graph, only large food categories (with 50 or more subsumed hyponyms) are shown, and the semantic orientation scores were normalized to the range (-1.0=male to +1.0=female). The reader will note that the top-level categories, $food_1$ and $food_2$ are oriented toward the female pole– this is due to the fact that there were 54% more food-related vocabulary invoked in female blogs than in male blogs. The category axis was drawn through 0.18 because that is the average semantic orientation of all the food terms– doing so permits the reader to inspect male and female gravitations toward particular sorts of food. Dairy, desserts, and produce were the most feminine foods, while alcoholic beverages, meats, and sauces were the most masculine foods. Sweets and healthy foods were typically feminine, while liquids and hearty foods were typically masculine. Another trend that can be inferred from these results is that women paid more attention to the details and intricacies of food, evident in the female-leaning of granular descriptors in $ingredient_3$, $foodstuff_2$, $flavorer$, while men thought about foods more abstractly as gestalts $sauce_1$, and more biologically as necessary nutrients rather than as sensual pleasures. This finding is sympathetic to findings about gender and time, which also suggested that there may exist a particularity-generality dichotomy distinguishing women from men, respectively.

4.3 Color

In analyzing the genders’ uses of color description in their writings, we obtained results consistent with the particularity-generality pattern that was observed for the time and food dimensions. To make a measurement of men’s and women’s preferences for colors, we started with the widely-used X11 color lexicon[5] because it tends to prefer folk color names over composite names (e.g. ”chartreuse” rather than ”brilliant-greenish-yellow”). The X11 color lexicon is constituted by 144 color names, but we only considered a subset of 53 color names that were one-worded. We mapped those color names into linguistic features extracted from the blog corpus to determine the semantic orientation of each color toward masculine or feminine and found that overall, use of any color description leans toward the feminine (0.18). Looking at the invocation of particular colors in writings, the use of ”navy,” ”gold,” and ”silver” were most telling of masculine writing; the use of ”purple,” ”tan,” and ”pink” were telltale of feminine writing. Figure 3 (top) shows the six most masculine and six most feminine colors.

Next, we used average-semantic-orientation and the known RGB values of the X11 colors to cluster color usage along dimensions such as saturation, hue, brightness, redness, and color temperature but found that no true pattern emerged. A final dimension that we checked for patterns along was color ”order” – a concept in color theory which prescribes every color as being either primary, secondary, tertiary (3rd order), quaternary (4th order), and so on. Following the RGB system, red, green, and blue are the three primary colors; yellow, magenta, and cyan are three secondary colors because they can be produced by mixing two primary colors; tertiary colors are those produced by mixing primary and secondary colors; and so forth. As such, color order gives a sense of a color’s complexity and nuance –for example, it takes greater color sensitivity to correctly perceive a higher-order color than a lower-order one. We assigned color order values to our X11 colors by performing a nearest-neighbor mapping of their RGB into the known RGB centroids of primary, secondary, tertiary, and quaternary colors. Unfortunately the color order system does not usually account for shading (adding of black) and tinting (adding of white) of colors, so we took some liberty to factor this into our complexity calculation by adding white and black as primary colors; thus a shaded green and tinted red will be assigned to the second order.

Figure 3 (bottom) shows the result of sorting colors by their order. Grey dots indicate raw data points for the semantic orientation of each color. Black dots indicate the averaged semantic orientation

[3] In WordNet, each sense-disambiguated word is associated with a set of surface linguistic expressions, known as a synset

[4] Throughout the paper, we use a subscript notation to indicate the WordNet sense number.

[5] http://tools.ietf.org/html/rfc1198

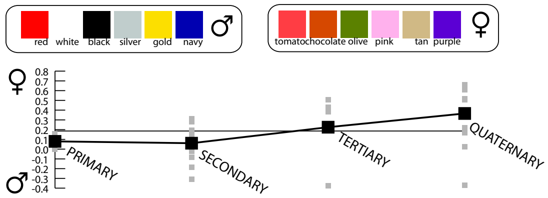

Fig. 3: *Masculine and feminine colors and color order*

of all colors of a particular color order. The x-axis crosses value axis at 0.18 because that is the average semantic orientation of all the X11 colors that were considered (recall that women used color description more often than men). The graph formalizes the intuition embodied in the color swatches of Figure 3 (top) as it shows that men tended to think in terms of bolder and simpler primary and secondary colors, while women tended to think in terms of more complex and more nuanced tertiary and quaternary colors. By regarding primariness as a type of generality, and by regarding color complexity as a type of specification, we can interpret these findings as consistent with the particularity-generality dichotomy we are increasingly building a case for.

4.4 Size

How big or small are the objects, people, and ideals that men and women prefer to focus upon? According to Lakoff and Johnson [14], size is one of only a handful of fundamental spatial metaphors that enframes our thoughts and guides us toward certain interpretations of the world. Based on the linguistic concept of "gradability," we devised an experiment to estimate the scale of things that men and women think about. Gradability means the willingness with which a noun phrase accepts a graded modifier (e.g. "soft shoes") or an adjective accepts an intensifier (e.g. "extremely hot"). Previous research in computational linguistics has studied the significance of adjective gradability to the identification of a text as being subjective rather than objective [8]. Here, our technique is to test the "size-gradability" of the adjectives and nouns that are most predictive of feminine and masculine writing by querying Google for graded terms, in order to arrive at a most general impression of size of masculine and feminine objects and ideals. We began with a list of the most frequent 4000 unigram features (annotated with their gender-discrimination power) from the blog corpus, and compiled two lists– the top 1/3 most masculine words, and the top 1/3 most feminine words. To test for gradability, we generated five size-graded expressions for each word in these lists using an inventory of five most common size adjectives, previously used in [24]. For example, the feminine feature "skirt" generated the terms: "tiny skirt," "small skirt," "average skirt," "big skirt," "huge skirt."

Feeding these graded expressions into Google, we recorded the number of results for each term –accepting that count as an estimate of the commonality of an object or idea being a certain size. The technique of using Web search counts to estimate the semantic orientation of words was first described in [28]. In a previous blog analysis [19], we successfully employed the technique to study patterns in everyday life. To derive the size distribution for each term T, we performed unit normalization over the log of the raw counts, shown in the expression:

$$Distribution_{size}T = \frac{Count(T) - \overline{log(Count(T, size))}}{\sigma(log(Count(T, size)))} \quad (3)$$

where $size$ can take any of the values tiny, small, average, big, or huge.

The overall 'tinyness' of masculine things was estimated by averaging across the 'tinyness' of the most masculine terms, and so on for 'small', 'average', 'big', and 'huge'. The results of this analysis are shown in Figure 4.

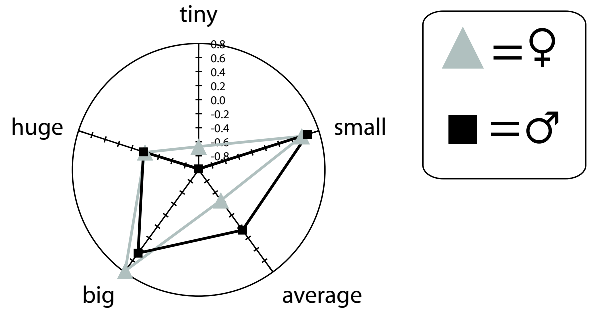

Fig. 4: *Estimating the size-gradability of masculine and feminine unigram features*

The results were quite telling. Feminine affairs were more likely to be size-gradable than were masculine affairs– finding a well-rounded distribution across the five sizes that were tested. On the other hand, men rarely considered tiny things, and were much more likely than women to consider average things– from 'average', we can infer that a thing is hard to size, perhaps because the thing is abstract, unspecific, or difficult to reify. The fact that women consider tiny things much more than do men, and the fact that men are more likely to consider unsizeable things supports the particularity-generality gender dichotomy that we have been finding across all the dimensions being studied.

4.5 Socialness

How social are the genders? Long-standing research in socio-psychology reminds us that there is some truth to be had in the old stereotype that women more than men are endowed with the ability for social gap and gossip [4]. It too has been written that men appraise their happiness in terms of the success and career they have achieved, while women assess their own happiness on the longevity and depth of their personal relationships [25]. So is the prevailing wisdom about the genders and socialness correct? Does evidence from the gendered blog corpus corroborate or contravene those findings?

In our previous but related blog study on the topic of happiness [19], we used Google and the web-based semantic collocation approach –just applied to estimate size gradability– to find that social contexts were linked much more to happiness than to sadness. In the present study, we aimed to paint a more detailed and telling picture of the genders and socialness.

First, we turned to WordNet to make a formal measurement of men and women's social focus. There are two most relevant hyponym subtrees. The hierarchy below $relative_3$ contains an extensive catalog of familial and kinship relations such as "aunty," "sibling," and "groom" –hence we mined 352 terms representing family and intimate socialness. Next, the hierarchy below $socialgroup_1$ yielded 2394 terms about more general groupings and associative units in society and culture (many of which appear to be work-related) such as "staff," "church," "bikers," and "tribe." Unlike $relative_3$, terms under $socialgroup_1$ seem to imply mostly weak social relations; while technically speaking a "faculty" or a "government" is social, these terms usually connote coldness, detachment, and anti-socialness. Our experimental technique was to again filter the masculine and feminine n-gram feature lists against the lexicon harvested from these two WordNet subtrees, and to calculate the overall gender leaning of $relative_3$ and $socialgroup_1$ as the average of the discriminating power of individual terms. The results of the experiment found that $relative_3$ saw an average orientation of 0.16 (on the aforementioned -1.0 to +1.0 scale), thus leaning toward the feminine; and $socialgroup_1$ saw an average orientation of -0.22, thus leaning toward the masculine. The findings agreed with the previous research in gender psychology [25], indicating that women were more likely to focus on relationships with immediate family members and loved ones whereas men were more likely to focus on sociality in the societal sense, opining on work-related, political, and cultural groups more than on close-knit relationships.

While the use of WordNet lexicons for comparative sensing of masculine and feminine texts is clearly a semantic measurement, it is also possible to glean the semantics of gender socialness from syntactic features. In particular, we are interested in how men and women make use of pronouns such as "I," "we," "you," "them," "its." In the semiotic study of textual stylistics [26], seemingly pure syntactic features –such as sentence length, choice of determiners, and choice of pronouns– are viewed systematically on a large scale, and are thought to reveal psychological and affective aspects of the writer. In the case of pronouns, for example, the dominance of first-person plural pronouns like "we," "us," "ourselves" could suggest that the author experiences relationship or group identity more than individual identity. To make a formal measurement of the role of pronouns in masculine and feminine texts, we considered the unambiguous set of 31 English pronouns, and labeled them into five characteristic groups– 1st-person singular e.g. "I"; 1st-person plural e.g. "we"; 2nd-person e.g. "you"; 3rd-person e.g. "he," "she," "it"; and possessives e.g. "my," "his." We referred to our list of dominant unigram features to assign discriminatory-power scores to the pronouns, and calculated the semantic orientation of each group as the average semantic orientation of its subsumed pronouns.

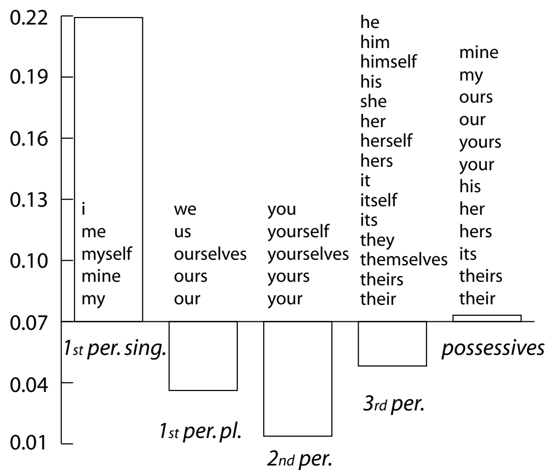

Fig. 5: *Semantic orientation of pronouns*

Figure 5 shows the results of pronoun analysis. The average semantic orientation of all pronouns was 0.07 on the -1.0 to +1.0 scale, hence women were more likely to use pronouns than men. Women made much more use of 1st-person singular pronouns (i.e. "I," "me," "myself," "mine," and "my") than did men. Men made more use of 1st-person plural, 2nd-person, and 3rd-person pronouns relative to all pronouns than did women. Of all 3rd-person pronouns, men preferred the impersonal pronouns i.e. "it," "itself," "its," whose average orientation was -0.10, while women preferred the personal pronouns, such as "she" (0.32), "him" (0.23), and "hers" (0.33). We were not surprised to find that men preferred impersonal pronouns while women preferred personal pronouns, because the above WordNet experiment found that women are invested in personal relationships while men are invested in abstract, societal groups. We were somewhat surprised to find women making such disproportional use of "I," "me," and "my," though this could indicate that women preferred to talk about their own immediate circumstances while men preferred to address "you," the more abstract reader or hypothetical person. We were also surprised that 1st-person plural pronouns i.e. "we," "us," "ourselves," "ours," and "our" did not lean more toward the feminine because we assumed that these pronouns would indicate close social context such as relationships and families; however, this assumption was perhaps unwarranted, as men could as easily speak "we" to mean membership in societal groups, e.g. from the corpus: "We conservatives have few who we could say are better advocates for us." The results of pronoun analysis are compatible with and indeed support our other findings for socialness of the genders.

4.6 Affect

Not all things, people, and ideals will naturally connote color, food, time, socialness, or size; but virtually every thing, person, and ideal entails a default affective context. That is because, as many psychologists have theorized, affect parameterizes and enframes all cognition [7], organizing thoughts into affective buckets so that they can be more efficiently accessed. Prima facie, one would intuit that men and women have different affective dispositions; indeed, psychologists report that the feminine is stereotyped as soft, gentle, and emotionally vulnerable, and the masculine is stereotyped as emotionally detached, rational, and aggressive [4].

To uncover the latent emotional lives of men and women via linguistic analysis, it is necessary to know the general affective context that is associated with things, people, and ideals. To this end, we made use of ANEW [3] –a set of normative affective ratings for 1034 common English words, obtained by psychometry over focus groups. ANEW rates words using the pleasure-arousal-dominance (PAD) model of emotion [18]. Employing ANEW as a knowledge base of affective ground truths, we analyzed the top 4000 unigram features from the blog corpus by filtering for words in ANEW. 823 out of ANEW's 1034 words were utilized in this set of 4000 features.

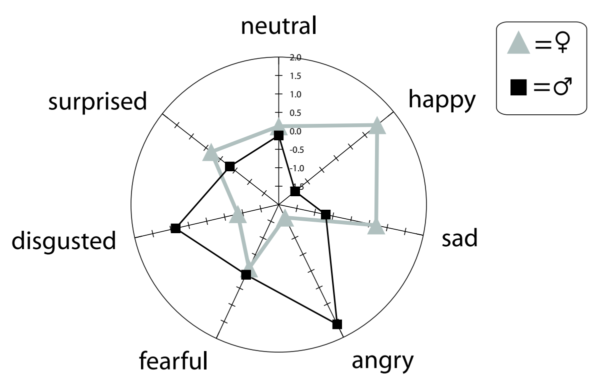

Fig. 6: *Emotional characteristics of masculine and feminine features*

Segregating the unigram features into masculine and feminine, we proceeded to calculate the average pleasure, arousal, and dominance level of men and women. We found that, with statistical significance to 90% confidence, women were occupied with more pleasurable topics than were men ($Pleasure_{male} = 0.047$; $Pleasure_{female} = 0.096$), while men were focused on topics that were more arousing ($Arousal_{male} = 0.048$; $Arousal_{female} = 0.014$) than women's. As for dominance, men and women could not be distinguished with statistical significance. As aggressivity has two components– arousal and dominance– these averages only lend partial support to the masculine stereotype reported in [4].

Next, we wanted to understand how men and women distributed their attentions across different named emotions. We opted for Ekman's [5] ontology of six universal emotions + 'neutral'. To create a mapping of Ekman's emotions into PAD-space, we assigned

Fig. 7: *A snapshot of* GenderLens

ground truth coordinates to each of the seven emotion states consistent with the intuition applied in [1]. We calculated the average unit-normalized Cartesian distance between the masculine point cloud and each emotion state, and repeated the process for the feminine point cloud. Under the assumption that emotions are conserved within men and women, we performed a second normalization. Figure 6 shows the resulting distribution of emotions for men and women.

These results show that women are most likely to focus on happy topics, while men were least likely to focus on them. Men instead relied on topics generally evoking anger and disgust. Women were more likely to opine about sad things. Neutral and fearful topics garnered equal proportions of men and women's attention.

5. A gender lens for information access

To evaluate the possible role of the gender features we observed in our corpus study for creating better interfaces, we designed GenderLens – a news filtering system that reranks the daily news based on the gender biases learned from the blog data set. Since men and women tend to have different interests, our hypothesis is that this fact will be reflected in their preference toward different stories from the day-by-day news.

GenderLens is reading the news feed from a major news aggregator (Google News), and is reordering the news according to each of the two gender-biased language models. We use the top 14,000 most discriminatory unigram features extracted from the blog dataset. We use their associated *gender scores* reflecting their saliency in the data, calculated according to the feature scoring mechanism described earlier in the paper.

Next, a gender score is computed for each of the stories in the news feed, determined as the average across all the gender features found in the news article. For efficiency considerations, the gender score is calculated for the summary of the news story, rather than the entire article, to ensure that the processing and reranking are performed in real-time. Similar to the scores computed for individual features, the gender score of a news story ranges from 0 to 1, with a value closer to 1 indicating a bias toward womens-specific features (learned from the female-authored corpus), and a value closer to 0 reflecting a bias toward men-oriented features (with high saliency in the male-authored data).

Finally, the news items are reordered based on their gender score, resulting in two columns – one (left column) where the news stories are ordered in decreasing order of their gender score, likely to correspond to the women's prioritization of news interestingness, and one (right column) where the stories are ranked in increasing order of the gender score, reflecting increased interest for a man reader. Figure 7 shows a snapshot of GenderLens, with the top ranked stories in each column.

5.1 Evaluation

To evaluate the effectiveness of GenderLens, we conducted a user study where 30 users (15 men and 15 women) were asked to indicate their preference for one of the two gender-biased news columns. We considered five news categories – top stories, world stories, science and technology, sports, and entertainment – allowing us to determine the possible role of gender preferences for different news topics.

For each news category, the users were shown the female-ranked and male-ranked news columns side-by-side, and were asked to indicate which column best fitted their own interests. In order to avoid any bias, the true purpose of the study was concealed as a topical news filtering study, all gender references were removed from the interface, and the left and right columns were randomly swapped across categories.

The results of the study were evaluated with respect to the agreement between the actual gender of a user and our predicted preference for one of the two female-biased or male-biased news streams. Figure 8 shows the agreement measured for each of the news categories, together with the Pearson correlation and the corresponding level of significance.

5.2 Discussion

These results show that the statistical gender models derived from weblogs were successful in predicting gender preference for news in four of five news categories. One hurdle which is often a factor but did not seem to impede the success of GenderLens is the use of models trained over one text genre (weblog) for prediction in a different genre (news). This may suggest the generality and

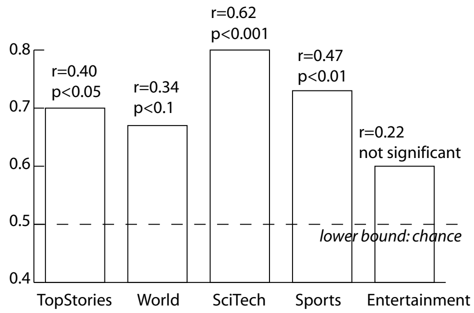

Fig. 8: *Results of the user evaluation of* GENDERLENS

reusability of weblog-derived models for interface personalization tasks. In fact, lending support to this suggestion is the fact that, when subjects were asked to state why they preferred a certain column, the vocabulary of their responses was very much in-line with the dominant feminine and masculine features that were learned from blogs. To report some of the representative descriptors, men valued news which they stated was ”informative,” ”powerful,” ”useful,” and ”funny.” Women valued news which was ”new,” ”cute,” ”chic,” and ”beautiful.” Men seemed to cite rational and objective factors for their preference, while women cited most sensitive and affective factors for their preference. Surprisingly, gender prediction for ’SciTech’ was most successful. Looking at the reasons stated for judgment, we found that women preferred science and technology that was ”fascinating,” ”cute,” and ”not technical,” while men preferred that which was ”latest,” ”innovative,” and ”geekish.” Gender prediction for entertainment news was notably not successful with statistical significance. Upon further inspection, there was a significant preference for the feminine side of entertainment news, with 12 of 15 female subjects and 6 of 15 male subjects preferring it. This suggests that perhaps the inherent success of certain kinds of information, such as entertainment news, is in its ability to titillate and inspire gossip, which seems to be more supported by the feminine model than by the masculine model. Looking broadly across all the categories, we made another observation– every category leaned slightly toward the feminine, averaging 20% bias. How could this be? Maybe the feminine lens is actually selecting for news that is more interesting to read. Or perhaps because news in general tends to be dry and stogy, feminine news tends to stand out at a glance because it does not give off typically stale airs that more formal news puts off.

6. Conclusions [gender-balanced]

In this paper, we tried to gain insights into how men and women perceive day-by-day events, and what they most value in their daily experiences, by looking at a very large number of diary entries extracted from the blogosphere. Our analysis of gender distinctions revealed that women's and men's sensibilities exhibited a particularity-generality dichotomy that swept all dimensions of gender space. Women focused on immediate time, nuanced colors, close-knit relationships, objects describable by size, the flavors of food, and were disposed to happiness and sadness. Men focused on months and years, primary colors, social hierarchies, abstract ideas, food as a tool for sating hunger, and were disposed to anger and arousal. These findings generally agreed with previous research in gender psychology, but do articulate gender tendencies with greater specificity apropos user interface design. We then used the most important traits learned from this study to create GENDERLENS – a real-time news system that reorders the stories in a news stream based on the facts more likely to be of interest for men and for women. A 30-person user evaluation of GENDERLENS found that the system was able to correctly predict the preferences of men and women for different news categories with statistical significance for four out of five news genres– thus demonstrating the promise of gender-based customization for improving user interfaces.

References

[1] C. Becker, S. Kopp, and I. Wachsmuth. Simulating the emotion dynamics of a multimodal conversational agent. In *Proceedings of Affective Dialogue Systems*, 2004.

[2] C. Boulis and M. Ostendorf. A quantitative analysis of lexical differences between genders in telephone conversations. In *Proceedings of ACL*, 2005.

[3] M. Bradley and P. Lang. Affective norms for english words (anew): Instruction manual and affective ratings. Technical report, The Center for Research in Psychophysiology, University of Florida, 1999.

[4] J. Doyle. *Sex and Gender: The Human Experience*. Wm. C. Brown Publishers, 1985.

[5] P. Ekman. Facial expression of emotion. *American Psychologist*, 48, 1993.

[6] D. Gefen and D. Straub. Gender differences in the perception and use of e-mail. *MIS Quarterly*, 21(4), 1997.

[7] D. Gelernter. *The Muse in the Machine : Computerizing the Poetry of Human Thought*. NY: The Free Press, 1994.

[8] V. Hatzivassiloglou and J. Wiebe. Effects of adjective orientation and gradability on sentence subjectivity. In *Proceedings of COLING*, 2000.

[9] J. Herlocker, J. Konstan, L. Terveen, and J. Riedl. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 1(22), 2004.

[10] S. Herring and J. Paolillo. Gender and genre variation in weblogs. *Journal of Sociolinguistics*, 2004.

[11] D. Huffaker and S. L. Calvert. Gender, identity and language use in teenage blogs. *Journal of Computer-Mediated Communication*, 2005.

[12] T. L. M. Kennedy, J. S. Robinson, and K. Trammell. *Does gender matter? Examining conversations in the blogosphere*. 2005.

[13] M. Koppel, S. Argamon, and A. Shimoni. Automatically categorizing written texts by author gender. *Literary and Linguistic Computing*, 4(17), 2002.

[14] G. Lakoff and M.Johnson. *Metaphors we live by*. University of Chicago, 1980.

[15] H. Liu. *Computing Point-of-View: Modeling and Simulating Judgments of Taste*. PhD thesis, Program in Media Arts and Sciences, Massachusetts Institute of Technology, 2006.

[16] H. Liu, H. Lieberman, and T. Selker. A model of textual affect sensing using real-world knowledge. In *Proceedings of ACM IUI*, 2003.

[17] H. Liu and P. Maes. What would they think? a computational model of attitudes. In *Proceedings of ACM IUI*, 2004.

[18] A. Mehrabian. Framework for a comprehensive description and measurement of emotional states. *Genetic, Social, and General Psychology Monographs*, 121, 1995.

[19] R. Mihalcea and H. Liu. A corpus-based approach to finding happiness. In *Proceedings of the AAAI Spring Symposium on Computational Approaches to Weblogs*, 2006.

[20] G. Miller. Wordnet: A lexical database. *Communication of the ACM*, 38(11), 1995.

[21] D. Passig and H. Levin. Gender preferences for multimedia interfaces. *Journal of Computer Assisted Learning*, 16(1), 2000.

[22] E. Rich. User modeling via stereotypes. *Cognitive Science*, 3:329–342, 1979.

[23] J. Schler, M. Koppel, S. Argamon, and J. Pennebaker. Effects of age and gender on blogging. In *Proceedings of AAAI Spring Symposium on Computational Approaches for Analyzing Weblogs*, 2006.

[24] S. Sharoff. Meaning as use: exploitation of aligned corpora for the contrastive study of lexical semantics. In *Proceedings of LREC*, 2002.

[25] D. Tannen. *You Just Don't Understand: Women and Men in Conversation*. London, Virago, 1991.

[26] T. Todorov. *Introduction to Poetics*. University of Minnesota Press, 1981.

[27] P. Turney. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings of ACL*, 2002.

[28] P. Turney and M. Littman. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems*, 4(21), 2003.