FontLex: A Typographical Lexicon based on Affective Associations

Tugba Kulahcioglu, Gerard de Melo
Rutgers University – New Brunswick
Department of Computer Science, Piscataway, NJ, USA
{tugba.kulahcioglu, gerard.demelo}@rutgers.edu

Abstract

The task of selecting suitable fonts for a given text is non-trivial, as tens of thousands of fonts are available, and the choice of font has been shown to affect the perception of the text as well as of the author or of the brand being advertized. Aiming to support the development of font recommendation tools, we create a typographical lexicon providing associations between words and fonts. We achieve this by means of affective evocations, making use of font–emotion and word–emotion relationships. For this purpose, we first determine font vectors for a set of ten emotion attributes, based on word similarities and antonymy information. We evaluate these associations through a user study via Mechanical Turk, which, for eight of the ten emotions, shows a strong user preference towards the fonts that are found to be congruent by our predicted data. Subsequently, this data is used to calculate font vectors for specific words, by relying on the emotion associations of a given word. This leads to a set of font associations for 6.4K words. We again evaluate the resulting dataset using Mechanical Turk, on 25 randomly sampled words. For the majority of these words, the responses indicate that fonts with strong associations are preferred, and for all except 2 words, fonts with weak associations are dispreferred. Finally, we further extend the dataset using synonyms of font attributes and emotion names. The resulting FontLex resource provides mappings between 6.7K words and 200 fonts.

Keywords: typography, font, emotion

1. Introduction

It is not uncommon for people to spend several minutes looking for the right font, but finally ending up using the default one (Fox, 2010). One would indeed be well-advised to spend some effort on typographic choices, as the choice of font has been shown to be able to affect the perception of the text as well as of the author (Juni and Gross, 2008; Shaikh, 2007b; Shaikh et al., 2007), or of associated products and brands (Fligner, 2013).

The task of selecting a suitable font is particularly burdensome for graphic designers, as their profession calls for such decisions to be made on a regular basis, and the mere use of a neutral font may negatively affect the perception of the related brand or product (Shaikh, 2007a; Shaikh, 2007b).

Even more severe than opting for a neutral font is to end up picking an ill-suited font typeface. This is commonly described as the font being *incongruent* with the underlying meaning or theme (e.g., writing the word "*happy*" with a font perceived as unhappy). As previous studies have revealed, the use of incongruent fonts not only increases the response time of users (Lewis and Walker, 1989; Hazlett et al., 2013), but can also have a particularly detrimental effect on the perception of the related product (Fligner, 2013; Childers and Jass, 2002; Van Rompay and Pruyn, 2011).

To make things worse, the task of font selection is becoming ever more challenging as the number of available fonts keeps increasing. Google Fonts[1] as of October 2017 provides a catalog of 822 font families, while broader font sharing websites[2] typically serve several thousands.

Despite the obvious need, the assistance offered by current tools remains very limited. Some websites (dafont.com, 2017; Bloch, 2017), and recently also word processing tools such as Microsoft Word, provide a categorized presentation of fonts for users to explore, based on visual attributes as well as also certain semantic ones. In addition to the exploratory approach, O'Donovan et al. (2014) present a method to recommend fonts that are semantically similar to the current font selection. In Qiao (2017), vector representations are used to generate font pairs. The visualization by Data Scope Analytics (2017) aims to help users in discovering aesthetically pleasing font pairs. Previous work, however, neglects the content of the text to be formatted, and in particular neglects the affective dimension of human perception.

Towards the aim of supporting the development of font recommendation tools based on the *textual content* and the associated *affect* of the message, in this study, we produce FontLex, a typographical lexicon that maps 6,721 words to a set of 200 fonts. As our main method, we rely on word–emotion and font–emotion associations, and connect words with fonts via their affective associations. This gives rise to 200-dimensional vector embeddings that capture the strength of the association between a given word with each of the 200 considered fonts. In an additional step, we also gather synonyms of the font attribute names from WordNet, which enables us to obtain font vectors for these words more directly from the font vectors of the related attribute.

The rest of this paper is organized as follows. First of all, Section 2. reviews related work on semantic attributes of fonts and on font recommendation techniques. Section 3. presents our method to predict emotion–font scores and evaluates it through a user study. Section 4. presents our method to predict word–font scores using the previously obtained emotion–font scores, and evaluates it through a further user study. Subsequently, Section 6. describes the semantic extension of the dataset using synonym relation-

[1] `https://fonts.google.com`
[2] For instance, `https://www.dafont.com/` and `http://www.1001fonts.com/`

ships. Finally, Section 7. concludes the paper and outlines plans for future work.

2. Related Work

We begin with a review of previous studies and tools that have approached the topic of semantic attributes of fonts or the goal of recommending fonts.

2.1. Semantic Attributes of Fonts

Through a crowdsourced study, O'Donovan et al. (2014) associate 200 fonts with 37 semantic attributes (e.g., happy). They ask users to pick one of two presented fonts for a given attribute, and then based on these selections assign scores between 0 and 100 for each font–attribute combination. The resulting dataset is publicly available[3] and will be discussed further in Section 3.

Kulahcioglu and de Melo (2018) extend the above crowdsourced dataset using deep Convolutional Neural Network (CNN) embeddings as a means of obtaining a similarity measure between fonts. To predict semantic attribute scores for a font outside the dataset, the authors take weighted averages of the nearest four font scores, as determined by the embeddings. Based on leave-one-out cross validation test results, the method is able to predict scores with around 9% mean absolute error.

In an online survey conducted by Shaikh et al. (2006), the characteristics of 20 fonts are assessed with respect to 15 adjective pairs (e.g., stable – unstable). The fonts are presented using alphabetic, numeral, and common symbols.

Further studies (Velasco et al., 2014; Velasco et al., 2015) analyze the relationship between visual font characteristics and taste attributes (sweet, sour, etc.) through user studies. They conclude that round fonts exhibit an association with sweet taste.

Finally, many font-focused websites (Sam Berlow and Sherman, 2017; dafont.com, 2017; Bloch, 2017) allow contributors to tag fonts with attributes, some of which are more semantic than visual.

2.2. Font Recommendation

O'Donovan et al. (2014) present a method of proposing fonts that are similar to a given font that is currently being used. In their experiments, they find that semantic attributes are more conducive to predicting the similarity of fonts than geometrical features. Thus, making use of a set of semantic attributes, they learn a font similarity metric based on crowdsourced comparisons, in which users need to assess which of two presented fonts is more similar to a provided reference font.

Wang et al. (2015) rely on a deep learning approach to find similar fonts. It is claimed that a qualitative comparison of both methods reveals this approach as producing better results than the former one by O'Donovan et al. (2014).

Using vector representations, Qiao (2017) aims to identify fonts that are both contrasting and complementary. The system can either propose a novel pair of fonts, or suggest a second font for an already specified one. The vector representations are provided online[4].

The force-directed graph visualization[5] developed by Data Scope Analytics (2017) displays 458 fonts and 1,807 co-usages gathered by Sam Berlow and Sherman (2017). The visualization[6] in Ho (2017) displays around 800 font embeddings mapped into a 2D space.

Several websites, including those of Sam Berlow and Sherman (2017), Canva.com (2017) and Mills (2017), provide font pair suggestions gathered from users or from other web sources.

2.3. Impact of Font Choices

A number of Stroop-style studies have been conducted to investigate the effect of font characteristics on perception. Hazlett et al. (2013) asked users to judge whether a displayed word is positive or negative, comparing 5 fonts and 25 words that are all strongly associated with positive or negative emotion. The results indicate that congruent typefaces yield faster responses. Lewis and Walker (1989) ask users to press a left hand key if the words *slow* or *heavy* appear, versus a right hand one if *fast* or *light* appears. In a second experiment, they display related words (e.g., *fox*) instead of the original words (e.g., *fast*) to ensure that the user needs to grasp the meaning of the displayed word. In both experiments, they repeat the tasks with congruent and incongruent fonts, finding that the former significantly reduce the response time.

In terms of survey-style studies, Juni and Gross (2008) present newspaper articles using two different fonts. Their survey reveals that the same text is perceived as more humorous or angry when read in a certain font compared to another. Shaikh (2007b) presents documents to participants using congruent, incongruent, and neutral fonts, while soliciting ratings to assess the perception of the document (e.g., as *exciting*) as well as the perceived personality of the author (e.g., in terms of trustworthiness). The findings show strong effects across the assessed font types with respect to the perception of documents, whereas congruent and neutral fonts appear to evoke similar perceptions of an author's personality.

Shaikh et al. (2007) study the effect of the choice of font on email perception. Their results suggest that fonts with low congruency may result in different perceptions of an email than fonts with medium to high congruency. A similar study on the perception of a company website (Shaikh, 2007a) demonstrates that neutral and low congruency fonts can negatively affect a company's perception in terms of professionalism, believability, trust, and intent to act on the site.

Many studies in marketing analyze font effects, especially in packaging design. For instance, Fligner (2013) shows that fonts associated with the attribute *natural* increase the perceived *healthfulness* of products when used in their packaging, particularly if the products' intrinsic cues (e.g., being fat-free) and extrinsic ones (e.g., being sold at Whole

[3] http://www.dgp.toronto.edu/~donovan/font/

[4] https://github.com/Jack000/fontjoy

[5] https://datascopeanalytics.com/fontstellations/

[6] http://fontmap.ideo.com/

	anger	anticipation	disgust	fear	joy	negative	positive	sadness	surprise	trust
1	¬calm	fresh	clumsy	bad	happy	bad	strong	¬happy	dramatic	strong
2	clumsy	formal	bad	capitals	playful	strong	¬bad	gentle	happy	calm
3	capitals	dramatic	sloppy	¬calm	graceful	sharp	happy	¬graceful	¬sharp	¬bad

Table 1: Top three closest attributes, where ¬ indicates attributes that are negated

Foods Market) also concur. Childers and Jass (2002) establish that the semantic attributes of a font bear an impact on user perception for both high and low engagement levels. Through experiments using bottled water of a fictional brand, Van Rompay and Pruyn (2011) finds additional evidence that the congruence between fonts and other design elements influence the perception of brand credibility, aesthetics, and value.

3. Emotion Mapping

In this section, we describe our method to obtain font scores for the emotion attributes that shall later, in the following section, be used to obtain font scores for words in an existing emotion lexicon.

3.1. Method

Our method assumes as input a set of fonts $\mathcal{F}$ that are described in terms of a set of font attributes $\mathcal{A}$. For this, we rely on the crowdsourced data from O'Donovan et al. (2014), which for a given font $f \in \mathcal{F}$ provides scores in $[0, 100]$ for each attribute $a \in \mathcal{A}$. From this data, we derive $|\mathcal{F}|$-dimensional vectors $\vec{a} \in [0,1]^{|\mathcal{F}|}$ for each font attribute $a \in \mathcal{A}$. For this, we simply transform the dataset to consider the fonts for a given font attribute, normalizing scores to $[0, 1]$.

Then, to induce FontLex, we first generate $|\mathcal{F}|$-dimensional font vectors for a set of emotion attributes $\mathcal{E}$. Subsequently, using existing word–emotion associations, we will infer $|\mathcal{F}|$-dimensional font vectors for words such that each component of such a vector quantifies the strength of the association between a word and a font.

As the set of emotions $\mathcal{E}$, we consider the ten emotion attributes used in EmoLex (Mohammad and Turney, 2013). Our first step is to map these $e \in \mathcal{E}$ to vectors $\vec{e} \in R^{|\mathcal{F}|}$ that characterize their association with fonts $f \in \mathcal{F}$ in our data. To achieve this, we proceed as follows. For each emotion $e \in \mathcal{E}$, we determine the $k = 3$ most similar font attributes $a \in \mathcal{A}$, as shown in Table 1. To decide on this value, we have carried out leave-one-out tests on the crowdsourced seed dataset (O'Donovan et al., 2014). Although the average overall success of the method in terms of the mean error was slightly higher for higher k than 3, we found that for $k = 3$ the most attributes attained their highest scores. Also considering the complexity of the negation decisions as will be described shortly, we opted to use the closest $k = 3$ neighbors.

We rely on word2vec distances $d(e, a)$, using cosine distances on the standard word2vec Google News pretrained model[7], to determine similarity scores $\mathrm{sim}(e, a)$ between emotion names and font attribute names as below:

$$\mathrm{sim}(e, a_i) = \frac{1}{k-1} \frac{\sum_{\substack{j=1 \\ i \neq j}}^{k} d(e, a_j)}{\sum_{j=1}^{k} d(e, a_j)} \tag{1}$$

One aspect that needs to be addressed, however, is the widely known fact that distributional models of semantics tend to conflate synonyms with antonyms. Hence, we first define

$$\vec{\mu}(e, a) = \begin{cases} \vec{1} - \vec{a} & \text{if } a \text{ is assessed as an antonym of } e \\ \vec{a} & \text{otherwise,} \end{cases} \tag{2}$$

where $\vec{1}$ is an $|\mathcal{F}|$-dimensional vector of ones. Thus, for those words that are assessed as antonyms, we do not use the regular font vector $\vec{a}$, but instead consider an inverted vector, in which we subtract each value from the maximum value of 1. The assessment is performed manually. For relationships such as between *anger* and *calm*, determining antonym relationships was straightforward. However, for some more challenging decisions, such as *negative* and *sharp*, we evaluated both options and discussed the obtained results with a graphic designer before making the final decision. In Table 1, attributes labelled as antonyms are marked with a "¬" symbol.

To obtain font vectors $\vec{e}$ for emotions $e \in \mathcal{E}$, we compute

$$\vec{e} = \sum_{i=1}^{k} \mathrm{sim}(e, a_i)\, \vec{\mu}(e, a_i) \tag{3}$$

where the a_i are the k most similar attributes, as described above. Thus, the font vectors are a weighted average of the vectors for related attributes, after possibly inverting their respective vectors.

3.2. Results

Figure 1 depicts the top 3 fonts that are most strongly associated with the ten emotion attributes, whereas Figure 3 shows the three fonts for each emotion that are found to have the weakest associations. Figure 2 shows sample fonts that are predicted to be neutral in terms of the respective emotion, which are ranked in the middle of the ranked font list. In all figures, the emotion names are rendered using the corresponding fonts.

The fonts that are strongly associated with emotions share some special characteristics. For instance, for *joy*, we encounter handwriting-style typefaces, whereas for *disgust*, we find display fonts with salient stylization. It should also be noted that not all fonts that share these characteristics are

[7] https://code.google.com/archive/p/word2vec/

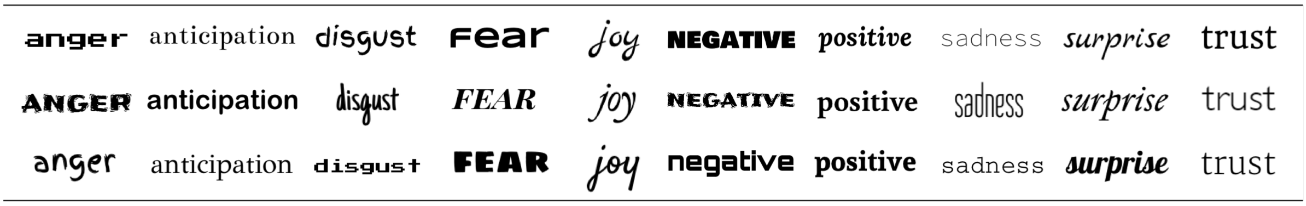

Figure 1: Emotion attributes rendered using the three most congruent fonts as predicted by our method. The renderings on the first line uses the fonts ranked 1st, the second line uses fonts ranked 2nd, and the third line uses fonts ranked 3rd.

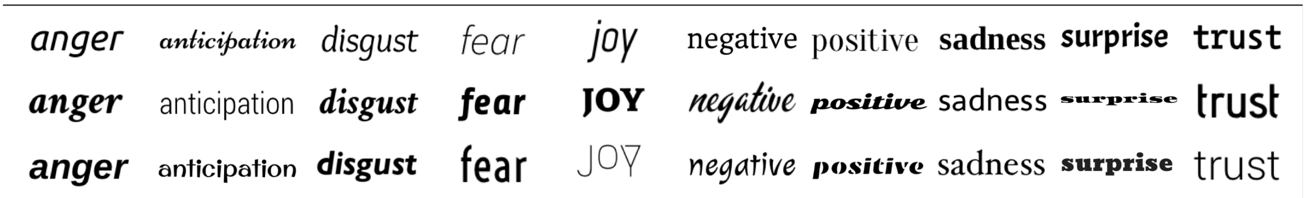

Figure 2: Emotion attributes rendered using the neutral fonts as predicted by our method. The renderings on the first line use the fonts ranked 99th, the second line uses fonts ranked 100th, and the third line uses fonts ranked 101st.

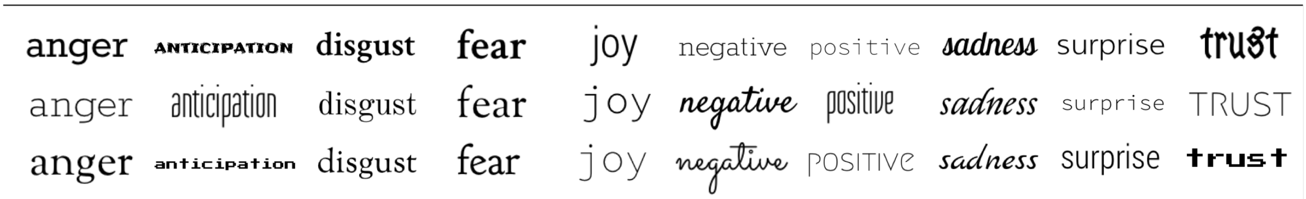

Figure 3: Emotion attributes using the three most incongruent fonts as predicted by our method. The renderings on the first line use the fonts ranked 198th, the second line uses fonts ranked 199th, and the third line uses fonts ranked 200th.

strongly associated with these emotions, since the relationships between emotion attributes and font characteristics are not straightforward (Kulahcioglu and de Melo, 2018).

Figure 4: An example task for *positive*. The second and fifth fonts are congruent, the third and fourth is incongruent and the first is neutral.

3.3. Evaluation

To assess the quality of the obtained emotion font score predictions, we carry out a user study.

3.3.1. User Study

For each of the ten emotion attributes, we generated four tasks with different random font choices. An example is given in Figure 4. Each task includes 5 fonts, two congruent fonts selected randomly among the top-scoring 10 fonts for that emotion, two incongruent fonts selected randomly among the lowest-scoring 10 fonts for that emotion, and one neutral font selected randomly among the ten fonts that are in the middle of the ranked list of fonts. In each task, the user is requested to select the image that best represents the word. As described above, the available options include the same word presented using five different fonts.

	Congruent	Neutral	Incongruent
Expected value	40.00	20.00	40.00
anger	74.04	14.42	11.54
anticipation	28.85	34.62	36.54
disgust	70.19	10.58	19.23
fear	78.85	5.77	15.38
joy	91.35	4.81	3.85
negative	59.80	15.69	24.51
positive	60.00	23.81	16.19
sadness	46.15	16.35	37.50
surprise	72.12	8.65	19.23
trust	62.50	20.19	17.31
Average	64.38	15.49	20.13

Table 2: Evaluation Results (in %) for Emotions

Each task is carried out by 30 participants in Mechanical Turk, all from the United States, with at least 5,000 approved hits and an overall approval rating of 97% or more. We used counterbalancing, i.e., half of the users received the tasks in the reverse order from the other half. We also used three validation tasks, and eliminated results of three participants who incorrectly answered all three of them.

3.3.2. Evaluation Results

Table 2 summarizes the results of this user study. The *congruent* column lists the percentages of selections in which

Figure 5: Selected words rendered using the three most congruent fonts as predicted by our method. The renderings on the first line uses the fonts ranked 1st, the second line uses fonts ranked 2nd and the third line uses fonts ranked 3rd.

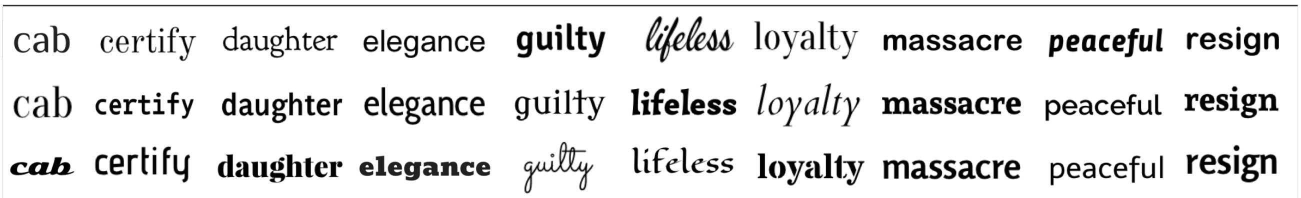

Figure 6: Selected words rendered using the three most congruent fonts as predicted by our method. The renderings on the first line uses the fonts ranked 99th, the second line uses fonts ranked 100th and the third line uses fonts ranked 101st.

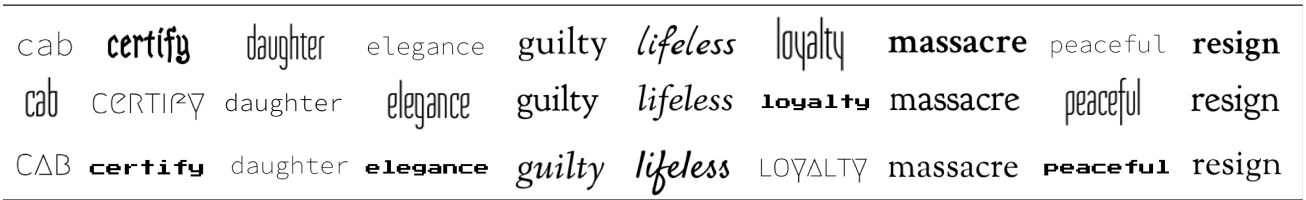

Figure 7: Selected words rendered using the three most incongruent fonts as predicted by our method. The renderings on the first line uses the fonts ranked 198th, the second line uses fonts ranked 199th and the third line uses fonts ranked 200th.

the congruent fonts (those in the top 10 for that word) are preferred. Similarly, the *neutral* and *incongruent* columns list the percentages of choices of neutral and incongruent fonts, respectively. The first row lists the expected value assuming the null hypothesis of a uniform distribution over the five choices, of which 2 are congruent, 1 neutral, and 2 incongruent.

The average is 64.38% for congruent font preferences. Compared to the expected value of 40%, this shows a strong trend toward the fonts predicted to be congruent, hence validating our results in general. Similarly, the preferences for the fonts that are found to be incongruent by our method was much lower than the expected value, with an average of only 20.13%.

However, a detailed look at the values for individual emotion attributes reveal that the performance differs between them. The strongest preference is obtained for *joy*, with a value of 91.35%, whereas the lowest is for *anticipation* with 28.85%. Another comparably low value is obtained for *sadness*, with a congruency of 46.15%. This suggests that different emotions may differ in how saliently and uniquely they are associated with visual font characteristics (cf. Section 5.).

4. Lexical Mapping

The next phase involves computing font vectors for words.

4.1. Method

EmoLex (Mohammad and Turney, 2013) provides binary emotion association indicators between words and the emotion attributes $e \in \mathcal{E}$ listed in Table 1. There are 6,468 words with at least one emotion association in their data. For words w in this set, we consider their data as providing vectors $\vec{w}_{\mathrm{E}} \in [0,1]^{|\mathcal{E}|}$.

To generate a font vector $\vec{w}_{\mathrm{F}}$ for a word w, we compute

$$\vec{w}_{\mathrm{F}} = \frac{1}{\|\vec{w}_{\mathrm{E}}\|_1} \mathbf{M}_{\mathrm{E}} \vec{w}_{\mathrm{E}} \tag{4}$$

where $\|\vec{w}_{\mathrm{E}}\|_1$ denotes the ℓ_1 norm of $\vec{w}_{\mathrm{E}}$ and $\mathbf{M}_{\mathrm{E}} = [\vec{e}_1 \dots \vec{e}_{|E|}]$, i.e., a matrix with columns that capture the font vectors for the emotions $e \in \mathcal{E}$ (in the same order as captured in $\vec{w}_{\mathrm{E}}$).

4.2. Results

Figure 5 shows the top three congruent fonts associated with ten sample words, Figure 7 shows the most incongruent three fonts for the same words, and Figure 6 shows sample fonts that are predicted to be neutral for the respective words. In all images, the words are rendered using the corresponding fonts. These words are among those used in the evaluation user study in the following section.

4.3. Evaluation

We evaluate the dataset through a user study. In the following, we provide details on the design and the results of this study.

	Congruent	Neutral	Incongruent	Corresponding Emotion Attributes									
Expected value	40.00	20.00	40.00	AG	AN	D	F	J	N	P	SA	SU	T
appreciation	70.59	15.69	13.73	0	0	0	0	1	0	1	0	0	1
cab	53.85	11.54	34.62	0	0	0	0	0	0	1	0	0	0
certify	79.59	6.12	14.29	0	0	0	0	0	0	0	0	0	1
conformance	61.54	19.23	19.23	0	0	0	0	0	0	1	0	0	0
congenial	42.86	20.41	36.73	0	0	0	0	0	0	1	0	0	0
daughter	70.59	13.73	15.69	0	0	0	0	1	0	1	0	0	0
elegance	76.00	16.00	8.00	0	1	0	0	1	0	1	0	0	1
guiilty	49.02	21.57	29.41	1	0	0	0	0	1	0	1	0	0
instruct	55.77	28.85	15.38	0	0	0	0	0	0	1	0	0	1
kill	75.00	5.77	19.23	0	0	0	1	0	1	0	1	0	0
lifeless	32.00	26.00	42.00	0	0	0	1	0	1	0	1	0	0
loyalty	76.92	9.62	13.46	0	0	0	0	0	0	1	0	0	1
massacre	56.00	16.00	28.00	1	0	1	1	0	1	0	1	0	0
medley	40.38	36.54	23.08	0	0	0	0	0	0	1	0	0	0
murky	82.35	5.88	11.76	0	0	1	0	0	1	0	1	0	0
noble	72.55	15.69	11.76	0	0	0	0	0	0	1	0	0	1
oracle	52.00	16.00	32.00	0	1	0	0	0	0	1	0	0	1
outcome	64.71	17.65	17.65	0	0	0	0	0	0	1	0	0	0
peaceful	64.00	12.00	24.00	0	1	0	0	1	0	1	0	1	1
persistent	65.38	9.62	25.00	0	0	0	0	0	0	1	0	0	0
precedence	56.86	15.69	27.45	0	0	0	0	0	0	1	0	0	1
resign	20.00	18.00	62.00	1	0	1	1	0	1	0	1	0	0
shameful	50.00	28.85	21.15	0	0	0	0	0	1	0	1	0	0
tickle	63.46	9.62	26.92	0	1	0	0	1	0	1	0	1	1
verified	64.71	23.53	11.76	0	0	0	0	0	0	1	0	0	1
Average	59.85	16.78	23.37										

Table 3: Evaluation Results (in %) and Emotion Associations for Words in the User Study. (AG: Anger, AN: Anticipation, D: Disgust, F: Fear, J: Joy, N: Negative, P: Positive, SA: Sadness, SU: Surprise, T: Trust)

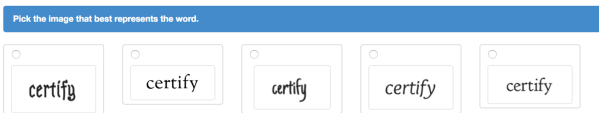

Figure 8: An example task for the word *certify*. The second and fifth fonts are congruent, the first and third is incongruent, and the fourth is neutral.

4.3.1. User Study

For our study, we consider 25 words randomly selected from the set of words with at least one salient font association. For this purpose, we consider any of the 3,882 words that have a score of 0.75 or higher in any of the components of their respective font vectors. For each of the random 25 words, we generated two tasks with different random font choices. We have reduced the number of tasks to two, compared to the four tasks used in the previous section, to keep the total number of tasks reasonable for each participant.

An example task for the word *certify* is given in Figure 8. Each task includes 5 fonts, two congruent fonts selected randomly among the top-scoring 5 fonts for that word, two incongruent fonts selected randomly among the lowest-scoring 5 fonts for that word, and one neutral font selected randomly among the three fonts that are in the middle of the ranked list of fonts for the word. The decision to use 5 fonts as opposed to 10 is again based on considerations regarding the workload per user.

Each task involves a user being requested to select the image that best represents the word. As described above, the available options include the same word presented using five different fonts. Each task is carried out by 30 participants in Mechanical Turk, all from the United States, with at least 5,000 approved hits and an overall approval rating of 97% or more. We used counterbalancing and eliminated results of one participant that accidentally completed both of the original and reversed task sessions. We have also used three validation tasks, and eliminated results of one participant that incorrectly answered both of the two validation tasks.

4.3.2. Evaluation Results

Table 3 summarizes the evaluation results for the 25 randomly selected words as described above. The *congruent* column lists the percentages of selections in which the congruent fonts (those in the top 5 for that word) are preferred. Similarly, the *neutral* and *incongruent* columns list the percentages of choices of neutral and incongruent fonts, respectively.

The average is 59.85% for congruent font preferences, which shows that the consensus between our data and the users were strong. The strongest preference is obtained for the word *murky*, with a value of 82.35%, whereas the lowest is for the word *resign* with 20.00%. Similarly, the average for the incongruent preferences was only 23.37%, bearing further witness to the quality of the results. Only two out of twenty-five words, namely *lifeless* and *resign*, received congruent preferences that are less than the expected value of 40%. Such results are expected, given that different words may differ in the strength and uniqueness of their associations (cf. Section 5.).

Table 3 also displays the corresponding emotions for the words used in the evaluation, allowing us to analyze the relationship between the success of the two datasets. In some cases, words associated with the same set of emotions obtained similar user ratings, such as *instruct*, *noble*, *precedence*, and *verified*. Whereas in some cases, words with the same emotion set obtained quite divergent ratings: *massacre* and *resign*.

5. Discussion on Results

We have introduced two datasets that connect emotions and words with fonts in terms of real-valued scores. Besides showing strong support for the datasets, the user evaluations also revealed that the performance varies for different emotions and words. Below, we discuss the potential sources for these differences.

For the emotion–font dataset, one reason for the differences between results could be the varying potential of fonts to represent or evoke different emotions (Kulahcioglu and de Melo, 2018). This could be observed in the results for *anticipation*, for which determining a font type may prove difficult even for an experienced graphic designer. It is also observed that emotions with higher arousal, namely *anger*, *disgust*, *fear*, *joy*, and *surprise*, received higher congruent user preferences compared to other emotions, which may be a direction that merits further analysis.

The second reason may be a lack of appropriate similar attributes in the crowdsourced seed dataset. Looking at Table 1, it could be argued that *joy* has semantically close neighbors in the dataset, whereas this is not the case for *anticipation*.

For the word–font dataset, checking the underlying emotion connections using Table 3 may shed some light on the differences. Recalling that the lowest performing emotion–font scores are for *anticipation* and *sadness*, one might expected that words associated with these emotions are prone to showing fewer user preferences that are congruent. The words associated with *anticipation*, namely *elegance*, *oracle*, *peaceful*, and *tickle*, do not seem to possess the same difficulty, as the lowest preference for these words is 52% (for *outcome*), which shows a strong preference.

On the other hand, among the words associated with *sadness*, the words *lifeless* and *resign* do not show such strong preferences. One might conjecture that this stems from low-performing emotion–font associations. However, a detailed look reveals that *kill* and *massacre* have the same underlying emotion associations as *lifeless* and *resign*, respectively. The fact that the fonts for *kill* and *massacre* received strong support from users indicates that the word–emotion associations might have played a role. Some words may have inaccurate or missing emotion associations, or some words may have weaker emotional associations than others, which is not reflected in the binary scheme used by EmoLex. Using a dataset with real-valued scores instead of binary associations might help to capture the latter case.

Fortunately, overall, both datasets have received strong support from users, with around 60% and 64% of the average user preferences towards the fonts found to be congruent by our datasets. Only for two words out of twenty-five, incongruent fonts are preferred more frequently than chance would predict, i.e. $\frac{2}{5} = 40\%$. In contrast, for 23 words, congruent fonts are preferred more frequently than chance would predict. Despite the subjective nature of font preferences and associations, we observe that there is a clear correspondence between the fonts chosen by our method and those assessed as appropriate by the human participants.

6. Extension via Semantic Relationships

Finally, we extend the dataset and increase its accuracy by accounting for semantic relationships given by WordNet (Fellbaum, 1998). For all attribute words in $\mathcal{E} \cup \mathcal{A}$, in total 47 attributes (37 original font attributes and 10 emotion attributes for which our study has computed font vectors), we gather the set of words that share a common synset with the attribute names (such as the words *deadening*, *dull*, *ho-hum*, *irksome*, *slow*, *tedious*, *tiresome* and *wearisome* for the font attribute *boring*). We then go through this list manually to exclude any synonyms with an irrelevant meaning (such as the word *building complex* for the font attribute *complex*). The remaining synonyms are assigned the font vectors of the words in $\mathcal{E} \cup \mathcal{A}$. This results in 364 additional word-font assignments, 112 of which override the ones from the methods in Sections 3. and 4. While small in number, these provide for particularly salient associations.

7. Conclusions and Future Work

To the best of our knowledge, no existing tool or resource provides semantic font recommendation support in which the meaning of the text is computationally matched with the semantic attributes of the fonts. Our study aims to support the development of such font recommendation tools.

Following this aim, we have created FontLex[8], a dataset that maps 6.7K words to 200 fonts. These derive mainly from the affective associations between words and fonts. Our evaluation shows an average of 55.95% of selections evincing a preference for the fonts recommended by the dataset. This is a strong result given the subjective nature of such preferences. Our ongoing work is broadening this even further based on further semantic relationships.

As part of the future work, we plan to further expand the dataset by making use of font attributes such as *thin*, *wide*, and *angular*, and their connections with objects, as opposed to the more abstract focus in this paper.

Acknowledgments

Gerard de Melo's research is supported in part by the DARPA SocialSim program. Tugba Kulahcioglu presents her thanks to Anthony Porturas for discussions on an earlier version of this study.

[8] http://gerard.demelo.org/fonts/

8. Bibliographical References

Bloch, M. (2017). 1001 Fonts. `http://www.1001fonts.com`. Accessed: 2017-07-29.

Canva.com. (2017). Canva font combinations. `https://www.canva.com/font-combinations/`. Accessed: 2017-07-29.

Childers, T. L. and Jass, J. (2002). All dressed up with something to say: Effects of typeface semantic associations on brand perceptions and consumer memory. *Journal of Consumer Psychology*, 12(2):93–106.

dafont.com. (2017). `http://www.dafont.com`. Accessed: 2017-07-29.

Data Scope Analytics. (2017). Fontstellations. `https://datascopeanalytics.com/blog/fontstellations/`. Accessed: 2017-07-29.

Fellbaum, C. (1998). Wordnet: an eletronic lexical database. cambridge, massachusetts, eua.

Fligner, A. (2013). *The effect of packaging typeface on product perception and evaluation*. Ph.D. thesis, The Ohio State University.

Fox, D. (2010). Free will for typeface selection: Myth or reality? *Usability News*, 12(1).

Hazlett, R. L., Larson, K., Shaikh, A. D., and Chaparo, B. S. (2013). Two studies on how a typeface congruent with content can enhance onscreen communication. *Information Design Journal*, 20(3).

Ho, K. (2017). Organizing the world of fonts with ai. `https://medium.com/@kho98`. Accessed: 2017-07-29.

Juni, S. and Gross, J. S. (2008). Emotional and persuasive perception of fonts. *Perceptual and motor skills*, 106(1):35–42.

Kulahcioglu, T. and de Melo, G. (2018). Predicting semantic signatures of fonts. In *Proceedings of the 12th IEEE International Conference on Semantic Computing*. IEEE.

Lewis, C. and Walker, P. (1989). Typographic influences on reading. *British Journal of Psychology*, 80(2):241–257.

Mills, H. (2017). Font pair. `http://fontpair.co/`. Accessed: 2017-07-29.

Mohammad, S. M. and Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29(3):436–465.

O'Donovan, P., Libeks, J., Agarwala, A., and Hertzmann, A. (2014). Exploratory font selection using crowdsourced attributes. *ACM Trans. Graph.*, 33(4):92:1–92:9.

Qiao, J. (2017). Fontjoy – Font pairing made simple. `http://fontjoy.com/`. Accessed: 2017-07-29.

Sam Berlow, S. C. and Sherman, N. (2017). Fonts In Use. `https://fontsinuse.com/`. Accessed: 2017-07-29.

Shaikh, A. D., Chaparro, B. S., and Fox, D. (2006). Perception of fonts: Perceived personality traits and uses. *Usability News*, 8(1):1–6.

Shaikh, A. D., Fox, D., and Chaparro, B. S. (2007). The effect of typeface on the perception of email. *Usability News*, 9(1):1–7.

Shaikh, A. D. (2007a). The effect of website typeface appropriateness on the perception of a company's ethos. *Usability News*, 9(2).

Shaikh, A. D. (2007b). *Psychology of onscreen type: Investigations regarding typeface personality, appropriateness, and impact on document perception*. Ph.D. thesis.

Van Rompay, T. J. and Pruyn, A. T. (2011). When visual product features speak the same language: Effects of shape-typeface congruence on brand perception and price expectations. *Journal of Product Innovation Management*, 28(4):599–610.

Velasco, C., Salgado-Montejo, A., Marmolejo-Ramos, F., and Spence, C. (2014). Predictive packaging design: Tasting shapes, typefaces, names, and sounds. *Food Quality and Preference*, 34:88 – 95.

Velasco, C., Woods, A. T., Hyndman, S., and Spence, C. (2015). The taste of typeface. *i-Perception*, 6(4).

Wang, Z., Yang, J., Jin, H., Shechtman, E., Agarwala, A., Brandt, J., and Huang, T. S. (2015). Deepfont: Identify your font from an image. In *Proc. ACM Multimedia*, pages 451–459. ACM.