A method to capture categorization strategies in order to understand user perception.

A case study on user categorization strategies for typefaces.

Abstract: The following paper discusses a modified card sorting procedure; which was used to capture the typeface classification strategies used by a set of people. This method attempts to address the problems raised by rigid card sorting by allowing the categories to be more intuitive and realistic while retaining its statistical validity. The overall goals of this study were to understand how people perceive Devanagari typefaces. Data was collected and treated in a manner that allowed for it to be subjected to a variety of qualitative as well as quantitative techniques. This paper shows how the data was consolidated using agglomerative hierarchical cluster analysis within which Ward's-minimum variance was used to calculate the distance between the clusters. The paper suggests that user's are more comfortable when they are given the option of placing items in multiple groups, and that it is possible to validate such data.

Key words: Typeface Classification, User-Perception, Hierarchical Clustering, Card sorting.

1. Introduction

Studying how people classify a group of items has long been a way to understand how people perceive a given set of items and the relations between them. In recent years, card sorting has been a popular tool for researchers trying to understand people's mental models. Primarily two variations of card sorting are used: Open card sorting, wherein participants are asked to create classes and closed card sorting where participants are given a set of items and a set of classes. Open card sorting is used to generate classes whereas closed card sorting is used more to verify classes created by open sorting. Some authors [8, 11] have written introductions to card sorting and the differences between open and closed card sorts. However, within these two techniques if the given data contains too many interlinkages amongst its elements, participants often express the need to place a single item in two or more classes. In order to incorporate this need; in this study participants were allowed to place a single item into more than one category if they so wished, so as to create fuzzy categories.

Card sorting data can be analyzed by a variety of methods; the selection of the method depends upon the goal or intent of the study. Cluster analysis is at times used to statistically analyze card sorting data and to visualize its results if the intent of the study is to create a highly structured setting. Factor analysis has also been used [1] to analyze card sort data where the intent of the study was to identify attributes or properties used by participants to assess categories. Visual examination or "eyeballing" [10] has also been used to gauge the overall organizational structure of card sorts. Since one of the goals of this study was to create a well-defined classification system for linear mediums such as font catalogues, books and menu structures for word processing softwares (a more intuitive typeface selection scheme); cluster analysis was used to come up with distinct clusters for Devanagari typefaces.

1.1 Goals of the Study

The study was aimed at examining how various individuals classify a set of Devanagari fonts. An experiment was designed to capture the different Devanagari font classification strategies used by people through a modified card sorting methodology. The primary objectives of this experiment were:

1. To understand the overall strategies that participants use to classify Devanagari typefaces.
2. To identify the significant parameters used to classify Devanagari typefaces.
3. To record the nomenclature given by participants to various font classes as well as parameters.
4. To know which parameters are considered and given preference while creating classes and which are overlooked.
5. To come up with a classification system for linear mediums such as font catalogues, books and linear menu structures within softwares for a user-friendly selection of typefaces.

The first four objectives listed above are part of a larger study and hence have not been discussed in detail in this paper. This paper focuses mainly on the accomplishment of the fifth objective.

2. Pilot Experiment

During the initial stages of the study, a pilot experiment was conducted to test various experimental procedures and data recording techniques. The pilot was done on four subjects, and was independently evaluated by two experts. The stimulus card size for the pilot experiment was 26cm x 18cm, four lines of nonsense Devanagari text at 72 point size (normalized *kana* height 1.25cm) (for a reference to the anatomy and terminology of Devanagari letters, refer [3]).

Fourteen nonsense words were created with the above characters and were set in fifty font samples. They were constructed in such a manner so as to be the conceptual equivalent of the English pangram, “the quick brown fox jumps over the lazy dog”. Typeface sampling was partially random (from a list of typefaces published by CDAC [Centre for Advanced Computing]) and partially selective (fonts in popular use were selected). The nonsense Devanagari text for the pilot was:

झिसुपभ ऐउऋकु घिडेग ओचीऊत र्मधळ ट्रेएषअः।

ईनएँट्र अक्षव आशम औदीजुब इथढु अंज्ञरफ्रै छणलठि त्रॅफीयू ॥

These sample lines contained: Independent vowels (अ आ इ ई उ ऊ ऋ ए ऐ ओ औ अं अः), dependent vowels signs (◌ा ◌ि ◌ी ◌ु ◌ू ◌े ◌ै ◌ो ◌ौ ◌ृ), consonants (क ख ग घ च छ ज झ ट ठ ड ढ ण त थ द ध न प फ ब भ म य र ल व श ष स ह ळ), three frequently used conjuncts (क्ष ज्ञ त्र), diacritic marks (◌ँ ◌ॅ ◌ं ◌ः), *halant* sign (◌्), three र marks (*Rephar*, *Rashtra* mark and diagonal line to show conjunction with र), and punctuation marks (। ॥)

The evaluation of the pilot experiment revealed that the sample size and stimulus size was large; and participants had difficulty (physically placing the items) in handling and arranging the given samples. It was also observed that participants initially had difficulty in comprehending the task to be performed, but they picked up later on as the experiment progressed. As a result of this the stimulus was cut short to one line for the main experiment. The samples were reduced to thirty fonts and a warm up classification exercise was created, which was to be administered before the actual exercise took place.

3. Main Experiment

3.1 Stimulus

The stimulus size for the main experiment was 25cm x 4cm, (see Fig.1) one line of non-sense Devanagari text at 72 point size (normalized *kana* height 1.25cm). The nonsense line with the letters and its signs combined was:

झिखृफुद्धी बुठ्ठभीश् किट्रेर्न आंग्रै

Sampling of the thirty font samples was partially random (from a list of typefaces published by CDAC) and partially purposive (fonts in popular use and distribution were selected). The selection of the fonts was as follows:

Randomly selected fonts were: Aakash, Akshar, Alankar, Basant, Bhima, Dhruv, Gandhar, Ganesh, Kishor, Latika, Megha, Meghadoot, Mohini, Prakash, Sahaj, Samata, Shridhar, Subodh, Vallabh, Vasundhara, Vinit, Yamini, and Sanskrit 2003.

Purposely selected fonts (and the reason for selection) were: Mangal (default Microsoft font for Windows XP), Natraj (used in Hindi and Marathi state textbooks), Yogesh (selected for its popularity), Surekh (selected for its popularity), Raghu (open source typeface supplied in some Linux distributions), NID-Mahendra (typeface designed for use at the National Institute of Design), Arial Unicode MS (supplied with Microsoft Office 2002 and 2003). The sample card sheets were randomly numbered, with the number being placed on the lower right corner of the sample sheet.

3.2 Participants

The exercise was administered on thirty-eight participants (twenty males and eighteen females). Their age ranged from 21 years to 42 years, with an average of 25 years (S.D. = 2 Years). All the participants had completed a minimum level of education and had formally studied Hindi, Marathi or Sanskrit languages till the 10th grade.

3.3 Experimental Procedure

The experiment began with the participants being given a warm up exercise. In the warm up exercise, each participant was given a set of thirty three small cards. On each of the card was a name of an animal. Participants were told to classify them into as many categories as they wished; and that they could have multiple categories within a category. It was also stated that one animal could belong to more than one category and that they should arrange the cards according to the categories that they have made. After they had completed classifying the animals, they were asked to name the groups and state the reasons (basis) for which they were categorized together. The warm up exercise ensured that the participants cleared up all their doubts about the fuzzy classification exercise.

The main procedure for the experiment was conducted after the warm up exercise. The participants were given thirty cardboard mounted font sample cards. They were then asked to classify them according to any logic that they pleased; with no restriction on the number of categories that they could make. It was also stated that they could have as many sub-categories as they wanted within the main categories that they created. The classification procedure took around 30 to around 120 minutes to complete. After they had completed classifying participants were asked to give names to each of the classes which had been created; and elucidate the basis on which each of the groups was made. They were also asked to state all the properties of the classes that they had created. When the participants had completed classifying and naming the groups they were asked to reconfirm their classification scheme, and see if they wanted to make any changes (to see if their classification was consistent with the applied logic). Once the classification scheme was confirmed, the categories and the properties attributed to them were recorded on a data collection sheet.

It was essential here that users were given the option of placing an element into two or more classes. There is sufficient proof [7, 13] that mental categorization is a graded and fuzzy phenomena. In typefaces, a particular typeface can be categorized considering several properties, and a typical mental model could have multiple rationales, and viewpoints towards the grouping of typefaces. For example, consider the typefaces displayed in figure 3: the first letter (a) represents the traditional Devanagari text (written with a right canted reed pen with modulated stroke thickness), the third letter (c) depicts the same letter drawn with a mono-linear pen (flat terminals with perceptually un-modulated stroke thickness). These two groups, traditional Devanagari and mono-linear Devanagari can be considered to be two primary groups of typefaces. It may be noted that the second letter (figure 3 (b)) contains properties of both the earlier formed groups (a right cant at the terminal, as well as an un-modulated stroke). Users wish to place this typeface in both of the above groups. Many users combined multiple rationales for typeface classification, i.e. typefaces were grouped partially based on formal elements (such as written with a traditional Devanagari pen) and partially on the basis of usage (like decorative text or headline text) or subjective keywords (such as energetic text, professional text). Multiple rationales necessarily demanded most of the fonts to be placed in two or more groups.

3.4 Data Recording

Data was recorded in a tabular fashion, with the categories and subcategories being written in a top to bottom fashion. During the preparation of the samples, each of the samples was randomly assigned a number. After the participants had completed classifying the given samples, the groups were recorded with the help of the numbers assigned to each of the samples. The top most sample numbers were recorded first in a tabular fashion; the name and the description provided by the participants was then recorded besides the recorded numbers along with the associated properties of each class.

3.5 Treatment of Data

Clustering algorithms require data to be specifically formatted before it can be processed. Converting card sort data into these formats is not an obvious process and is not usually discussed in literature on card sorts [1]. Capra has discussed one such method for the preparation of card sort data for factor analysis. A method to prepare card sort data for cluster analysis is presented here.

The recorded tabular data was used to create a proximity matrix of the thirty samples. The matrix was created so as to be treated by a clustering algorithm. The matrix here is a similarity matrix rather than a dissimilarity matrix. The proximity matrix p, is hence an m by m matrix (m is sample size and is thirty in our case) containing all the pair wise similarities between the samples considered. If x_i and x_j are the i^{th} and j^{th} objects respectively, then the entry at i^{th} row and j^{th} column of the proximity matrix is the similarity s_{ij} between x_i and x_j. Here in our case, the number of times a typeface sample was grouped together at any level by the participant increments the similarity in the proximity matrix.

A sample creation of the proximity matrix for a hypothetical 8 font sample test is shown as:

Consider participant one has grouped the sample numbers as follows,

Group One

(2, 5, 6, 7) (level 1)

(2, 6) and (5, 7) (level 2) and

Group Two

(1, 3, 4, 8) (level 1)

(1, 3) and (4, 8) (level 2)

The initial proximity matrix would be as follows,

	One	Two	Three	Four	Five	Six	Seven	Eight
One	X	0	0	0	0	0	0	0
Two	0	X	0	0	0	0	0	0
Three	0	0	X	0	0	0	0	0
Four	0	0	0	X	0	0	0	0
Five	0	0	0	0	X	0	0	0
Six	0	0	0	0	0	X	0	0
Seven	0	0	0	0	0	0	X	0
Eight	0	0	0	0	0	0	0	X

The building of the proximity matrix is initiated by incrementing by one the similarity value of the first pair of samples, consecutively till all the sample pairs have been iterated through.

During the first pass the first group is processed, at level 1, the following table is built.

	One	Two	Three	Four	Five	Six	Seven	Eight
One	X	0	0	0	0	0	0	0
Two	0	X	0	0	1	1	1	0
Three	0	0	X	0	0	0	0	0
Four	0	0	0	X	0	0	0	0
Five	0	1	0	0	X	1	1	0
Six	0	1	0	0	1	X	1	0
Seven	0	1	0	0	1	1	X	0
Eight	0	0	0	0	0	0	0	X

At the first group, level 2, the following table is achieved.

	One	Two	Three	Four	Five	Six	Seven	Eight
One	X	0	0	0	0	0	0	0
Two	0	X	0	0	1	2	1	0
Three	0	0	X	0	0	0	0	0
Four	0	0	0	X	0	0	0	0
Five	0	1	0	0	X	1	2	0
Six	0	2	0	0	1	X	1	0
Seven	0	1	0	0	2	1	X	0
Eight	0	0	0	0	0	0	0	X

The pseudo code for the entire processing of the data of a selected group at any given level is as:

initialize proximity_matrix

```
count = sample_group.count
selection = sample_group.value
For i = 1 to count
    For j = 1 to count
    proximity_matrix(selection(i,1),selection(j,1))+=1
    Next j
Next i
Transpose proximity_matrix
```

The above operation was carried out for each category created by the participants. This matrix was then used as an input for the cluster analysis algorithm.

4. Experimental analysis

4.1 Cluster Analysis:

Cluster analysis groups objects based on information found in the data describing the objects or their relationships [16]. The goal is that the objects in a group should be similar to one another and different or dissimilar from the objects in the other groups. The greater the similarity within the group and the greater the dissimilarity between the outside groups, the better the clustering. Since one of the objectives of this study was to create distinct typographic categories for practical use; hierarchical cluster analysis was used.

Hierarchical Methods:

A hierarchical clustering method works by grouping data elements into a tree of clusters. Hierarchical clustering methods are of two kinds: agglomerative clustering where the hierarchical decomposition is formed in a bottom up merging manner and divisive clustering where the hierarchical decomposition is formed in a top-down (splitting) fashion.

Agglomerative hierarchical clustering: [5] The bottom-up scheme is initiated by placing each element in its own cluster and then merging these atomic clusters into larger clusters, until all of the elements part of a single cluster.

Divisive hierarchical clustering: [5] The top-down strategy is the opposite of agglomerative clustering, and is initiated with all elements being in one cluster. It then subdivides the cluster into smaller clusters, until each object forms a cluster on its own or until it satisfies certain termination conditions, such as a desired number of clusters is obtained. A tree structure known as a *dendrogram* is used to visually represent the process and result of hierarchical clustering.

Five widely used measures for distance between clusters are as follows, where $|p - p'|$ is the distance between two objects or points, p and p', m_i is the mean for cluster, C_i and n_i, is the number of objects in C_i.

Minimum distance: $d_{min}(C_i, C_j) = min_{p \in C_i,\ p' \in C_j} |p - p'|$

Maximum distance: $d_{max}(C_i, C_j) = max_{p \in C_i,\ p' \in C_j} |p - p'|$

Mean distance: $d_{mean}(C_i, C_j) = |m_i - m_j|$

Average distance: $d_{avg}(C_i, C_j) = \frac{1}{n_i n_j} \sum_{p \in C_i} \sum_{p' \in C_j} |p - p'|$

An algorithm using the minimum distance between clusters, it is known as the nearest-neighbor clustering algorithm. Average linkage has a tendency to join clusters with small variances, and it is slightly biased toward producing clusters with the same variance; since it considers all members in the cluster rather than just a single point, however, average linkage tends to be less influenced by extreme values than other methods. Complete linkage is strongly biased toward producing compact clusters with roughly equal diameters, and it can be severely distorted by moderate outliers. Complete linkage ensures that all items in a cluster are within some maximum distance of one another.

If the clustering process is terminated when the distance between nearest clusters exceeds an arbitrary threshold, it is called a single-linkage algorithm. Single linkage sacrifices performance in the recovery of compact clusters in return for the ability to detect extended and irregular clusters Also, single linkage tends to chop off the tails of distributions before separating the main clusters. This has many desirable theoretical properties but has fared poorly in Monte Carlo studies [13] and hence was rejected for this study.

In Ward's minimum-variance method, the distance between two clusters is the ANOVA sum of squares between the two clusters added up over all the variables [2]. While creating each cluster the within cluster sum of squares is minimized over all partitions obtainable by combining two clusters from the previous creation. The sums of squares are easier to interpret when they are divided by the total sum of squares to give proportions of variance (squared semi-partial correlations). Ward's method tends to join clusters with a small number of observations, and strongly tends towards producing clusters with the same shape and with roughly the same number of observations.

For the i^{th} cluster, the Error Sum of Squares is defined as ESS_i = sum of squared deviations from the cluster centroid. If there are C clusters, the Total Error Sum of Squares is defined as:

$$\mathrm{ESS}(C_i) = \sum_{a=1}^{n_i} |x_a - \frac{1}{n_i}\sum_{b=1}^{n_i} x_b|^2$$

Consider the union of every possible pair of clusters. The two clusters would then be combined whose combination results in the smallest increase in ESS. The distance between clusters can hence be calculated as

Ward's Minimum-Variance Method: $d_{\mathrm{ward}}\left(C_i, C_j\right) = ESS\left(C_i, C_j\right) - [ESS(C_i) + ESS\left(C_j\right)]$

For this study Ward's minimum variance method was chosen over the others, because:

1. It is good at recovering cluster structure, and yields unique and exact hierarchy.[4]
2. It does not leave any "loose ends". No clusters with only one or a few elements. All data is grouped in bite size chunks, which can be studied further.[6]
3. Aberrant points are also grouped together, which might not have anything in common with each other except for the fact that they are dissimilar from the other objects.

The Clusters generated (*dendrogram*) by using Ward's method on the earlier created proximity matrix, are shown in figure 2.

4.2 Cluster Validation

Validating the results of cluster analysis requires subjective decision making [12]. Besides primary validity, which measures how valid the cluster analysis is overall, there are a few measures of secondary validity which assess whether the clusters have certain desirable properties. The following cluster validation techniques were used to evaluate this study:

1. *Agreement of Different Multivariate Methods*:[12] As a measure of validity some researchers have used the agreement of classifications produced from the same data matrix processed by different multivariate methods—for example, methods such as cluster analysis, principal component analysis, multidimensional scaling, and factor analysis. In accordance to the earlier studies, the collected data was also analyzed by factor analysis. A method similar to the one used by Capra [1] was used. A high level of concurrence was found between the groups formed through factor analysis and the basic categories generated through cluster analysis.

2. *Agreement of Classifications Based on Split Samples of Data:* Sinha [15] has suggested that the objects of the original data matrix be randomly split into two subsamples and separate cluster analyses be run on each to produce two separate classifications. To be judged valid, the two classifications should agree: the number of classes should be the same and their defining attributes should be the same. As part of the cluster validation scheme, the resulting clusters were validated incrementally, first on a set of twelve participants then on a set of twenty-four participants and then on the final set of thirty eight participants. During each of these stages the clusters remained fairly stable and only one or two members changed their position within the lowest sub-categories.

5. Results

The resulting dendrogram was subjected to the cutoff line (shown by the dashed line in the result sheet). The cut-off line was chosen by looking at the major (graphical) groups within the clusters, since "the cut-off points to identify clusters … are a matter of subjective judgment by the researchers" [9]. After the cut-off point the resulting clusters can be seen in fig 2.

Hierarchical cluster analysis yielded five stable and distinct clusters.

Fonts from the first cluster (fig.2 (1)) starting with the left have been labeled by many participants as "Traditional text", "Standard Text", "Textbook or Newspaper Text". The formal properties of the whole group indicate that all of them have been drawn with the help of a canted pen (8 right canted, 1 left canted), due to which one sees canted vertical terminals, at inclined axis and high contrast among its letters. This group is further classified into three subgroups; based on their visual characteristics one can notice that the first two groups have been differentiated according to their counters (open vs. closed).

The second cluster (fig.2 (2)) in the resulting set is that of the "scripts", the only common visual feature that all the elements share is that their vertical terminals have a "swoosh" which either goes to the right or to the left in a smooth manner. Two of them have an inclination towards the right. They are further bifurcated in to two other subgroups, the first one seems to the one where the fonts have a darker grey value and have a narrower appearance than the regular fonts. The second of the sub groups are the ones which are wider in their appearance and have a lighter grey value. Two fonts in this group have an inclination towards the right. One font instead of having a "swoosh" as the vertical terminal has a rounded (felt-tip pen) terminal.

In the third group of clusters (fig.2 (3)), the significant common visual feature seems to be the serif. These are all fonts influenced by Latin typefaces. There is a considerable amount of variation in the serifs of the fonts; from single sided to two sided, thin, thick and inclined. There also is a considerable difference in the width of the fonts; two narrow fonts form a sub group, while three broad fonts form the other sub group. There is further differentiation in this group as fonts with medium grey value and fonts with dark grey value.

In the fourth cluster (fig.2 (4)) the prominent visual feature of this group is the fact that all the fonts in this group are mono-linear i.e. the strokes are of uniform thickness. All the elements have a horizontal vertical terminal except in one font sample which has a traditional Devanagari inclined terminal. This group is further subdivided into two categories based on the typographic color; the medium weights form one group while the darker weights form the other group.

The fifth cluster (fig.2 (5)) consisting contains typefaces which are either dark in their grey value, and/or are very broad or tall. The description that some of the participants have given is that of "display typefaces" or typefaces that one would use for headlines. They are again sub-grouped according to their weights; the first subgroup is that of the thick, dark or bold fonts; while the other is group of varying widths i.e. narrow and broad fonts.

6. Conclusion

Understanding how people perceive a set of objects and the relationships that they associate between those objects is a complex phenomena which cannot be addressed by quantitative analysis alone. In order to have a complete understanding of the users' mental model a combination of qualitative and quantitative techniques must be used. In this paper a method has been presented to capture fuzzy categorization schemes, through a modified card sorting procedure. It also demonstrates a way to prepare the recoded data for statistical analysis. Hierarchical clustering is one way by which a fuzzy card sort data can be transformed in to distinct classes. Capra[1] has already shown how factor analysis can be used to created to fuzzy classes for similar data. The fuzzy (multiple) input ensures a more intuitive and realistic data collection technique for users. When the study is oriented towards a practical application, such as the creation of a rigid classification system then hierarchical clustering seems to be an appropriate method. This method has an advantage over eyeballing card sort data, which can get quite cumbersome when the data is quite large.

7. References

[1] Capra, M. G. (2005) Factor analysis of card sort data: an alternative to hierarchical cluster analysis. In *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting,* 691–695.

[2] Clark, V. (2004) *Sas/Stat 9.1: User's Guide* SAS Publishing.

[3] Dalvi, G. (2009) Anatomy of Devanagari Typefaces. In *Design Thoughts Vol. I,* Industrial Design Centre, IIT Bombay.

[4] Fox, E. A. (N.D.). *CL Algorithm Details*. Retrieved January 26, 2008, from http://ei.cs.vt.edu/~cs5604/f95/cs5604cnCL/CL-alg-details.html

[5] Han Jiawei, (2006). *Data Mining: Concepts and Techniques*. Morgan Kaufmann.

[6] Kleiweg, P. (N.D) [Online Article] How to determine the best clustering of an area. [Accessed 11, September 2007], Available at <http://www.let.rug.nl/kleiweg/kaarten/MDS-clusters.html.en>

[7] Lakoff, G. (1990). *Women, Fire, and Dangerous Things.* University Of Chicago Press.

[8] Maurer, D. &. (2004) Card sorting: a definitive guide, from Boxes and Arrows: [Accessed 14, December 2005] Available at <http://www.boxesandarrows.com/archives/card_sorting_a_definitive_guide.php>

[9] Mukherjee, R. (1983) *Classification in social research*. SUNY Press.

[10] Nielsen, J. (1994) User Interface Design for Sun Microsystem's Internal Web. In *Electronic Proceedings of the Second World Wide Web Conference '94: Mosaic and the Web.*

[11]Robertson, J. (2001) *Information Design Using Card Sorting* [Accessed 12, April 2005] Available at Step Two Designs <http://www.steptwo.com.au/papers/cardsorting/>

[12] Romesburg, C. (2004) *Cluster Analysis for Researchers* Lulu.com .

[13]Rosch, E.H. (1973) Natural categories, In *Cognitive Psychology 4 pp* 328-50.

[14] Schmidhammer, J. L. (N.D.) *Applied Multivariate Methods* Retrieved [Accessed 10, August 2008] Available at <www.bus.utk.edu/stat/Stat579/Hierarchical%20Clustering%20Methods.pdf>

[15] Sinha, R. N. (1977). *Uses of multivariate methods in the study of stored-grain ecosystems* In Environmental Entomology. 6: 185-192

[16] Wille, L. (2004) *New Directions in Statistical Physics: Econophysics, Bioinformatics, and Pattern Recognition* Springer.

7. Figures

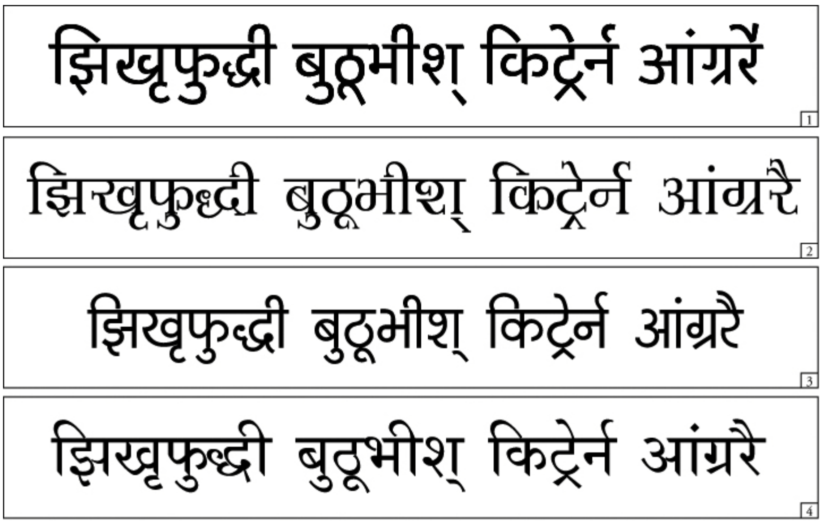

Figure 1: Samples given participants during the study.

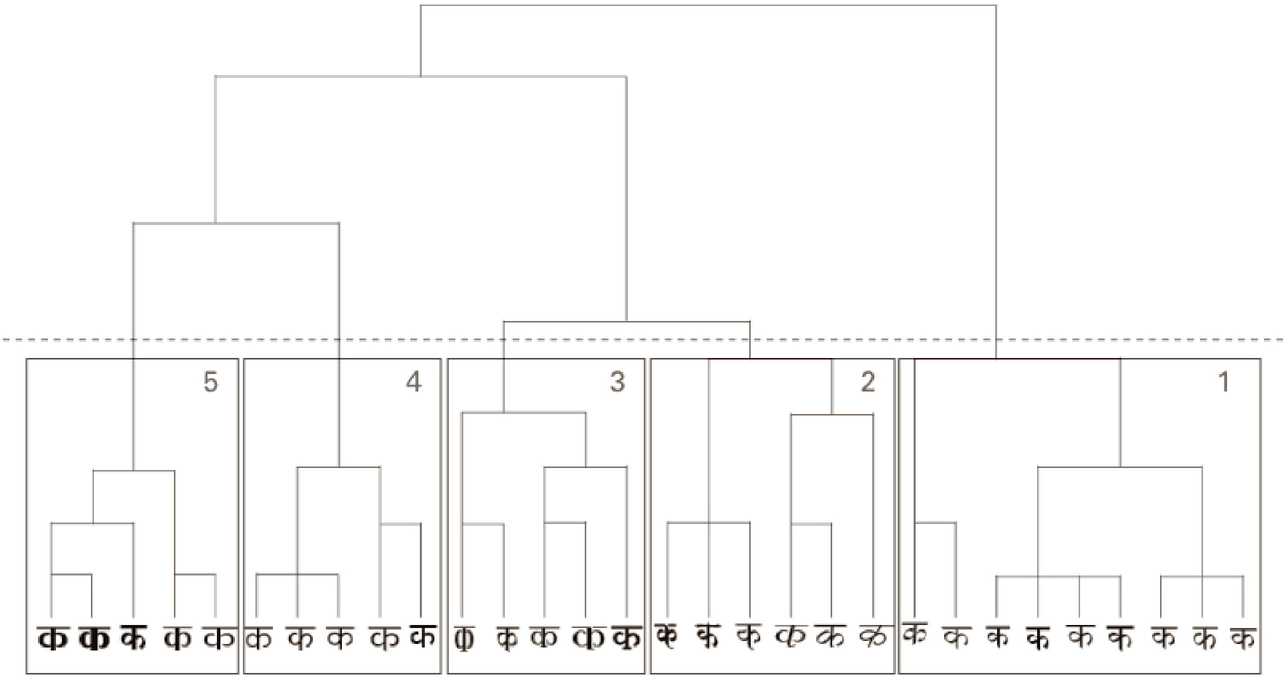

Figure 2: Resulting tree, cutoff, and final clusters.

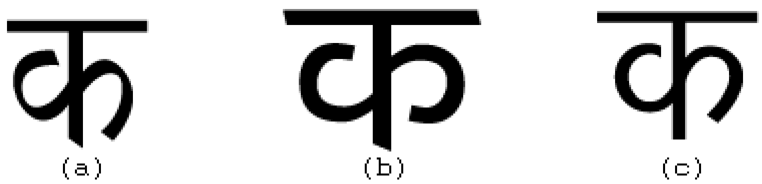

Figure 3: Samples of three fonts, with a possible fuzzy font in the middle.